U.S. DEPARTMENT OF ENERGY

Office of Science

2014 Genomic Science Contractor-Grantee Meeting XII

Abstract Book

February 10-12, 2014

Sponsored by the
U.S. Department of Energy
Office of Biological and Environmental Research
Biological Systems Science Division

This document was produced under contract number DE-AC05-06OR23100 between the U.S. Department of Energy (DOE) and Oak Ridge Associated Universities (ORAU).

The research grants and contracts described in this document are supported by the DOE Office of Science, Office of Biological and Environmental Research, Biological Systems Science Division.

Welcome

Department of Energy
Washington, DC 20585
February 10, 2014

Dear Colleague:

On behalf of the Biological Systems Science division within Department of Energy's (DOE) Office of Biological and Environmental Research (BER), welcome to the 2014 Genomic Science Annual Contractor-Grantee Meeting! We are looking forward to another outstanding meeting and are eager to meet with you and discuss your latest scientific accomplishments. I'm pleased to announce that we have full representation from the DOE National Laboratory programs engaged in Genomic Science Programs (GSP) research this year, unlike last year. We are still operating under rather significant restrictions for hosting this meeting but hope this does not detract too much from the scientific discussions.

I would like to extend a special welcome to new PIs to program. Use this opportunity to confer with colleagues in the program, discuss new ideas and survey the resources available to you within the program. This is an annual opportunity to view your research in the context of the larger program and among colleagues engaged in complementary research. I would also like to extend a welcome to our colleagues from other DOE offices and Federal agencies. Please feel free to discuss your program(s) with researchers at this meeting working in similar areas.

This year's meeting will highlight exciting research by individual investigators, interdisciplinary research teams, user facilities such as DOE's Joint Genome Institute (JGI), and the Bioenergy Research Centers (BRCs), all of which are supported by the Genomic Science program. The BRCs have entered their second 5-year phase of operations, continuing their extremely productive programs. The three centers will report on exciting new bioenergy research in the plenary sessions as well as the poster sessions.

The biosystems design projects are into their second year and several projects will report on the latest advances in metabolic engineering and redesign of biological processes for biofuel production. The field as a whole has seen several major breakthroughs this past year and we are pleased to have Dr. George Church, one of the most impactful leaders in the field, as this year's keynote speaker. Also, the final plenary session will feature recent Early Career award recipients, whose projects are focused on development of a broader set of plant and microbial platform organisms for biosystems design research.

This year's breakout sessions will cover a broad range of BER mission-relevant topics, from modeling and understanding biological networks to systems biology analysis of the interaction between organisms and their interface with the environment. One area we would like to emphasize is how to integrate and translate the detailed genome-based understanding of plants and microbes into broader predictions of the sustainability of bioenergy production practices and/or ecosystem responses to environmental change. A draft report on a recent workshop (Oct 2103) entitled "Research for Sustainable Bioenergy" highlighting the scientific challenges for sustainable bioenergy research will be presented.

The DOE Systems Biology Knowledgebase (KBase) will roll out its next set of capabilities focusing on the "Narratives" component of the project, a social networks-inspired interface to facilitate and promote scientific collaboration among GSP researchers. Through the Narratives, KBase will provide access to powerful genomic analysis and modeling tools allowing rapid analyses via cloud-based computing on

DOE's state of the art computational systems. An overview of KBase accomplishments and vision for the future will be presented by Dr. Adam Arkin (lead PI) followed by an interactive breakout session including scientific demonstrations and tutorials organized by the KBase developers. These presentations will highlight KBase's current capabilities and vision for the future.

The Genomic Science program provides community access to unique technological capabilities at DOE Office of Science National Scientific User facilities. Plenary talks from the directors of the JGI and the Environmental Molecular Sciences Laboratory (EMSL) will highlight the latest high-throughput genome analysis capabilities and technologies. Imaging and analytical technologies including those provided by JGI and EMSL will be featured in a breakout session on technology-enabled systems-biology research.

Thank you for all your efforts to help make this program a success. Scientific advances from this program continue to have a profound impact on solutions to some of the most daunting energy and environmental challenges facing DOE, the nation and indeed the world. It is our pleasure to work in partnership with you on these challenges. Enjoy this year's PI meeting.

Sincerely,

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

Agenda

Genomic Science Program
2014 Contractor-Grantee Meeting
February 10-12, 2014

SUNDAY, FEBRUARY 9TH

5:00 PM – 8:00 PM	**Early Registration and Poster Setup**

MONDAY, FEBRUARY 10TH

7:00 AM – 8:30 AM	**Registration** Location: Independence Foyer
7:30 AM – 8:30 AM	**BREAKFAST ON YOUR OWN**
8:30 AM – 9:00 AM	**Welcome and Introduction to the Meeting** Location: Ballroom E/F
8:30 AM – 8:40 AM	**Sharlene Weatherwax**, Associate Director, DOE Office of Biological and Environmental Research Opening Remarks
8:40 AM – 8:55 AM	**Todd Anderson**, Director, Biological Systems Science Division, DOE BER Introduction to the Genomic Science Program
8:55 AM – 9:00 AM	**Pablo Rabinowicz**, Program Manager, DOE BER Announcements
9:00 AM – 10: 30 AM	**Plenary Session: DOE Bioenergy Research Centers** Location: Ballroom E/F Moderator: Kent Peters

Speakers:

9:00 AM – 9:30 AM	Sarah Blumer-Schuette (Bioenergy Science Center - North Carolina State University), Caldicellulosiruptor Species –Unique Functions, Genetics, and Comparative Genomics for an Extremely Thermophilic Cellulolytic CBP Microbe
9:30 AM – 10:00 AM	Seema Singh (Joint Bioenergy Institute), Deconstruction – Ionic Liquid Pretreatment: Are We There Yet?
10:00 AM – 10:30 AM	Brian Fox (Great Lakes Bioenergy Research Center), Discovery of Multi–Functional Enzymes and Their use in Time–Resolved Studies of Biomass Deconstruction

10:30 AM – 11:00 AM	**BREAK**

MONDAY, FEBRUARY 10TH

11:00 AM – 12:00 PM

Keynote Presentation: George Church (Harvard Medical School), New Genome Engineering Technologies and Applications
Location: Ballroom E/F
Moderator: Pablo Rabinowicz

12:00 PM – 2:00 PM

LUNCH ON YOUR OWN

2:00 PM – 5:00 PM

Breakout Session A: Systems Biology Approaches for Microbial Community Analysis
Location: Ballroom E
Moderator: Joseph Graber

Description of Session:
Systems biology research is rapidly moving beyond single organism studies and extending these approaches to complex microbial communities. Increasingly sophisticated techniques for meta–omics analyses, molecular–scale process visualization, and data integration offer new opportunities for investigating the structure and functional properties of microbial communities and their interactions with their environments. The presentations in this session will highlight recent breakthroughs in –omics enabled analysis of microbial communities and discuss how these approaches are refining our understanding of microbially–mediated processes.

Speakers:

2:00 PM – 2:30 PM	Eric Alm (Massachusetts Institute of Technology), The ENIGMA Science Focus Area 100 Well Survey: Linking Biology and Geochemistry in a Contaminated Watershed
2:30 PM – 3:00 PM	Jennifer Pett–Ridge (Lawrence Livermore National Lab), Systems Biology of a Microbial Mat: Energy Flows at Both Population and Single Cell Scales
3:00 PM – 3:30 PM	**BREAK**
3:30 PM – 4:00 PM	Mila Chistoserdova (University of Washington), Methane Metabolism in Lake Sediments: A Community Perspective Through Microcosm Manipulations
4:00 PM – 4:30 PM	Trent Northen (Lawrence Berkeley National Lab), Systems Approaches for Understanding Soil Metabolite Foodwebs
4:30 PM – 5:00 PM	Steve Lindemann (Pacific Northwest National Lab), Interrogating the Interspecies Interactions that Drive Microbial Community Assembly: From Phylotype to Function

2:00 PM – 5:00 PM

Breakout Session B: Understanding and Modeling Biological Networks
Location: Ballroom F
Moderator: Jay Fitzgerald

Description of Session:
To enable the next wave of organismal engineering, a deeper understanding of the complex biological networks governing metabolism is necessary. It is increasingly clear that genetic or environmental alterations produce wide–ranging effects on living organisms best understood in the context of regulatory and metabolic models. This breakout session is focused on highlighting recent breakthroughs in understanding and

MONDAY, FEBRUARY 10TH

modeling biological networks in diverse systems, ranging from microbes and microbial consortia to plants as well as the development of predictive computational frameworks.

Speakers:

2:00 PM – 2:05 PM	Introductory Comments
2:05 PM – 2:35 PM	Costas Maranas (Pennsylvania State University), Using MetRxn for Metabolic Model Reconstruction, Flux Elucidation and Redesign
2:35 PM – 3:05 PM	Laura Hug (University of California, Berkeley), Rapid Binning and Metabolic Profiling of Subsurface Microbial Community Metagenomic Data Via an Interactive Online Knowledgebase
3:05 PM – 3:30 PM	**BREAK**
3:30 PM – 4:00 PM	Andrew Allen (J.C. Venter Institute), Regulation of Cellular Nitrogen Metabolism in the Model Marine Diatom *Phaeodactylum tricornutum*
4:00 PM – 4:30 PM	Sergei Maslov (Brookhaven National Lab), Universal Distribution of Component Frequencies in Biological and Technological Networks
4:30 PM – 5:00 PM	John Morgan (Purdue University), Kinetic Modeling in Plant Secondary Metabolism

2:00 PM – 5:00 PM

Breakout Session C: Technology–Enabled Systems–Biology Research
Location: Washington AB
Moderator: Roland Hirsch and Dean Cole

Description of Session:
The Genomic Science program supports basic research that includes the application and development of a variety of imaging and analytical technologies. The biological challenge for these technologies is to simultaneously measure multiple chemical and biological species at multiple spatial and temporal scales within complex, heterogonous cellular and environmental systems. This breakout session will discuss the current capabilities of key technologies and relate how they can be used to address significant biological problems of interest to the genomic science community.

Speakers:

2:00 PM – 2:05 PM	Introductory Comments
2:05 PM – 2:30 PM	Chris Hunt (University of Wisconsin), Confocal Fluorescence Microscopy/Modeling of Lignolytic Mechanisms
2:30 PM – 2:55 PM	Sunney Xie (Harvard University) and Shi-You Ding (National Renewable Energy Laboratory), Stimulated Raman Scattering (SRS) and Atomic Force Microscopy (AFM) of Enzymatic Deconstruction of Plant Cell Walls
2:55 PM – 3:20 PM	Paul Langan (Oak Ridge National Lab), Neutron and X–Ray Experiments and Computational Modeling of Pretreatment of Biomass

MONDAY, FEBRUARY 10TH

	3:20 PM – 3:40 PM	**BREAK**
	3:40 PM – 4:05 PM	Ruth Richardson (Cornell University), Application of Proteomics to Understand Microbial Iron and Sulfate Reduction
	4:05 PM – 4:30 PM	Jeremy Semrau (University of Michigan), Metal Uptake by Methanotrophs: Genetic Basis for the Biosynthesis of A Novel Chalkophore and Molecular Spectroscopic Analyses of Mercury Detoxification
	4:30 PM – 5:00 PM	Haw Yang (Princeton University), Optical Imaging/Quantum Dots for Studying Cellular Uptake of Nanoparticles
5:00 PM – 7:00 PM	**Poster Session (odd-numbered posters)** Location: Independence Center	

TUESDAY, FEBRUARY 11TH

7:30 AM – 8:30 AM	**BREAKFAST ON YOUR OWN**	
8:30 AM – 10:00 AM	**Plenary Session: Biosystems Design** Location: Ballroom E/F Moderator: Pablo Rabinowicz	
	Speakers:	
	8:30 AM – 9:00 AM	Gregory Stephanopoulos (Massachusetts Institute of Technology), Design Criteria for Engineering Overproducing Oleaginous Microbes
	9:00 AM – 9:30 AM	Ryan Gill (University of Colorado, Boulder), New Technologies to Advance the Design, Build, and Test Cycle for Engineering Microbial Systems
	9:30 AM – 10:00 AM	John Cushman (University of Nevada, Reno), Engineering Crassulacean Acid Metabolism (CAM) to Improve Water–use Efficiency of Bioenergy Feedstocks
10:00 AM – 10:30 AM	**BREAK**	
10:30 AM – 12:00 PM	**Plenary Session: DOE User Facilities & Community Resources** Location: Ballroom E/F Moderator: Dan Drell	
	Speakers:	
	10:30 AM – 11:00 AM	Eddy Rubin (Lawrence Berkeley National Lab), JGI Capabilities Enabling Biological Science
	11:00 AM – 11:30 AM	Allison Campbell (Pacific Northwest National Lab), EMSL Capabilities Enabling Biological Science
	11:30 AM – 12:00 PM	Adam Arkin (Lawrence Berkeley National Lab), The DOE Systems Biology Knowledgebase (KBase): Progress Towards a System for Collaborative and Reproducible Inference and Modeling of Biological Function
12:00 PM – 2:00 PM	**LUNCH ON YOUR OWN**	

TUESDAY, FEBRUARY 11TH

2:00 PM – 5:00 PM

Breakout Session D: Biological Systems for the Production of Biofuels
Location: Washington AB
Moderator: John Houghton

Description of Session:
Although biofuel production has been demonstrated, developing an advanced cellulosic biofuel sector will require transformational science to improve current production processes. The presentations in this session will describe results of recently supported research on a variety of topics relating to the biological production of biofuels. These topics vary from understanding how to improve biomass characteristics to the generation of biofuels by fungi and bacteria.

Speakers:

2:00 PM – 2:30 PM	Brian Davison (Oak Ridge National Lab), Feedstock to Fermenter: How Microbes Respond to Variant Biofeedstocks
2:30 PM – 3:00 PM	Daniel Amador–Noguez (University of Wisconsin), Metabolomics in Clostridial Biofuel Producers
3:00 PM – 3:30 PM	Hans Blaschek (University of Illinois), Understanding Fundamental Aspects of Butanol Production by *Clostridium beijerinckii*
3:30 PM – 4:00 PM	**BREAK**
4:00 PM – 4:30 PM	Audrey Gasch (University of Wisconsin), Comparative Fungal Genomics for Improved Biofuel Production
4:30 PM – 5:00 PM	Sam Hazen (University of Massachusetts Amherst), Systems Level Regulation of Rhythmic Growth Rate and Biomass Accumulation in Grasses

2:00 PM – 5:00 PM

Breakout Session E: Soil Microbes and Microbial Communities: Interactions and Role(s) in Sustainability
Location: Ballroom E
Moderator: Cathy Ronning

Description of Session:
Recent advances in systems biology and genomics technologies are beginning to map the complex networks underlying molecular mechanisms of plant and microbial growth, development, and metabolism. However, applying these tools to understanding the impacts of a fully sustainable, bioenergy agriculture on whole ecosystems remains a challenge, further complicated by climate variability and change. The ability to predict plant and microbial species' responses to a changing environment will be critical to understanding potential environmental impacts as well as for optimizing feedstock production. In this session, we will focus on the role(s) of microbes and microbial communities in the development of sustainable bioenergy agriculture systems, delivery of ecosystem services, and effective stewardship of the environment.

Speakers:

2:00 PM – 2:10 PM	Introductory Comments

TUESDAY, FEBRUARY 11TH

2:10 PM – 2:40 PM	David Weston (Oak Ridge National Lab), Predicting Plant Function from Endophyte Community Assembly and Member Abundance
2:40 PM – 3:10 PM	Mary Firestone (University of California, Berkeley), Root-Microbial C and N Interactions
3:10 PM – 3:40 PM	Kirsten Hofmockel (Iowa State University), Composition and Distribution of Core Carbohydrate-Active Microbial Genes in Biofuel Soils
3:40 PM – 4:10 PM	**BREAK**
4:10 PM – 4:40 PM	Philip Robertson (Michigan State University), Microbial Contributions to Greenhouse Gas Balances in Biofuel Landscapes: A Biogeochemical Perspective
4:40 PM – 5:00 PM	Roundtable Discussion

2:00 PM – 5:00 PM

Breakout Session F: An Introduction to Using KBase to Analyze and Model Microbes, Plants and Their Communities
Location: Ballroom F
Moderator: Pablo Rabinowicz

Description of Session:
The KBase project aims to provide the computational capabilities needed to predict and ultimately design biological function. KBase enables users to collaboratively integrate the array of heterogeneous datasets, analysis tools and workflows needed to achieve a predictive understanding of biological systems. It incorporates functional genomic and metagenomic data for thousands of organisms, and diverse tools including (meta)genomic assembly, annotation, network inference and modeling, thereby allowing researchers to combine diverse lines of evidence to create increasingly accurate models of the physiology and community dynamics of microbes and plants. KBase will soon allow models to be compared to observations and dynamically revised. A new prototype Narrative interface lets users create a reproducible record of the data, computational steps and thought process leading from hypothesis to result in the form of interactive publications. This workshop will describe some of the scientific analysis capabilities of KBase and introduce how to access them via the Narrative interface.

Speakers:

2:00 PM – 3:00 PM	Paramvir Dehal (LBNL) and Christopher Henry (ANL), Introduction to the KBase Narrative with a Microbial Use Case
3:00 PM – 3:15 PM	Q&A plus Community Feedback
3:15 PM – 3:30 PM	**BREAK**
3:30 PM – 4:00 PM	Folker Meyer (ANL) and Dylan Chivian (LBNL), Comparison of Microbial Communities and Creating Metabolic Models for Microbial Communities
4:00 PM – 4:15 PM	Q&A plus Community Feedback

TUESDAY, FEBRUARY 11TH

	4:15 PM – 4:45 PM	Doreen Ware (CSHL), David Weston (ORNL) and Priya Ranjan (ORNL), From Sequence to Consequence: A Demonstration of KBase Resources to Identify Plant Trait to Gene Relationships
	4:45 PM – 5:00 PM	Q&A plus Community Feedback

5:00 PM – 7:00 PM **Poster Session (even-numbered posters)**
Location: Independence Center

WEDNESDAY, FEBRUARY 12TH

8:00 AM – 9:00 AM **BREAKFAST ON YOUR OWN**

9:00 AM – 9:30 AM **DOE Report Update**
Location: Ballroom E/F
"Research for Sustainable Bioenergy: Report from the October 2013 Workshop"
Cathy Ronning, DOE–BER

9:30 AM – 11:30 AM **Plenary Session: DOE Early Career Research Program**
Location: Ballroom E/F
Moderator: Pablo Rabinowicz

Speakers:

9:30 AM – 10:00 AM	Michelle O'Malley (University of California, Santa Barbara), Engineering Anaerobic Gut Fungi for Lignocellulose Breakdown
10:00 AM – 10:30 AM	Heather Coleman (Syracuse University), Extreme Expression of Cellulases in *Populus*
10:30 AM – 11:00 AM	Brian Pfleger (University of Wisconsin, Madison), Application of Next–Generation Sequencing to Engineering mRNA Turnover in Cyanobacteria
11:00 AM – 11:30 AM	Dominique Loque (Lawrence Berkeley National Lab), Developing Synthetic Biology Tools to Improve Nutrient Acquisition of Energy Crops

11:30 AM **CLOSE–OUT AND ADJOURNMENT**

Table of Contents

1. Experimental Systems to Model Nutrient and Carbon Exchange in Plant-Microbe Symbiosis

Peter Larsen* (plarsen@anl.gov)[1], Shalaka Desai[1], Geetika Trivedi[2], Avinash Sreedasyam[2], Leland Cseke[2], and **Frank Collart**[1]

[1]Biosciences Division, Argonne National Laboratory, Lemont, IL
[2]Department of Biological Sciences, University of Alabama in Huntsville, Huntsville, AL

http://www.bio.anl.gov/molecular_and_systems_biology/protein_expression/

Project Goals: The Argonne Environment Sensing and Response (ESR-SFA) program addresses the hypothesis that cellular behavior can be modeled through an understanding of the biological interface with the environment and the cellular responses that originate from the cell/environment interaction. The overall objective of the ESR-SFA program is to identify the molecular basis of cellular transport and sensory pathways that mediate the response to environmental nutrients. This study describes experimental model systems comprised of Aspen, Laccaria bicolor, and Pseudomonas species will be used to minimize system complexity that permit manipulation of nutrient parameters under controlled conditions. All three components of this model system have completed genome sequences and are able to be cultured as monocultures, co-cultures, and tertiary or community-based cultures, thus enabling characterization of molecular interactions within and between the individual systems.

The FY13-15 ESR-SFA science plan uses a experimental model system comprised of Aspen, *Laccaria bicolor*, and *Pseudomonas* sp. The initial experiments proposed examined growth parameters using four selected Pseudomonas fluorescens strains (Pf-5, Pf0-1, SBW25, and WH6). These bacterial strains were isolated from the environment and previous studies indicate these are Plant Growth Promoting Bacteria (PGPB) that occupy the rhizosphere of many plant species and may improve ecosystem productivity. A three-component system is a complex experimental model but enhances the relevance for extrapolation of characterized molecular responses to organism interactions in natural ecosystems. The first objective was implementation of systematic experiments to characterize various experimental systems at a pilot scale to provide a foundation for the design and analysis of large-scale community experiments. In addition to characterizing the experimental models and community properties, the experiments were designed to validate methods for efficient extraction of community RNA and to measure organism/community parameters.

In one approach, we compared Phytagel and expanded perlite growth matrices for 3-component co-cultures. The Phytagel box system is well characterized for plant growth in Magenta boxes and provides reproducible, controlled results. Introduction of each bacterial strain is efficient in this system; however, access to the roots is limited by the density of the Phytagel. A vertical plate 3-component configuration allows clear visualization of the root systems as they interact with both fungi and PGPB. This system greatly improves recovery of plant root samples as they can be peeled from the surface of the gel; however, the plants within this vertical system do not develop as much biomass as either the Phytagel or perlite systems, indicating reduced overall plant health. The perlite system provides the best overall results. While plants need to be removed from the perlite to be visualize root development plant, fungal and bacterial growth and health are optimal, most likely due to the increased root aeration. Roots in particular exhibit a dramatic increase in biomass when grown in perlite; biomass is further increased when colonized by either *Laccaria* alone or *Laccaria* in combination with PGPB strains Pfl01 or PF-5. Such enhancements are not as apparent in the other two systems.

A parallel series of experiments determined the impact of *Pseudomonas fluorescens* PGPB strains on Aspen growth in phosphorus- or nitrogen-limited media. Control and nutrient-limited seedlings were grown with and without bacterial inoculation on vertical petri-plates in controlled environmental conditions. We examined the impact of several *P. fluorescens* strains (WH6, SBW25, Pf0-1, and PF-5) on the *in vitro* growth of Aspen seedlings at control and limited concentrations of nitrogen (4mM versus1mM) and phosphorous (1.5mM versus 20µM). Growth rate and morphology were assessed as primary indicators of plant health and additional parameters such as number of leaves, plant shoot height, root length and root structure such as branching pattern and rootlets, were also recorded to allow for more detailed evaluation of PGPB impact.

Nutrient limitation of both nitrogen and phosphorus was observed to decrease the number of leaves, but increase root length and branching structure. The number of seedling rootlets decreased significantly during phosphorus stress, but increased during nitrogen stress. All plants in nutrient-limited media showed and improvement in growth metrics when inoculated with PGPB relative to uninoculated controls under nutrient stress. The specific effects of PGPB colonization on plant growth and root morphology were found to be strain dependent and included alteration of plant root morphology, total root biomass, and aboveground biomass. Root length, for example, increased in control plants colonized with PF-5 and Pf0-1, while plants colonized with SBW25 exhibited enhanced root branching relative to the controls. An increase in the number of leaves was observed in control seedlings colonized with all bacterial strains. Similar strain-dependent effects were observed under both nutrient limiting and nutrient replete conditions. These preliminary results demonstrate some of the profound effects of PGPB on plant growth and morphology. Further understanding of nutrient limitation effects is expected to increase our insight into community structure and carbon cycling in terrestrial ecosystems.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357.Brookhaven National Laboratory is operated under Contract No. DE-AC02-98CH10886.This contribution originates from the "Environment Sensing and Response" Scientific Focus Area (SFA) program at Argonne National Laboratory. 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).

2. Functional Attributes of ABC Transporter Complexes from *P. fluorescens* That Mediate the Cellular Response to Changes in Environmental Nutrients

Eliza Zielazinski[1], Sarah Zerbs[1], Lisa M. Miller[2], Vito Graziano[2], Peter Larsen[1], Phil Laible[1], and **Frank Collart**[1]*(fcollart@anl.gov)

[1]Biosciences Division, Argonne National Laboratory, Lemont, IL
[2]National Synchrotron Light Source, Brookhaven National Laboratory, Upton, NY

http://www.bio.anl.gov/molecular_and_systems_biology/protein_expression/

Project Goals: The Argonne Environment Sensing and Response (ESR-SFA) program addresses the hypothesis that cellular behavior can be modeled through an understanding of the biological interface with the environment and the cellular responses that originate from the cell/environment interaction. The overall objective is to identify the molecular basis of cellular transport and sensory pathways that mediate the response to environmental nutrients. This study provides an overview of molecular approaches used to assign functional attributes to sparsely annotated transporter and sensor proteins with sufficient detail to predict the cellular response to specific environmental ligands. Experimental characterization of microbial transport and sensor proteins will facilitate the development of system-level predictive models for cellular response to nutrient availability. The ultimate goal of this work is to enable synthetic biology applications that will allow the manipulation of a system's response in a predictable way in order to maximize bioenergy production while minimizing negative environmental impact.

The Functional Characterization subgroup of the ESR-SFA applies a suite of biochemical and biophysical methods for protein production and functional annotation of transport machinery and regulatory networks. The experiments focused initially on characterization of ABC import complexes for sulfur compounds in four Plant Growth Promoting Bacterial (PGPB) strains of *Pseudomonas*. This pilot project focuses on methionine as the availability of this compound impacts cell metabolism via three major routes: as a sulfur source, its use in protein translation initiation and synthesis, and synthesis of the methyl group donor S-adenosylmethionine. Multiple ABC transporter complexes implicated in the import of methionine are encoded in the genomes of the four target strains: PF-5, Pfl01, SBW25, and WH6 strains.

Primary objectives of this subgroup are the profiling the ligand binding capabilities, mapping the interaction domains/peptides of the component proteins, and characterization of structure function relationships using biophysical methods. To achieve these objectives we have devised soluble and membrane protein expression strategies for the solute binding protein (SBP), permease, and ATPase component proteins of the ABC transporter complex. The majority of these components, comprising ~40 individual genes, have been successfully expressed and purified either individually or as complexes and functional characterization studies are ongoing. Profiling of the ligand binding specificity of the individual SBPs demonstrate these organisms can import a variety of sulfur-containing compounds with some organisms having multiple bacterial transporters with predicted specificity for the same ligand. This set of methionine importers also contains ATPase components with regulatory domains identified using the Pfam database and many of these ATPase proteins have been expressed in soluble form. An important question we will undertake is the examination the ligand binding specificity of these ATPase regulatory domains as it is not known if the ligand-binding capacity changes with specificity identical to that of the solute-binding protein.

The set of sulfur compound transporters from *P. fluorescens* PGPB strains was additionally used as a model ABC transport system for the production of intact, transmembrane transporter assemblies for functional characterization experiments. Here, success requires co-expression of permease and ATPase components. Multiple approaches were successful in producing purifiable quantities of the membrane-localized complex that contained affinity tag fusions for use in chromatography, cellular localization, and quantitation. Conditions that fostered efficient purification of permease-ATPase complexes utilizing affinity-tag fusions were discovered and optimized using the automated Maxwell purification system (Promega). Standardized protocols have been developed to produce samples of intact transporters with > 90% purity from whole cells in less than 4 hours. Expression yields approach 0.5 mg/L of cell culture and have been shown to be scalable. The most unexpected finding discovered in the course of methods development for transmembrane systems is a potential for cross-talk between components derived from different operons within the same organism. We used co-purification to demonstrate the feasibility for generation of hybrid methionine transporter assemblies from multiple ATPase components and an affinity-tagged permease. Because they are regulated differently, it is unclear if these heterologously- expressed *in vivo* findings from altered regulatory scenarios have any relevance to structures that are formed and operating in *P. fluorescens* cells, but we intend to examine fully the extent of these promiscuous interactions and to characterize how prominent their presence and utility is in Nature.

As many of the proteins involved are membrane-bound, or have membrane-bound components in multi-subunit complexes we evaluated the utility of SAXS/WAXS which can be performed on membrane proteins in their native lipidic environment, such as micelles, bilayers, and nanolipoprotein particles (nanodiscs). This model independent low resolution structural analysis by SAXS demonstrates the first successful step for determining membrane protein structures in nanodiscs. Current studies are focused on development of the data analysis methodology. Most *ab initio* SAXS analysis algorithms are based on proteins in buffer solution and do not consider the different electron density distribution in the phospholipid bilayer. This presents a major challenge in that existing programs cannot be extended to model nanodiscs. The problem is further complicated by the presence of embedded guest proteins in the nanodisc. As a consequence, we are developing custom software for modeling and fitting routines.

Specifically, we are utilizing a course-grained modeling approach to model the different scattering bodies within the nanodiscs, i.e. proteins, lipid head groups, and lipid tails which all have different electron densities and scatter differently. To do so, we plan to assemble course-grained scattering bodies as two protein belts wrapped around a cylindrical phospholipid bilayer. Next, a number of lipid bilayers are removed from the center of the nanodiscs and the void is filled with protein electron density scattering bodies. Using molecular constrained ab initio shape determination methods, the scattering curves are calculated and the model is refined iteratively to fit the experimental data. The goal of this work is to calculate the low resolution structures of membrane proteins incorporated into nanodiscs, and to determine conformational changes upon binding ligands or other soluble proteins.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357.Brookhaven National Laboratory is operated under Contract No. DE-AC02-98CH10886.This contribution originates from the "Environment Sensing and Response" Scientific Focus Area (SFA) program at Argonne National Laboratory. 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).

3. Plant Microbe Interfaces: Proteomic Characterization of Endophyte and Rhizosphere Microorganisms and their Impacts on Plants

Gregory B. Hurst[1*] (hurstgb@ornl.gov), Zhou Li,[1,4] Olaf Czarnecki,[2] Karuna Chourey,[1] Jun Yang,[2] Chongle Pan,[3] Jin-Gui Chen,[2] Dale A. Pelletier,[2] Sara Jawdy,[2] David J. Weston,[2] Jennifer Morrell-Falvey,[2] Amber Bible,[2] Shen Lu,[2] Keiji G. Asano,[1] Jessy Labbé,[2] Tatiana Karpinets,[2] **Gerald A. Tuskan**,[2] and **Mitchel J. Doktycz**[2]

[1]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831; [3]Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831; [4]Graduate School of Genome Science and Technology, University of Tennessee-Oak Ridge National Laboratory, Knoxville, TN 37996

Project Goal: 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.

http://PMI.ornl.gov

Protein expression patterns, and changes in those patterns, provide information on metabolic and other strategies employed by bacteria, fungi, and plants that participate in natural communities such as the root microbiomes of *Populus* trees. Ongoing projects in the PMI SFA are studying the proteome responses of these various organisms to treatments ranging from application of hormones to interactions with other members of the *Populus* microbiome community.

Populus proteome response to a constructed soil microbial community. We have acquired leaf proteome data from *Populus* specimens inoculated with a 3-member bacterial community (see poster by Weston et al.), as well as from axenic control *Populus* specimens. The goal of this study is to determine whether observed phenotype differences are accompanied by changes in the *Populus* proteome. Preliminary analysis reveals that slightly less than half of the approximately 4400 detected protein groups are common to all samples, with ongoing work aimed at identifying proteins of altered abundance in the inoculated plants.

Proteomes of mycorrhizal roots in *Populus*. We have performed proteomics characterization of *Populus* roots that exhibit varying degrees of colonization by the ectomycorrhizal fungus *Laccaria bicolor*. Identification of both fungal and plant proteins will provide biological insights into their interactions. To determine figures of merit for proteins that are potentially important in the colonization process, but may be difficult to measure because of their small size or transient expression, we have shown that one such protein (MiSSP7) is amenable to detection by our LC- MS-MS protocol as a pure standard, but may be masked by matrix effects when present in a full proteome.

Effects of plant hormones on protein expression. As a n ew class of plant hormones, strigolactones (SLs) act as a key inhibitor of shoot branching, stimulate seed germination of root parasitic plants, and promote hyphal branching and root colonization of symbiotic arbuscular mycorrhizal fungi. They also regulate many other aspects of plant growth and development. We have perfomed quantitative proteomics using an isobaric chemical labeling reagent, iTRAQ, to identify the proteome regulated by SLs in *Arabidopsis* seedlings. We found that in addition to regulating the abundance of proteins implicated in SL pathways, SLs also regulate the expression of a number of proteins that have not been previously assigned to SL pathways. Furthermore, we observed a drastic difference between the SL-regulated transcriptome and the SL-regulated proteome. These findings provide a new tool to investigate the molecular mechanism of action of SLs.

Proteomes of *Populus* endophyte and rhizosphere bacterial isolates. The increasing availability of genome sequences for PMI bacterial isolates (see poster by Pelletier et al.) has enabled proteomics investigations of selected members of this growing collection. Changes in protein expression resulting from an applied treatment (e.g., environmental stress, an applied chemical, or presence of another organism) can reveal metabolic pathways involved in the response to the treatment. Examples of studies of PMI isolates include responses of *Pseudomonas* sp. strain GM41 to tryptophan, and of *Pantoea* sp. strain YR343 to *Populus* root exudate. It will be interesting to consider detected proteomes across bacterial strains in light of comparative genomics studies being performed elsewhere in the PMI (see poster by Ussery et al.) To facilitate these measurements, we are refining our LC-MS-MS protocols to achieve higher throughput.

Future studies. We are planning metaproteomics measurements on natural rhizospheric communities of *Populus*, and integration with metagenomics analyses (See poster by Schadt *et al.*).

Current and new datasets are being made available across the PMI project using the Proteomics Workflow implemented in the PMI Knowledgebase (see poster by Ussery *et al*.).

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

4. Plant-Microbe Interfaces: PMI Data Consolidation: Estimating Data Quality in Large Datasets

Dave Ussery[1,2*] (usserydw@ornl.gov), Miriam Land,[1] Doug Hyatt,[2] Guruprasad H. Kora,[1,4] Se-Ran Jun,[1,5] Loren Hauser,[1,2,3] **Gerald A. Tuskan,[1]** and **Mitchel J. Doktycz[1]**

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge TN 37831; [2]Genome, Science and Technology Graduate Program, University of Tennessee, Knoxville TN 37996; [3]Computing & Computational Sciences, Oak Ridge National Laboratory, Oak Ridge TN 37831; [4]Department of Microbiology, University of Tennessee, Knoxville TN 37996; [5]Joint Institute for Computational Sciences, University of Tennessee, Knoxville, TN 37831

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.

Too much data: There has been an explosion of new methods for generating massive amounts of data. We are developing methods for evaluating data quality across large biological datasets (http://genomes.ornl.gov). As a first step, we have downloaded all public prokaryotic genome sequences from 6 different databases, with a total of more than 30,000 unique sequences. Most of the genomes are of draft quality (more than 80% of the current bacterial genome sequences). Within the Plant Microbe Interface (PMI) project, there is a need to know how reliable the genome sequences are for soil bacteria being used for comparisons. We examined the microbial DNA sequences available for complete, draft, and Short Read Archive genomes in GenBank as well as three other databases (Patric, KBase, and Broad) and assigned quality scores for more than 30,000 unique prokaryotic genome sequences. Scores were assigned using four categories (sequence quality, the presence of full length rRNA genes, tRNA composition (is there at least one tRNA for each of the 20 amino acids?) and the presence of a set of 120 conserved genes in all prokaryotes, and a combined score. Genomes with a quality score above a suitable threshold are being used to construct phylogenetic trees for the PMI project.

Community Structures: There are several major clades of bacteria associated with plants and soil samples. One commonly found environmental group of bacteria is the *Pseudomonas* genus. We have extracted all 258 *Pseudomonas* genomes from DOE KBase (in November, 2013). From these, two sets of trees were generated, one based on 16S rRNA, the other on ribosomal proteins. Twenty of these genomes were sequenced as part of the PMI project, and 19 formed a cluster with other *Pseudomonas fluorescens* related genomes, and one (*P. putida* related) was distantly related. A set of 47 genomes was selected to define the phylogenetic relationships of the 20 PMI genomes. Trees based on gene content and average amino acid identity (AAI) were constructed for these 47 genomes. The

19 PMI *P. fluorescens* related genomes cluster in a group of 24 genomes, and based on the AAI, there are 20 distinct species, with the 19 PMI species representing 17 distinct species. Within this cluster, there are a total of 22,000 different gene families, and a set of 2221 core gene families. For each of the 20 species, sets of species-specific gene families and their possible function have been identified.

Data Consolidation: One of the main priorities of PMI project has been consolidation of data resources across the project. The data consolidation initiative aims to help researchers collect data from multiple, disparate sources, and integrate into a single consolidated knowledgebase. As the project grows, demand for a single point of data access, data sharing, collaboration and provenance increases. PMI Knowledgebase now has an integrated workflow subsystem (http://pmi.ornl.gov) that enables PMI researchers to aggregate their project-generated data. The system provides tools and data-driven workflows that improves access to data resources, enhances data quality along with increased data protection. It has helped pipelining and to eliminate data duplication, thus providing a platform to consolidate data assets into a central repository resulting in efficient management and use of project-generated data.

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

5. Plant-Microbe Interfaces: Tripartite plant-fungal-bacterial interactions

Jessy Labbé,[1] Rytas Vilgalys[2], Gregory Bonito[2], Jessie Uehling[2], Claire Veneault-Fourrey[3], Aurélie Deveau[3*] (deveau@nancy.inra.fr) Cora Guennoc1,[3], Wellington Muchero[1], Le Tacon François[3],Olaf Czarnecki[1], Jay Chen[1], Jonathan Plett[4], Yohan Daguerre[3], Sebastian Wittulsky[3], Sarah J. Melton[1], Annegret Kohler[3], Jennifer Morell-Falvey[1], Annick Brun[3], Francis Martin[3], **Mitchel J. Doktycz[1]**, **Gerald A. Tuskan[1]**

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA; [2]Biology Department, Duke University, Durham, NC, USA; [3]INRA, Tree-Microbe Interactions, UMR1136, Champenoux, France; [4]Hawkesbury Institute for the Environment, University of Western Sydney, Richmond, Australia

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 fro 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 molecular events leading to recognition and colonization of a host plant by beneficial microorganisms are poorly understood. Our ongoing research is aimed at identifying and isolating microbes associated with natural *Populus* ecosystems in order to determine molecular, genetic and cellular events involved in recognition and establishment of beneficial microbial interactions with *Populus.* Several research investigations aimed at enhancing our understanding of plant-microbe interactions are underway, focused on elucidating the genetic and molecular mechanisms of the interactions of the host plant and the bacterial network associated with fungal partners within natural *Populus* ecosystems.

The mycorrhizal symbiosis is the most widespread plant-microbe association that supports forest growth and sustainability. Based on the *Populus-Laccaria* ectomycorrhizal model, this project is focused on identifying specific host-derived genetic determinants by characterizing a core set of genes regulating *Populus-Laccaria* interactions. To this end, we have combined quantitative trait loci analyses (QTL) of *Laccaria* colonization phenotypes in a *P. trichocarpa* x *P. deltoides* F_1 pedigree, transcripts level expression analyses under the same colonized condition, high- throughput genotyping and whole genome resequencing to characterize sequences in the target genomic regions. We have identified a species-specific whole-gene deletion in *P. deltoides*, which co-segregated with colonization efficiency. This gene was absent in 60 *P. deltoides* genotypes from diverse geographical origins but was highly conserved in 673 *P. trichocarpa* genotypes evaluated. Variants resulting in up to 2X more colonization by the fungal symbiont have been identified in *Populus*. Transgenic lines in *Arabidopsis* and in *Populus* have been constructed to investigate the functionality of a couple of target genes.

Plants have developed a complex defensive response system to protect themselves against invasion by detrimental organisms, often mediated by plant hormones. Invading organisms, in turn, have developed various methods to circumvent the plant's defenses or control plant cell function to their benefit. Pathogens attempt to manipulate the plant response by producing effectors that target different components of the JA and ET signaling pathways in such a fashion that colonization is favored. Similarly mutualistic fungi affect plant hormone signaling cascades to achieve colonization although knowledge of the mechanisms behind most of these differences is in its infancy. In this work, we demonstrate that MiSSP7, an effector protein produced by *L. bicolor*, targets plant-encoded JAZ proteins, in particular PtJAZ6, and interacts with it in the nucleus of the plant. Through this interaction, MiSSP7 is able to block the activity of MeJA and promote the proliferation of both bacteria and *L. bicolor* in plant tissues. This effect is likely due to the ability of MiSSP7 to stabilize the JAZ protein and reduce the jasmonic acid induced degradation of the JAZ protein. Thus, the jasmonic acid responsiveness of the host plant would be affected by the microbial environment to foster symbiotic interactions.

In complex soil ecosystems, fungi are surrounded by diverse microbial communities, which modulate the mycorrhizal symbiosis. These include the so-called mycorrhiza helper bacteria (MHB), which are thought to assist mycorrhiza formation and symbiosis. Moreover, some mycorrhizal and root-associated fungi possess bacterial endosymbionts. Because very little is known about the role of these helper and endosymbiotic bacteria in *Populus*-fungi interactions, this project is aimed at dissecting the signaling mechanisms underlying *Populus*–fungal–bacterial interactions. To this end, we have sequenced genomes of bacterial endosymbionts and helper bacteria, constructed mutant libraries and have begun mutant phenotype screening. We have cleared the endosymbiont from several fungal strains and observed fitness costs to the host under specific conditions. We demonstrate that some helper bacterial strains influence *Populus–L. bicolor* colonization and some mutants are affected in their beneficial effect. This study provides new insights into the mechanism of multi-partite interaction between *Populus* and its complex microbial communities.

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

6. Plant-Microbe Interfaces: Understanding the factors shaping microbial community structure within root and rhizosphere microbiomes of *Populus* species

M. S. Robeson, II,[1] M. Shakya,[1] Z. K. Yang,[1] M. Kerley,[1] M. Podar,[1] G. M. Bonito,[2] S. Tringe,[3] W. Muchero,[1] **G. A. Tuskan,**[1] R. Vilgalys,[4] and C. W. Schadt[1*] (schadtcw@ornl.gov)

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, [2]National Herbarium, Royal Botanic Gardens, Melbourne, Victoria, 3141, Australia, [3]Joint Genome Institute, Walnut Creek, CA. [4]Biology Department, Duke University, Durham, NC.

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 spp. (Poplar and Cottonwood) are a genetically diverse genus of tree species that are broadly adapted to the temperate environments of North America and typical of important riparian habitats, making them ideal as ecological model species. Additionally, their fast growth rates and the ease with which *Populus* clones and hybrids can be grown on land otherwise not suitable for food production, make them good candidates for production bioenergy research. For these reasons, it is important to understand the role of the root microbiome as it relates to the health and productivity of *Populus*. We have been examining the root, rhizosphere and soil communities of *Populus deltoides* in natural riparian habitats in the Eastern US, as well as controlled common garden populations of *P. trichocarpa* in the Western US, using a Roche-454 and Illumina-MiSeq analysis of rRNA amplicons respectively. Microbiome data for both bacteria and fungi are analyzed against the corresponding bulk soil properties, tree phenotype, and tree genotype data in order to understand how such properties influence microbiome structure.

In our studies of *P. deltoides* we have shown that the rhizo- and endosphere environments feature highly developed, diverse and to a large degree often exclusive communities of bacteria and fungi. Endophytic bacterial diversity is found to be highly variable, but on average tenfold lower than the rhizosphere, suggesting root tissues provide a distinct environment supporting relatively few species more heavily dominated by Actinobacteria and γ-Proteobacteria when compared with the rhizosphere. *Populus* spp. especially appear to be highly enriched for *Pseudomonas fluorescens*-like species/OTUs when compared to other *Populus* habitats as well as endophytic habitats of surrounding (non-*Populus*) tree species. Fungal endophytic species are more numerous than bacteria, but also less than rhizosphere spp. Both fungal and bacterial rhizosphere samples showed distinct phylogenetic composition patterns compared to the more variable endophyte samples. Contrary to initial expectations, both *Populus* spp. have low natural levels of colonization by ectomycorrhizal (ECM) and arbuscular

mycorrhizal fungi, but high levels of presumed fungal endophytic taxa such as *Nectria*, *Mortierrella*, and members of the *Atractiellales*.

Overall, the *P. trichocarpa* rhizosphere communities studied in the Western US separate based on the two Oregon common gardens in which they were sampled: Clatskanie and Corvallis. Conversely, the endophyte communities between the common gardens were similar. Pseudomonad OTUs dominated both the rhizosphere and endophyte samples while being nearly undetectable within bulk soil. Finally, the effects of *P. trichocarpa* genotype on the composition of its root microbiome appear to be limited compared to the effects of local soil environment. These results are similar to what we have observed for natural populations in the Eastern US for *P. deltoides*.

Future research plans will 1) expand both the number and geographic range of species of *Populus* examined, 2) move beyond the rooting zone to total microbiome studies of *Populus*, as well as 3) move towards additional functional examinations enabled by developments in metagenomic analyses.

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

7. Plant Microbe Interfaces: Defining the functional diversity of the *Populus* root microbiome

Dale A. Pelletier* (pelletierda@ornl.gov), W. Nathan Cude, Michael S. Robeson II, Zamin K. Yang, Steve L. Allman, Mircea Podar, Tse-Yuan S. Lu, Christopher W. Schadt, David W. Ussery, **Gerald A. Tuskan** and **Mitchel J. Doktycz**

Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

http://PMI.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community 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 fro 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 beneficial association between plants and microbes exemplifies a complex, multi- organism system that is shaped by the participating organisms and the environmental forces acting upon it. Studying the integral plant–microbe system in native, perennial environments provides a great opportunity for discovering plant–microbial system functions relevant to DOE missions related to bioenergy and carbon-cycle research and understanding of ecosystem processes. Therefore, *Populus* and its associated microbial community are being studied as part of our Plant-Microbe Interfaces project (http://pmi.ornl.gov). *Populus* trees are host to a variety of microorganisms within their endosphere and rhizosphere that can influence host phenotypes. Our goal is to understand the phylogenetic and functional diversity within the *Populus* microbiome and to elucidate the metabolic and molecular mechanisms responsible for shaping the *Populus*-microbial interface. To begin to untangle this complex ecosystem, we have applied cultivation dependent and cultivation independent techniques to capture and characterize the *Populus* root microbiome. Utilizing direct plating methods we have isolated and begun to characterize a large collection of *Populus* rhizosphere and endosphere bacterial strains. Through a JGI-CSP project, we are currently sequencing the genomes of many of these isolates and applying comparative genomics to identify functions important for the formation of mutualistic relationships with the host. However, our isolate collection may not be representative of the *Populus* microbiome. Population distributions from rRNA gene-based approaches on *Populus* suggest that the high-GC Gram-positives, *Planctomycetales*, *TM7*, *Crenarchaea*, and *Acidobacteria*, may be underrepresented in culture-based efforts. Further, because of the low microbial biomass relative to plant biomass, metagenomes have been difficult to obtain. Finally, certain endophytic groups have been difficult to isolate and culture in laboratory settings. To address these issues, a method of enriching live endophytes from *Populus* root homogenates for isolation and metagenomic investigation is being developed. Endophytic bacterial communities were enriched using differential and density gradient centrifugation. Total DNA was extracted from enriched and unenriched samples, and the endophytic bacterial community composition was determined by 16S rDNA sequencing, using the MiSeq platform. Our enrichment protocol reduced

the number of contaminating chloroplast DNA reads by approximately 10 fold. The enrichment also significantly increased the reads of Actinobacteria, Planctomycetia, and Alpha- and Gamma-proteobacteria classes. Live bacterial enrichments were also inoculated to agar plates for isolation and flow sorting for single-cell genomics. The ability to perform single-cell multi-omic analyses will allow for in-depth characterization of rare endophytic bacteria residing within the natural root system of *Populus.*

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

8. Plant-Microbe Interfaces: Probing the Molecular Mechanisms of Plant-Microbe Interactions

Amber Bible,[1] Yasu Oda,[2] Dale Pelletier,[1] Amy Schaefer,[2] Jennifer Morrell-Falvey,[1] Carrie Harwood, [2] and **Mitchel Doktycz**[1*] (doktyczmj@ornl.gov)

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Department of Microbiology, University of Washington, Seattle, WA

http://PMI.ornl.gov

Project Goals: We are focused on understanding *Populus*–microbe interactions at the molecular level to dissect the signals and pathways important for initiating and maintaining symbiotic relationships with *Populus*. Our goal is to elucidate the molecular, spatial, and temporal dynamics involved in *Populus*-microbe interactions using systems biology approaches and directed analytical methodologies. We are interested in how bacteria selectively respond and become associated with *Populus*; and how microbially induced molecular and cellular events impact plant growth, health, and fitness. Ultimately, these data will be used to construct model plant-microbial communities to better understand the underlying rules to community assembly and the functional contributions that result from arrangements of multiple organisms.

Our current research is focused on dissecting the signaling pathways involved in plant-microbe interactions using select Populus-derived isolates that were chosen based on phenotypic screens and genomic inventory data. One area of focus in these isolates is cyclic-di-GMP signaling, which often controls exopolysaccharide production, motility, and other colonization factors. In the robust root colonizer *Pantoea* YR343, we have employed promoter libraries and live imaging to identify three c-di-GMP signaling genes, *orf2884, orf3006,* and *orf3134,* which are highly expressed during *Populus* root colonization but not in laboratory culture. These genes encode diguanylate cylases, which are enzymes that synthesize c-di-GMP, that were more highly induced by growth on *Populus* and/or wheat roots. We have engineered strains to overproduce c- di-GMP in *Pantoea* YR343 and the *Pseudomonas* strains GM17 (an inhibitor of *Laccaria* fungal growth), GM41 (a helper of *Laccaria* fungal growth), and GM30 (an inducer of plant root proliferation). Ongoing RNAseq analyses of the c-di-GMP-overexpressing strains aim to elucidate putative c-di-GMP-controlled exopolysaccharide genes believed to play a role during root colonization. A second signaling system of interest is that of acyl-homoserine lactone signal quorum sensing (QS). Our recent work has shown the genes encoding QS signal synthases (*luxI* genes) and QS signal receptors (*luxR* genes) are prevalent in members of the *Populus* microbiome. We also observed many examples of a recently described subfamily of orphan *luxR*-genes encoded in the genomes of *Rhizobium* and *Pseudomonas* strains isolated from *Populus*. This LuxR subfamily is unusual in that it is believed to respond to an unknown plant- derived signal, not a bacterially produced acyl-homoserine lactone signal. We have created reporter fusions in order to follow LuxR activity in a *Populus* isolate, *Pseudomonas* GM79, and found that plant macerate is required for LuxR activity. These reporter fusions now enable experiments aimed at the elucidation of the plant compounds that serve as a LuxR ligand.

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

9. Plant-Microbe Interfaces: Discovery of Small Secreted Proteins in *Populus* in Response to Symbiotic Fungus *Laccaria bicolor*

Xiaohan Yang[1*] (yangx@ornl.gov), Hengfu Yin,[1] Jonathan Plett,[2] Ting Li,[1] Rick Chen,[1] Annegret Kohler,[3] Emilie Tisserant,[3] Gregory B. Hurst,[4] Jessy Labbé,[1] Wellington Muchero,[1] Priya Ranjan,[1] Sara Jawdy,[1] Timothy J. Tschaplinski,[1] Claire Veneault-Fourrey,[3] Francis Martin,[3] **Mitchel John Doktycz,**[1] and **Gerald A. Tuskan**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Hawkesbury Institute for the Environment, University of Western Sydney, Richmond, Australia; [3]Ecogenomics of Interactions Lab, UMR 'Interactions Arbres/Micro-Organismes',INRA-Nancy, France; [4]Chemical Sciences Division, Oak Ridge National Laboratory,
Oak Ridge, TN

http://PMI.ornl.gov

Project Goals: The goal of the Plant-Microbe Interfaces SFA is to gain a deeper understanding of the diversity and functioning of mutually beneficial interactions between plants and microbes in the rhizosphere. The plant-microbe interface is the boundary across which a plant senses, interacts with, and may alter its associated biotic and abiotic environments. Understanding the exchange of energy, information, and materials across the plant-microbe interface at diverse spatial and temporal scales is our ultimate objective. Our ongoing efforts focus on characterizing and interpreting such interfaces using systems comprising plants and microbes, in particular the poplar tree (*Populus*) and its microbial community in the context of favorable plant microbe interactions. We seek to define the relationships among these organisms in natural settings, dissect the molecular signals and gene- level responses of the organisms using natural and model systems, and interpret this information using advanced computational tools. PMI research is multidisciplinary by design and multi-institutional in composition.

Ectomycorrhizal (ECM) symbiosis is a mutualistic association between the roots of many plant species and fungal partners found in soil. Mycorrhizal symbiosis offers various benefits including 1) increasing nutrient availability, 2) improving water use efficiency, 3) enhancing carbon sequestration in terrestrial ecosystems and 4) remediating degraded soils. All of these beneficial aspects make tree-mycorrhizal association an excellent strategy for improving the sustainability of woody crop production. *Populus* is an important woody crop that has been developed for pulp and paper manufacturing, phytoremediation, carbon sequestration, and biofuels production. The soil fungus *Laccaria bicolor* is able to form symbiotic associations with many temperate forest trees including *Populus*. The reference genome sequences along with rich genetic and genomic resources are available for both *Populus* and *Laccaria*. Therefore, the *Populus-Laccaria* interaction is an excellent model system for studying mycorrhizal symbiosis. Understanding the molecular mechanisms underlying the *Populus-Laccaria* interaction would provide potential solutions to protecting and maximizing the value of forest ecosystems, which may lead to novel breeding targets, new sustainable silviculture strategies and better utilization of woody tree species in both industrial and ecological settings.

Small proteins in plants play important regulatory roles in various biological processes such as stress response, flowering, and hormone signaling. However, our knowledge about plant small proteins in relation to mycorrhizal symbiosis is very limited. To address this limitation, we performed genome-wide analysis of *Populus* small proteins, with a focus on small secreted proteins, in response to *L. bicolor* inoculation. Based on RNA- seq data analysis, we identified 1,242 computationally predicted

Populus small proteins (SmPs) that were up-/down-regulated in response to *Laccaria* inoculation. In the small protein set, 417 proteins (33.6%) were predicted to be small secreted proteins (SSP). Gene ontology analysis revealed that some SmPs were involved in biological processes relevant to plant-microbe interactions, such as response to fungus, jasmonic acid metabolism, salicylic acid-mediated signaling pathway, and cell-to-cell communication.

P. trichocarpa is much more heavily colonized by *L. bicolor* than *P. deltoides*. Analysis of *Populus* genome resequencing data revealed several *Populus* SmP genes present in *P. trichocarpa* genotypes ("93-968" and "Nisqually-1") but absent in *P. deltoides* genotypes ("D124" and "ILL-101"), suggesting that these SmP genes contribute to the genotypic difference in *Populus-Laccaria* interaction. Importantly, we tested the 39 predicted SSPs using yeast trap system and 33% of them were confirmed to be secretory proteins. The high-confidence SSP candidate genes are being characterized using molecular and genetic approaches. This research generated new knowledge about the molecular basis of *Populus-Laccaria* 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.

10. Plant-Microbe Interfaces: Bacterial Community Effects on Host Plant Biomass Allocation through Experimentation and Modeling

Dave Weston[1*] (westondj@ornl.gov), Jeremiah Henning,[1,2] Collin Timm,[1] Nathan Cude,[1] **Gerald Tuskan,**[1] Dale Pelletier[1], and **Mitchel Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Genome Science and Technology Program, University of Tennessee, Knoxville TN

http://PMI.ornl.gov

Project Goals: The Plant-Microbe Interfaces project's goals include understanding how bacterial strains and communities impact the growth of the host plant through direct interaction or modification of the environment. In this project we investigate the effects of specific bacterial isolates from the *Populus* endosphere on host plant traits. Our specific goal is to use genomic information and experimental observations to predict the effect of a defined microbial community on plant biomass generation and allocation.

Diverse microbial species and complex interactions make the prediction of function of a microbial community an arduous task. This is especially relevant in plant microbial communities, where hyper-diverse species pools exist in the rhizosphere: the immediate soil spaces around a plant root. Microbes in the rhizosphere can affect soil conditions by fixing atmospheric nitrogen, solubilizing inorganic phosphorous, and secreting hormones or other factors that directly influence the plant host. In turn, the plant host provides carbon sources and can harbor some of these microbes as endophytes. Deterministic factors that may influence endophytic colonization include: plant selection (avoiding innate immune response), host genotype, microbial competition, microbial niche differences, and environmental differences (pH, soil moisture, etc.) In natural ecosystems, it is exceedingly difficult to tease apart all interaction processes, which are undoubtedly occurring simultaneously. To address this complex network of interactions, we use a bottom-up approach to understand microbe-microbe, microbe- host, and microbe-environment relationships that influence collective community function. By choosing bacterial isolates with known host-effects and known interactions with complementary isolates, we reassemble simplified communities and predict community function using quantitative models.

Growth of *Populus deltoides* with single microbial strain inoculations demonstrate that microbial strains can have variable impacts on plant biomass allocation (root:shoot ratio), leaf number, root area, and stem length. But when grown with a three microbe consortia, we find a strong dominance hierarchy between microbes, which suggests niche separation or competition or plant selection. Surprisingly, the most dominant microbe in our mixes (*Burkholderia sp. BT03*, 97- 99% relative abundance), which displayed the most unique plant phenotype when grown alone with plant host (high root:shoot ratio, decrease stem length), contributed very little to plant phenotype in the mixed community. Thus, preliminary results show microbe-microbe interactions may be important in structuring endophytic communities, however community function may be driven by bacterial strains that exist at low abundances. Ongoing measurements of plant gene expression in response to microbial communities will help identify mechanisms of interaction.

Plant biomass generation is primarily limited by availability of nitrogen or phosphorous in the environment. Nitrogen and phosphorous availability can be increased by the presence of nitrogen fixing or phosphate solubilizing bacteria in the soil and/or endosphere. We are developing metabolic

models of nitrogen fixing endophytes isolated from *Populus deltoides* to examine how environmental conditions and the microbial genotype affects rate of atmospheric N_2 fixation when associated with the plant. Similarly, the amount of phosphorous solubilizing bacteria impacts the availability of phosphorous to the plant and other microbes in the system. Due to the difficulty in measuring these rates directly in root systems, we are generating quantitative models of fixation and solubilization rates in monoculture. The models are then combined using measured or predicted biomass fractions to simulate the community function in the natural environment. Ultimately, these models can be used to design microbial community structure and function to optimize plant biomass generation rate.

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

11. Metagenomic Insight into the Rhizospheres of Three Biofuel Crops

James R. Cole[1*] (colej@msu.edu), Aaron Garoutte[2], Jiarong Guo[1], Jordan Fish[1], Adina Howe[1,4], Bangzhou Zhang[1], Chao Xue[1], Qiong Wang[1], Caio Rachid[1], Elaine Gomes[1], C. Titus Brown[2,3] and **James M. Tiedje**[1,2]

[1]Center for Microbial Ecology, [2]Microbiology and Molecular Genetics, [3]Computer Science and Engineering, Michigan State University, East Lansing, MI 48824; [4]Argonne National Laboratory, Lemont, IL 60439

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.

An important issue in producing biofuel feedstock is long-term environmental sustainability. In previous years, we developed approaches to explore the high diversity present in complex communities of agricultural soils, with an emphasis on genes involved in nitrogen availability and cycling. In Phase II of this project, we apply approaches to identify critical biogeochemical cycling genes and populations in a rigorous sampling of multiple localized rhizosphere soil communities (n=7) from three major bioenergy crops: switchgrass, Miscanthus, and corn (continuous), from GLBRC's Kellogg Biological Station (KBS) intensive sites. This depth of replicated sampling significantly extends the statistical power of previous studies to identify relevant drivers of beneficial plant-microbe interactions and nitrogen cycling genes.

Each of all 21 samples was shotgun sequenced as one lane in Illumina HiSeq by JGI and produced high yield and good quality sequence reads. To mine this large volume of data for gene information, we developed a set of data mining processes/techniques, including our novel digital normalization (diginorm) and partitioning method for big data assembly, SSU rRNA gene fragment finding and analysis method, our Xander tool for gene targeted assembly from shotgun data, and N-cycle gene finding tools from shotgun data. Together, these tools are being used to reconstruct representative genetic references for bioenergy crop soil microbial communities in the rhizosphere.

Overall, the taxonomic composition of rhizosphere samples were consistent between using our SSU rRNA fragment finding method and our diginorm/partitioning assembly pipeline. They both implicated corn as the most different among the three microbiomes. Compared to the two perennial grasses, switchgrass and Miscanthus, corn-samples were more abundant in Proteobacteria but less in Acidobacteria. Proteobacteria are usually fast-growing and Acidobacteria are slow growers, consistent with corn growing new roots every year providing more new C for selection. The OTU based diversity analysis also suggests that corn-associated communities are less rich and diverse and significantly different from switchgrass and Miscanthus microbiomes.

Soil metagenomes were annotated against known nitrogen genes in the MG-RAST database. Novel nitrogen genes were also assembled using our own tool (Xander). Annotations of 10M read subsets of the short read replicates by MG-RAST showed that the relative abundances of N- cycle genes of the three crops are significantly different and, again, corn has the least abundance of the three. We found our Xander gene-targeted assembler provides higher sensitivity in determining contents of specific genes. We used Xander on all seven Miscanthus replicates (a total 1.6 billion reads or 100 billion

unique 30-mers after paired-end assembly and quality filtering) and found 60 unique contigs, and that the majority of contigs had a mean coverage of less than three, demonstrating that this technique can assemble genes with low coverage (and missed by any global assembly method). These contigs showed close relatedness(dissimilarity < 10%) to their closest *nifH* reference sequences, with the top two groups similar to family Rhodospirillales (69.6%), which includes Azospirillum, while another 15.7% were most similar to the Rhizobiales. Xander was also used to assemble 995 contigs of *nirK* gene from 50G sequences and examined the targeted sites for 12 *nirK* gene PCR primers. We found all literature primers would miss about 95% of soil *nirK* sequences (≤ 2 mismatches), while validated best MSU-developed primer provided a calculated > 90% of coverage (≤ 2 mismatches).

We also performed amplicon sequencing and analysis using bacterial 16S (V4), fungal (28S), and *nifH* genes from soil cores collected in 2013 at intensive experimental sites at KBS and Arlington from five different biofuel cropping systems: Continuous corn (G1), Continuous corn plus Cover crop (G2), Switchgrass (G5), Miscanthus (G6), and Native prairie (G10). We found clear differences between corn main plots (stover removal) and microplots (stover non-removal) both in G1 and G2 plots. Furthermore, the bacterial communities in the three biofuel crops were very different from native prairie, and the communities in continuous corn were more different from other two biofuel crops than the latter were to each other, consistent with the shotgun data analyses of the rhizosphere samples.

A benchmarking study was carried out to identify genes expressed in soil associated with the biofuel crop Miscanthus at the KBS site. Initial results indicated an increased abundance of transcripts derived from housekeeping genes and phage in the later sample, suggesting the possible effect of Miscanthus on gene expression of the soil microbial community.

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

12. Nutrient Cycling for Biomass: Biochemical and Proteomic Profiling of a Nutrient Limited Poplar x *Laccaria* Ectomycorrhizal System

Leland J. Cseke[1*] (csekel@uah.edu), Avinash Sreedasyam,[1] Geetika Trivedi,[1] Bich Tran,[1] Stefanie Lasota,[2] Alistair Rogers,[2] Landon S. Wilson,[3] Helen Kim,[3] and Stephen Barnes[3]

[1]The University of Alabama in Huntsville, Huntsville, AL; [2]Brookhaven National Laboratory, Upton, NY; [3]The University of Alabama at Birmingham, Birmingham, AL

Project Goals: This project addresses the need to develop system-scale models at the symbiotic interface between ectomycorrhizal fungi (such as *Laccaria bicolor*) and tree species (such as poplar) in response to environmental nutrient availability/biochemistry. A multiple "omics" approach is implemented that integrates next generation sequencing transcriptomics, proteomics, biochemical analyses and ChIP-Seq analyses to construct ectomycorrhizal regulatory networks and computational modeling approaches to predict how atmospheric carbon is sequestered as plant and/or subsurface fungal biomass.

Poster Abstract:
Ectomycorrhizal (ECM) symbiosis significantly affects carbon (C), nitrogen (N) and phosphorus (P) uptake in many plants, including commercially important Poplar trees. In order to understand the molecular mechanisms underlying this essential ecological phenomenon, our project employs a controlled laboratory *Populus tremuloides* X *Laccaria bicolor* model system. Biochemical analyses were performed to quantify key C, N and P metabolites within the leaves, stems and roots of ECM trees subjected to nutrient limiting conditions, and differential shifts in C and N allocation were identified that are indicative of the tree's ability to produce biomass. This data was correlated with the biochemical pathways associated with each metabolite as a means to identify important proteins that drive significant biochemical processes. Out of the key pathways, ~600 selected proteins will be analyzed by quantitative nanoLC-pseudoMRM-MS targeted proteomics. To determine the proteotypic peptides needed for such protein profiling, we developed an updated global protein database for *P. tremuloides* by assembling previously obtained transcriptomic data (780 million RNA-Seq short reads). Using global nanoLC-MS analyses, the newly predicted protein sequences were compared to trypsinized proteins extracted from *P. tremuloides*, resulting in a 22.4% increase in successful protein identifications over previously available databases. A major objective of this project is to integrate transcriptomic, proteomic and biochemical data in order to build system-scale models of the molecular mechanisms involved in improved nutrient uptake and carbon sequestration in ECM systems under nutrient limiting conditions. Such models will be the basis for the prediction of ecosystem responses to environmental nutrient availability.

13. Nutrient Cycling for Biomass: ChIP Sequencing of Transcriptional Factors to Infer Gene Regulatory Networks for Carbon and Nitrogen Sequestration During Poplar x *Laccaria* Ectomycorrhizal Interactions

Avinash Sreedasyam[1*] (as0005@uah.edu), Geetika Trivedi,[1] Sarah Zerbs,2 Peter E. Larsen,[2] Frank R. Collart,[2] and **Leland J. Cseke**[1]

[1]The University of Alabama in Huntsville, Huntsville, AL; [2]Argonne National Laboratory, Lemont, IL

Project Goals: This project addresses the need to develop system-scale models at the symbiotic interface between ectomycorrhizal fungi (such as *Laccaria bicolor*) and tree species (such as poplar) in response to environmental nutrient availability/biochemistry. A multiple "omics" approach is implemented that integrates next generation sequencing transcriptomics, proteomics, biochemical analyses and ChIP-Seq analyses to construct ectomycorrhizal regulatory networks and computational modeling approaches to predict how atmospheric carbon is sequestered as plant and/or subsurface fungal biomass.

Poster Abstract:
RNA sequencing (RNA-Seq) and ChIP sequencing (ChIP-Seq) generate comprehensive transcript abundance data along with increased resolution on transcriptional regulation. However, these next-generation approaches are only beginning to be used to understand the mechanisms underlying increased plant biomass during beneficial plant-microbial interactions. In our study, we use these approaches to explore carbon and nitrogen sequestration by employing a laboratory-based *Populus tremuloides* X *Laccaria bicolor* ectomycorrhizal system. Based on past RNA-Seq analyses of this system, 16 transcription factors (TFs) were identified that are important in mycorrhiza-specific carbon and nitrogen flux metabolism. We performed ChIP-Seq analyses using antibodies generated against these TFs, including several MADS-Box TFs known to be involved in root and shoot development as well as a commercially available RNA- polymerase II antibody as an experimental control. 961 loci were identified as the potential targets of these MADS-Box regulators. In addition, the CC[A/T]$_6$GG motif, termed the CArG- box, is one known binding site for the MADS domain, and this motif was over-represented in the promoter sequences of the identified target genes, thus validating the ChIP-Seq analysis pipeline in *P. tremuloides*. Moreover, some of the identified target genes are involved in key biochemical and metabolic pathways associated with cellulose and lignin biosynthesis during growth and development. Thus, it appears that the ChIP-Seq approach will work well to enhance our understanding of biomass production in forest trees. Such data will be linked with transcriptome data to build gene regulatory networks and develop system-scale models that are predictive of the molecular mechanisms that control plant carbon management and allocation.

14. Mapping soil carbon from cradle to grave #3: Plant-microbial interactions regulate soil C cycling

Shengjing Shi[1,2*] (shengjing.shi@berkeley.edu), Donald Herman,[1] Erin Nuccio,[1,3] **Jennifer Pett-Ridge,[3] Eoin Brodie,[4] Trent Northen,[5] Zhili He,[2] Jizhong Zhou,[2] and Mary Firestone[1,4]**

[1]University of California, Berkeley, [2] University of Oklahoma, Norman, [3] Chemical Sciences Division, Lawrence Livermore National Laboratory, [4] Earth Sciences Division, Lawrence Berkeley National Laboratory, [5] Life Sciences Division, Lawrence Berkeley National Laboratory

Project Goals: Our project is focused on gaining a fundamental understanding of C cycling in soil, as mediated by soil microorganisms and their interactions with plants. Our work investigates how the interactions between roots and soil microorganisms affects transformation of root derived C, decomposition and loss as CO_2, as well as soil C sorption and stabilization at ambient and elevated levels of atmospheric CO_2.

Plants transfer atmosphere CO_2 into belowground soil C pools where soil microorganisms are primary mediators of C transformation and mineralization. Plant roots can strongly affect microbial transformation and mineralization of root-derived C but the molecular mechanisms underlying these plant-microbial interactions are poorly understood, as are the possible modulations caused by changing climate. We examined the effects of live *Avena fatua* roots (a common annual grass) on decomposition of ^{13}C-labeled root litter in a California grassland soil over two simulated growing-seasons. The presence of live roots consistently suppressed rates of litter decomposition; however this effect disappeared with plant senescence. Presence of live roots significantly altered the abundance, composition and functional potential of microbial communities (assessed by qPCR, MiSeq 16S and ITS sequencing, and GeoChip 4, respectively). Two possible mechanisms (preferential substrate utilization and drying stress) were identified for explaining the reduced rates of litter decomposition in the presence of live plant roots.

We also investigated the influence of elevated CO_2 (eCO_2) on C cycling through plant-microbial interactions. Plants grown with eCO_2 (700 ppm, $^{13}CO_2$) increased both total C allocated belowground and the amount of root-derived ^{13}C in the mineral-associated fraction, which is generally considered to be relatively stable soil C. Microbial communities associated with *Avena fatua* roots at different plant growth stages were analysed using Illumina 16S sequencing. Although eCO_2 caused little effect on the composition of bacterial communities in rhizosphere or bulk soil, the age of plant roots did play an important role in shaping rhizosphere microbial communities and driving their succession. Microbial diversity indices (richness, phylogenetic diversity, Shannon, Peilou's evenness) were significantly reduced in rhizosphere soil as compared to bulk soil, possibly due to bacterial responses to root carbon inputs and root-induced changes in soil microbial niches. Network analyses revealed successively greater complexity of microbial interactions in rhizosphere microbial communities compared to those in bulk soil. Understanding the mechanisms by which roots influence the assembly of rhizosphere microbiomes may allow delineation of generalizations useful in modelling terrestrial carbon processes.

15. Mapping soil carbon from cradle to grave #2: Multi-level omics analyses for parameterization of trait-based models of rhizosphere microbial community function

Eoin L. Brodie,[1,3*] (elbrodie@lbl.gov) Ulisses Nunes da Rocha,[3] Erin E. Nuccio,[1,4] Ulas Karaoz,[3] Nicholas Bouskill,[3] Shengjing Shi,[1,2] Zhili He,[2] Jason Shi,[2] Donald Herman,[1,3] Rachel Neurath,[1] **Peter S. Nico,[3] Mary Lipton,[5] Trent R. Northen,[3] Jizhong Zhou,[2,3] Mary K. Firestone[1,3] Jennifer Pett-Ridge[4]**

[1]University of California, Berkeley, [2]University of Oklahoma, Norman, [3]Lawrence Berkeley National Laboratory, Berkeley, [4]Lawrence Livermore National Laboratory, and [5]Pacific Northwest National Lab

Project goals: Our primary objective is to determine how organic C decomposition and stabilization processes in soil are impacted by the interactions between plant roots and the soil microbial community (bacteria, archaea, fungi, microfauna). To accomplish this requires information on the key features or traits of grassland soil microorganisms that impact their fitness as members of the rhizosphere and/or detritosphere. Through the identification of these traits, we propose to develop and parameterize trait-based models of microbial community function to interpret and predict carbon stabilization and turnover in the rhizosphere.

Theoretical approaches to understand the links between community diversity and ecosystem processes have become important tools in ecology. In particular, the application of trait-based modeling to explain complex patterns in taxonomic distribution across spatial and temporal scales and environmental gradients is increasingly common within studies of plant ecology. These approaches also show promise for improving hypothesis testing in microbial ecology, by developing frameworks linking microbial functional guilds (metabolically diverse organisms with a common function like lignocellulosic biomass decomposition) with the physiological and ecological traits that govern fitness independent of their phylogeny. We are adopting a Dynamic Energy Budget-based heterotrophic framework to incorporate metabolic theory into our ecological framework. When used to describe grassland soil microbial communities, this model has the potential to reproduce the diversity of organic molecules (e.g., exudates and polymer pools) and different exudate input rates and stoichiometry - thus selecting for different combinations of ecophysiological traits that maximize fitness under specific conditions resulting in a dynamic emergent community.

To begin to dissect the grassland soil microbial communities into its functional guilds we are using a range of 'omics approaches to characterize metabolic potential and niche preference. We have sequenced soil metagenomes taken at two physiologically relevant time points (peak plant activity and prior to wet-up) and are reconstructing genomes. Predicted genes relevant to soil biogeochemical cycling are being functionally annotated using a comprehensive suite of HMM models. We have also generated a large library of isolated bacterial heterotrophs and have sequenced 40 of those to date. Niche preference of bacteria, archaea and eukarya is being determined using non-targeted metatranscriptomic sequencing and phylogenetic marker reconstruction.

For our two soil metagenomes, between 250-400 Gb of high quality sequence data was obtained. Assembly and binning are ongoing, but initial classification shows both metagenomes are dominated by Actinobacteria and Alpha-proteobacteria. We have developed a functional gene based database and pipeline for metagenome sequence data analysis and are testing this for annotation of unassembled and assembled data. 290 bacterial isolates were obtained from multiple dilute media formulations incubated over 2.5 months. At the 97% homology level, the majority of OTUs were unique to one media type. Comparison with estimates of OTU abundance reconstructed from metagenomes using Emirge shows that

the isolates represented between 8 and 13% of the soil bacteria by relative abundance with a similar phylogenetic distribution to the complete bacterial community. Preliminary analysis of 14 bacterial isolate genomes (Actinobacteria, Alpha-, Beta-, Gamma-Proteobacteria, Firmicutes, Bactreroidetes) demonstrates a varied repertoire of carbohydrate active enzymes with little phylogenetic signal. Analysis of codon usage bias suggests differential minimum generation times that may relate to growth strategy.

In order to determine whether specific members of the soil microbial and microfaunal communities showed niche preference for litter-containing rhizosphere versus bulk soil, we extracted and sequenced total RNA from appropriate regions of greenhouse-maintained mini-rhizotrons cultivated with the annual grass *Avena fatua*. SSU rRNA genes were reconstructed using Emirge. Using this PCR-independent approach we determined that the litter in the presence of living roots selects for numerous Actinobacteria and Chloroflexi as well as fungi, while litter in bulk soil selected for Firmicutes and Bacteroidetes in addition to protists from the Amoebozoa and Alveolata amongst others.

Together these data are being used to determine the ecophysiological traits associated with rhizosphere enrichment and litter decomposition in grassland soils. Our overall goal is to use these and other information to assign functional roles to soil microorganisms and to develop mathematical models to predict their dynamics and contributions to soil carbon transformation.

16. Mapping soil carbon from cradle to grave #1: drafting a molecular blueprint for C transformation from roots to stabilized soil organic C

Erin Nuccio,[1,4] Rachel Neurath,[1] Shengjing Shi,[1,2] Zhili He,[2] Donald Herman,[1,3] Ulisses Nunes da Rocha,[3] **Peter Nico,**[3] **Mary Lipton,**[5] **Trent Northen,**[3] **Jizhong Zhou,**[2,3] **Eoin Brodie,**[1,3] **Jennifer Pett-Ridge,**[4] **Mary K. Firestone**[1,3*](mkfstone@berkeley.edu)

[1]University of California, Berkeley, [2]University of Oklahoma, Norman, [3]Lawrence Berkeley National Laboratory, Berkeley, [4]Lawrence Livermore National Laboratory, and [5]Pacific Northwest National Lab

Project Goals: The primary goal of this on-going project is to determine how organic C decomposition and stabilization processes in soil are impacted by the interactions of the soil microbial community with living roots. We are attacking this ambitious objective by testing three hypotheses:

H1: The microbiomes of the rhizosphere and detritosphere undergo a functional succession driven by the molecular composition and quantity of root-derived C.

H2: Elevated CO_2 impacts the function and succession of rhizosphere communities thus altering the fate of root-derived C.

H3: Microbial metabolism of root-derived C is a critical controller of sorption and desorption of organic C to mineral surfaces in soil.

We will draft a molecular blueprint for C transformation from root to stabilized soil organic C by:
1) Tracking the functional succession of microbial communities during utilization of root exudates and decaying root litter under ambient and elevated CO_2. We are using genomic, transcriptomic, and proteomic approaches (stable isotope assisted when possible) to identify key metabolic pathways responsible for C transformation and mineralization during root in-growth and root death/decay; we will assess how these rates and pathways will change in response to elevated CO_2 (eCO_2).

2) Measuring how the composition and quantity of root C affects the enzymatic capacities of the root microbiome and ultimately impacts C sorption/desorption from mineral surfaces. By coupling stable isotopes with metabolomic, transcriptomic and proteomic techniques, NanoSIMS (isotopic imaging), and STXM (synchrotron molecular imaging), we are identifying critical root C components, resulting enzymatic capacities and ultimately, the microbial mechanisms that impact sorption and desorption processes. A hierarchical design, ranging from intact rhizospheres to simplified model rhizospheres, are providing experimental systems in which proteomics, metabolite profiling, and Nimzyme-based enzyme assays are tractable but in which development and application of these emerging techniques to soils will be furthered.

3) Using data generated in the above experiments, we will parameterize a trait-based model of the microbial community functions which underlie carbon stabilization and turnover in the rhizosphere. The new trait-based model will provide a critical framework for organization and resolution of the extensive molecular and environmental data that will result from this research. The research proposed builds on an extensive foundation of past work by the PIs exploring root-soil-microbe interactions.

Plant roots and associated soil microbiomes comprise the primary nexus of belowground carbon cycling in terrestrial systems. Roots provide the source of C for stabilized organic matter in soil and it is now understood that most plant C is utilized or transformed by soil microorganisms en route to stabilization. Microorganisms supply enzymes that catalyze the decomposition and transformation of plant C into cell materials and products that are stabilized by physical protection (e.g. sorption to mineral surfaces). The proposed project directly targets the DOE goals of understanding C sequestration through plant-

microbial interactions. The information gathered will also benefit biofuels research as well as plant-stimulated bioremediation. Because microscale interactions between roots and soil microbes determine transformation of root C, decomposition and loss as CO_2, as well as C sorption and stabilization, our work will begin to reveal how microbial genes provide a blueprint for defined components of C flow in soil.

17. Compartmentalization of metabolic functions in microbial communities

William C. Nelson[1*](william.nelson@pnnl.gov), William Inskeep[2], Allan Konopka1, Margie Romine[1], Stephen R. Lindemann[1], Alex Beliaev[1], **James K Fredrickson[1]**

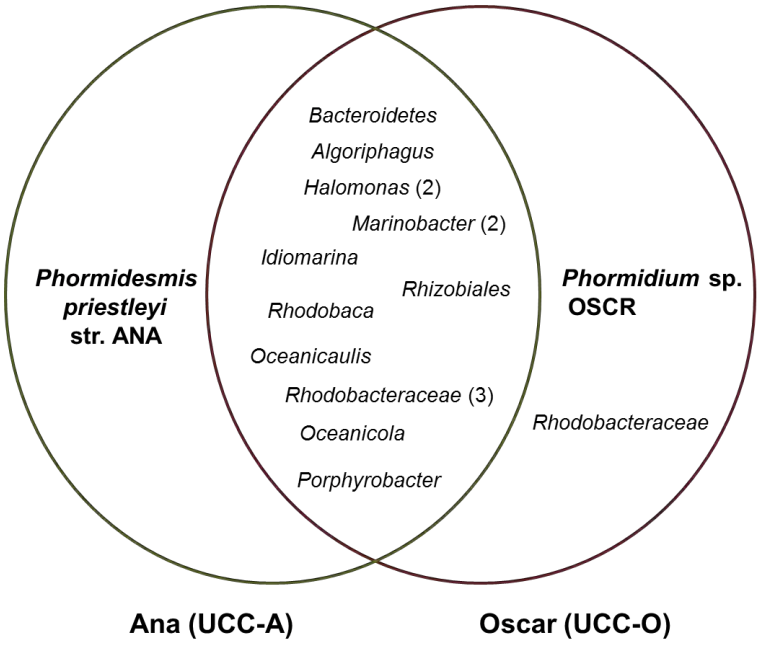
[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA; [2]Montana State University, Bozeman, MT

PNNL Foundational Scientific Focus Area, PIs: **Jim Fredrickson** and **Margie Romine**

Project Goals:

Microbial mats are laminated biofilm communities of bacteria, archaea, and microeukaryotes embedded in extracellular polymeric substances. The complex and compact spatial structure promotes metabolic networks among community members, leading to efficient energy utilization, nutrient cycling, and obligate mutualistic relationships. Assembly of these communities is driven by a combination of stochastic forces (e.g., colonization order) and deterministic forces (energy acquisition, nutrient availability, resource competition, microbial interactions), while spatial arrangement is constrained by gradients of physical and chemical parameters formed by environmental forces and microbial function. We hypothesize that functional compartmentalization is a mechanism which promotes community diversity and metabolic interaction. These interactions have a stabilizing effect upon diversity, and thereby lead to community resilience to stochastic environmental variation.

To examine niche partitioning and metabolic functional compartmentalization and the effects of key environmental variables upon community composition, we are using metagenomics to study two microbial communities: 1) a laboratory system of two uni-cyanobacterial consortia (UCC-A and UCC-O) derived from the hypersaline Hot Lake phototrophic mat, each containing a single photoautotroph and a stable heterotrophic assemblage, and 2) a field system of low-complexity chemotrophic Fe^{2+}-oxidizing mat communities from geothermal springs. In addition, we have axenic cultures and single-cell genomes (in progress) from each system. The simplified structure of these communities has allowed us to generate near-complete genome sequence for most members of each community. The genomic data is being used to predict the metabolic function of the individual species, which provides a foundation for predicting metabolic interactions. These predictions are subsequently tested through controlled manipulations of community-derived isolates and consortia. Field observations of environmental parameters have been correlated with changes in the community composition, spatial structure and gene content (metabolic potential) to identify key factors driving community assembly and stability.



Metagenomic sequence data collected from the UCC-A and UCC-0 cultures were assembled and segregated into taxonomic bins resulting in 18 distinct, near-complete (est. >90%) genome sequences, 15 of which are shared between the two communities. Each consortium contained a single cyanobacterium that is the sole obligate autotroph, capable of fixing inorganic carbon and N_2. Moreover, each consortium contains a heterotrophic assemblage comprised of members of *Bacteroidetes*, *Gamma-proteobacteria*, and *Alpha-proteobacteria*. Most organisms in the cultures can use urea as a nitrogen source and the gamma-proteobacteria can use cyanate, a by-product of the urea cycle. The alphaproteobacterium *Oceanicola* sp. possesses the most diverse set of carbohydrate catabolic genes and is capable of degrading a range of mono- and di-saccharides, organic acids and sugar alcohols. Conversely, the gammaproteobacterium *Idiomarina* and alphaproteobacterium *Oceanicaulis* spp. are putative amino acid fermenters. Isolate strains were tested for growth on various carbon sources to test these predictions. These predictions highlight examples of potential niche specialization and functional compartmentalization among community members that will be empirically validated using our consortia and isolates of organisms present within them.

Spatial gradients in environmental parameters also promote niche partitioning and functional compartmentalization within microbial mat communities. Within the Yellowstone One Hundred Spring Plain (OSP) Fe^{2+} mats, dissolved O_2 concentrations decrease rapidly with mat depth. As a result of biological consumption O_2 penetration is limited to 700 μm. This gradient correlates with changes in community composition and metabolic potential. The distribution of aerobic (*Hydrogenobaculum, Metallosphaera*) versus anaerobic (*Acidilobales*) populations varies with mat depth, suggesting that O_2 is a critical variable driving community spatial arrangement and stability. Transcript abundance of *Metallosphaera* heme copper oxidases (*foxA*) was higher in the surface layer of OSP mats, and metabolic reconstruction from genome assemblies shows different functional capabilities consistent with the observed spatial compartmentalization.

Together, results from two disparate community types show the importance of metabolic partitioning in promoting microbial interaction. This effect is observed as a function of both time and space and at the resolution of single organisms and entire communities. Our future goal is to develop a mechanistic understanding of niche specialization, spatial and temporal structuring of microbial communities, and individual and community response to changes in environmental variables. Measurement of metabolic potential and/or response in studies of natural communities and isolates/consortia will provide a basis for testing key metabolic interactions as a function of time and space, and how these attributes link with physicochemical properties of environmental systems.

18. FSFA Poster #2: Metabolic Interactions Structuring Microbial Communities

Alex Beliaev[1*] (alex.beliaev@pnnl.gov), Dmitry Rodionov,[2*], Andrei Osterman[2], Young-Mo Kim[1], Meng Markille[1], Steve Wiley[1], Tom Metz[1], Alice C. Dohnalkova[1], Allan Konopka[1], and **Margie F. Romine[1*]** (margie.romine@pnnl.gov).

[1]Pacific Northwest National Laboratory, Richland, WA
[2]Sanford Burnham Institute, CA

PNNL Foundational Scientific Focus Area, Jim K. Fredrickson and Margie F. Romine, PIs

The overarching goal of the PNNL Foundational SFA is to identify the principles that govern assembly, stability, and resilience of self-organized microbial communities to enable prediction of response to perturbation and control for maintenance of system function. Interactions between component members of communities are major forces maintaining community function in the face of perturbations. Complementary sequence-based functional predictions and experimental validation studies are being applied in iterative fashion to identify the spectrum of interactions that occur in natural communities, derived consortia, or synthetic co-cultures. Hypotheses- driven experimental approaches involve a suite of physiological, biochemical, and 'omics measurements performed in conjunction with controlled cultivation to systematically perturb organisms and measure response to determine network structure and identify potential interactions between members.

Through the application of next-generation sequencing technology in conjunction with controlled cultivation and metabolomic profiling, we carried out analysis of differentially regulated transcripts in a synthetic cyanobacterium-heterotroph consortium grown on either organic or inorganic C source. The obtained information led to the development of a conceptual model of interactions between phototautotrophs and heterotrophs as a function of carbon sources and flux directions (Fig. 1). Although the overall response was broad and complex, the global transcription patterns indicated competition for available carbon resources and maximization of metabolic capacity. This was manifested through the induction of transport and catabolic pathways involved in the assimilation of both monomeric and complex polymeric substrates, such as exopolysaccharides, peptides, and nucleic acids. This was particularly evident in the organism that was incapable of utilizing the primary carbon source (inorganic carbon vs. lactate). Substantial evidence supported metabolic coupling via the exchange of amino acids and co-factors. Co-cultivation with the heterotroph stimulated expression of facile and

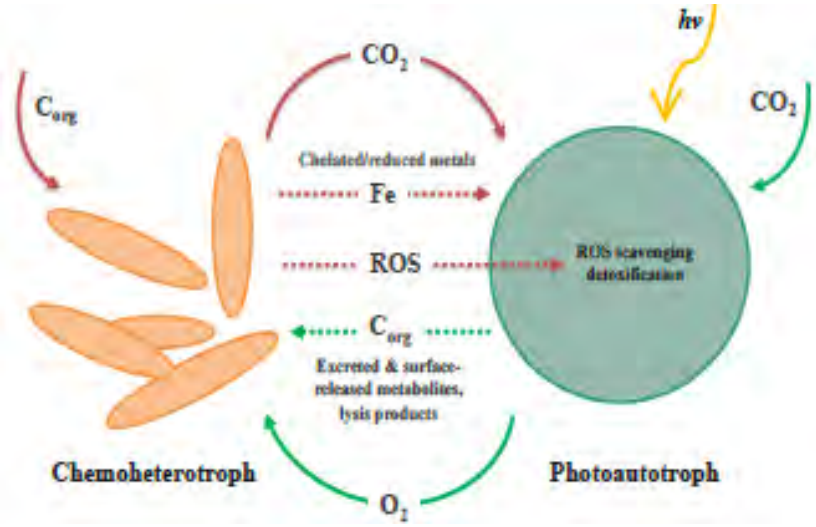


Fig. 1. Proposed mechanisms of interactions between of photoautotrophic and heterotrophic cells growing in co-culture. Arrows represent fluxes driven by photoautotrophic (green) and heterotrophic (red) metabolism.

energy-efficient pathways for Fe acquisition by the autotroph; Fe is required for the extensive electron transport pathways of photosynthesis. In contrast to other studies where heterotrophs were hypothesized to reduce phototroph oxidative stress via catalase-dependent ROS scavenging, the transcript levels of genes involved in oxidative stress response and scavenging of ROS radicals were down-regulated under co-culture conditions. This suggests that some phototrophs may be the provider of protection against oxidative stress in phototroph-heterotroph consortia. Isolates and consortia from hot spring and hypersaline mats are now allowing investigation of whether observed behaviors in synthetic consortia hold for functionally similar consortia derived from natural systems.

Complementary comparative analyses of genome sequences (derived from mat isolates and metagenome subassemblies) have developed predictions of resource sharing that are being tested experimentally using gene knock-outs/knock-ins, biochemical assay, or targeted synthetic chemical probes. B-type vitamin biosynthesis and salvage are broadly recognized as key interactions driving autotroph-heterotroph associations, but knowledge gaps in genes responsible for salvage and transcriptional control of co-factors prevent accurate prediction, probing, and modeling of exchange in microbial communities. Chemical probes have been developed for vitamins B_1, B_2, B_3, B_7, B_9, and B_{12}, and coenzymes derived from them, and have been shown to selectively target specific vitamins *in vitro*. Probes are currently being used on whole cultures (axenic or mixed) to assay uptake and identify proteins associated with uptake. Using comparative genomics we have identified the first archaeal transcriptional regulators of riboflavin, cobalamin, and thiamin biosynthesis; in bacteria control is mediated by riboswitches rather than via protein binding near promoters. An archaeal NAD-specific transcriptional regulator was also identified in Thermoproteales and Metallosphaera. Regulon and genome context analysis enabled discovery of three candidate salvage systems (transporters) for riboflavin and thiamin and one for niacin.

In both the chemotrophic and phototrophic communities it appears that vitamin auxotrophy is limited primarily to heterotrophs suggesting that autotrophic organisms are an important source of essential co-factors early in community assembly. Analysis of thermophilic iron mat metagenome subassemblies suggest that *Hydrogenobaculum*, an early colonizing chemoautrophic bacterium, is the only source of biotin. Similarly, in unicyanobacterial consortia derived from a saline mat, cobalamin is predicted to be synthesized by only two of 17 taxa (Phormidium and Rhizobiales) present in co-cultures grown in the absence of external vitamin supplementation.

Using our unicyanobacterial consortia (UCC) we are also examining the potential role of extracellular polysaccharides (EPS) in metabolic coupling of autotrophs and heterotrophs. Metabolite analysis of loosely associated EPS extracted from an 18 member UCC identified tartaric acid, glutamate, D-xylose, fucose, glucose, and mannose as primary carbohydrate constituents. Analysis of the metagenome subassemblies from the 18 taxa present in this UCC suggest that the potential to metabolize each of these carbohydrates are present in community members. Mannose metabolic potential was found in all but four of the taxa, which was consistent with this also being the most abundant metabolite detected in EPS. A mosaic distribution of catabolic pathways for the other carbohydrates was observed and hypothesized to minimize competition for EPS constituents as carbon sources.

19. Identifying environmental state variables governing the assembly of microbial communities

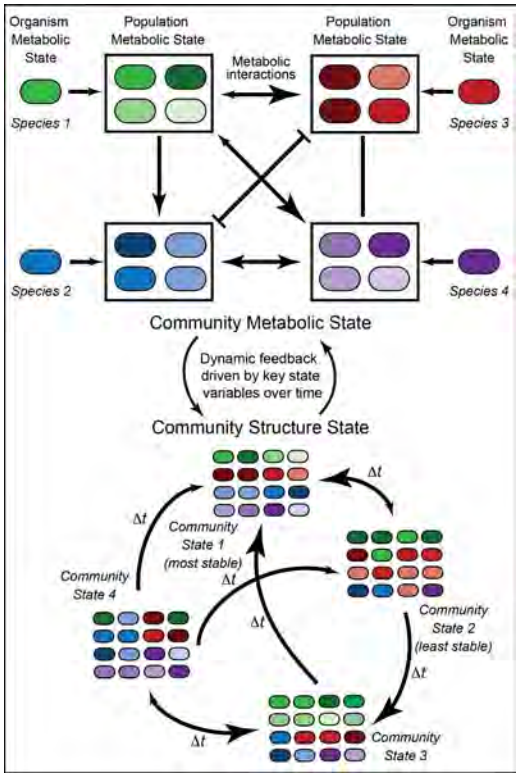
Stephen R. Lindemann[1*] (stephen.lindemann@pnnl.gov), William P. Inskeep[1,2*] (binskeep@montana.edu), Jacob P. Beam[2], Jessica K. Cole[1], Janine R. Hutchison[1], Ryan S. Renslow[1], and **James K. Fredrickson**[1].

[1] Pacific Northwest National Laboratory, Richland, WA. [2] Montana State University, Bozeman, MT

PNNL Foundational Scientific Focus Area, PIs: Jim Fredrickson and Margie Romine

Project Goals: To elucidate the underlying principles governing microbial community assembly by identifying the biotic and abiotic drivers determining recruitment and maintenance of phylogenetic and functional diversity in dynamic communities. Further, to examine how the interspecies interactions operating within a community are influenced by key environmental state variables, and how these effects feedback to community structure and function. To reveal how changes in community structure and function lead to alteration in higher-order community properties (e.g., resilience after perturbation) and predict environmental "tipping points" in community structural and functional state.

Microbial mats are compact ecosystems composed of highly interactive organisms in which complete energy and element cycles occur. These communities experience significant variation in key state variables (e.g., light quality and quantity, temperature, salinity, oxygen, pH) over both short (e.g., diel) and long (e.g., seasonal) time scales. To investigate overarching principles governing microbial community assembly, we examined temporal variation in disparate mat communities at multiple time points over seasonal cycles. We hypothesize that higher-order community properties (e.g., resistance to environmental change) emerge in microbial communities in predictable and generalizable ways based upon the structural and functional states of a community and that arise from the interspecies interactions operating between members. Two tractable and highly-structured microbial mat systems were chosen for this work (one phototrophic, one chemotrophic) to investigate system- independent principles of microbial interaction.



Hot Lake is a hypersaline lake that seasonally harbors a benthic, phototrophic mat that assembles each spring and disassembles each fall and is subject to greater than tenfold variation in salinity (primarily Mg^{2+} and SO_4^{2-}) and irradiation over the annual cycle. Similarly, mat communities exposed to overlying flow, such as those occupying low-pH geothermal outflow channels in Yellowstone National Park, exhibit strong spatial and structural control by pronounced geochemical gradients. The temporal self-organization of these structured microbial mats provides opportunities for

elucidating key state variables that drive interspecies interactions within dynamic communities and govern the interplay between community structure and function.

Relative abundances of microbial phylotypes were obtained using amplicon sequences of the 16S rRNA gene (V4 region), which were cross-referenced to near-full-length 16S rRNA sequences for increased phylogenetic resolution. Despite extreme variation in environmental conditions over the season, the composition of the Hot Lake phototrophic mat community exhibited year-round stability in its key autotrophic (*Leptolyngbya* and *Phormidium* spp.) and heterotrophic (primarily, *Rhodobacteraceae*) membership. In contrast, by mid-fall the mat began to show signs of disassembly, with mat community richness and evenness declining precipitously under conditions of increasing salinity and decreasing irradiance and temperature. Elucidation of the principles governing interspecies interactions that drive microbial community assembly is challenging because of the diversity, complexity and intractability of natural system to *in situ* manipulation. Therefore we isolated two filamentous cyanobacteria and their co- isolated heterotrophic consorts from the Hot Lake mat to study primary succession in a mat-relevant model phototrophic community. The relative simplicity and tractability of these unicyanobacterial consortia allowed the characterization of the spatial and community structures of their assembling biofilms under controlled environmental conditions. Both consortia retained essentially the same suite of heterotrophic species, representing ten of the most abundant OTUs within the Hot Lake mat community. The abundance and growth rate of autotrophs dominated early in assembly, yielding to increasing rates of heterotroph growth late in succession. The two consortia exhibited similar patterns of succession at low taxonomic resolution, but substantial differences in the abundances of specific heterotrophic members depending upon which cyanobacterium served as primary producer. Taken together, these data suggest that, although similar niches are created by the cyanobacterial metabolisms, the resultant network of autotroph- heterotroph and heterotroph-heterotroph interactions are specific to each primary producer.

Likewise, community assembly in thermal chemotrophic communities displays similar patterns of autotroph-heterotroph succession. We studied community structure as a function of time and key state variables by temporal colonization experiments and by sampling along geochemical gradients. The distribution of autotrophic and heterotrophic community members varies depending upon pH and concentration of key electron donors and acceptors (S^{2-}, Fe^{2+}, O_2). Temporal studies in Fe^{3+}-oxide systems have also shown that autotrophic populations (especially *Hydrogenobaculum* sp.) are important colonizers in early mat development (2-30 days), and that heterotrophic populations generally establish in deeper mats (> 1 mm) that develop on time frames > 100 d. Metagenomic analyses revealed that many of these key heterotrophs are likely auxotrophs for synthesis of required cofactors, amino acids, and nucleotides, a result also observed in the Hot Lake unicyanobacterial consortia.

These data suggest that similar governing principles shape the succession of microbial communities inhabiting disparate environmental systems. Early colonization of prototrophic organisms leads to generation of niches for auxotrophic heterotrophs, which likely increase the functional capacity and efficiency of the community. Future work will focus on identifying specific state variables governing the interspecies interactions that shape community succession dynamics and the emergence of higher-order community properties. These genome-enabled studies will involve an iterative field-lab approach in which observations of the mat in the field are paired with laboratory-based perturbation studies of consortia, both assayed using integrated global molecular techniques (e.g., transcriptomics, proteomics, and metabolomics).

Amplicon, genomic and metagenomic sequencing were provided for this study by DOE's Joint Genome Institute under Community Sequencing Project 701.

20. Influence of shallow soil strata and chronic N deposition on the fungal community in pine and maple forests

Cedar Hesse[1], Rebecca Mueller[1], Blaire Steven[1], La Verne Gallegos-Graves[1], Chris Yeager[1], John Dunbar[1], Carolyn Weber[1,3], Donald R. Zak[2], and **Cheryl R. Kuske[1*]**(kuske@lanl.gov)

[1] Los Alamos National Laboratory, Los Alamos, NM, [2] University of Michigan, Ann Arbor, [3] Idaho State University, Pocatello

Project Goals: Anthropogenic N deposition is a chronic and increasing condition in temperate regions that may strongly influence C cycling dynamics. One major theme of our Science Focus Area is to determine the influence of chronic N deposition on microbial C cycling processes in two major biomes of Earth's temperate regions, forests and arid grass/shrublands. In both biomes fungal and bacterial biomass is concentrated in shallow surface soil strata where C/N cycling is a major process. The goals of this project are to (a) correlate the resident fungal and bacterial communities, enzyme activities, and local geochemistry across shallow strata of strongly stratified forest soils and biocrust dominated soils of arid grass/shrub lands, (b) determine the impacts of chronic N amendment across multiple forest and arid grass/shrubland biomes using a suite of long-term field experiments, and (c) compare the ability of phylogenetic rRNA gene surveys, soil enzyme assays, and metatranscriptome surveys to detect shifts in community structure and concomitant changes in C cycling processes in response to altered N conditions.

This poster describes highlights from our studies of soil fungal communities in two forest types, loblolly pine in central NC (Duke forest FACE site) and hardwood forest dominated by maple in MI, USA. Studies at the long-term maple N deposition field site are led by Donald R. Zak (University of Michigan) and are supported by both the NSF and the DOE. We measured soil/litter chemistry, relative abundance of fungal rRNA genes (qPCR), and community richness and composition (454 sequencing of the fungal LSU gene).

In the pine forest, the fungal community was dominated by Basidiomycota taxa that are often ectomycorrhizal associates with pine (1). Soil chemistry and resident fungal communities were significantly different across the forest floor, O horizon (0-2 cm depth) and A horizon (2-5 cm and 5-10 cm depths) under ambient conditions. After 5 years of N amendment (11.2 g N/m^{-2} annually as ammonium nitrate), the soil chemical conditions were significantly altered, especially in the forest floor and O horizon. Relative abundance of fungi in each horizon was not affected by N amendment, but composition of the resident fungal communities were significantly different in N amended conditions, and the response of the fungal community to N amendment differed across the strata. Fungal taxa found to be responsive to N differed with soil stratum.

The impact of 16 years of N amendment (3 g N/m^{-2} annually as sodium nitrate) on fungal relative abundance and community composition was surveyed in the forest floor, in four maple forests across a 10 km latitudinal transect. In contrast to the pine forest, the forest floor fungal community was dominated by taxa in the Ascomycota, however our survey primers did not efficiently detect members of the Glomeromycota that may be arbuscular mycorrhizal associates of the maple. The concentration of N amendment is lower at the four maple sites than in the pine forest experiment, thus differences in forest floor soil chemistry with N amendment were slight. Fungal relative abundance in the forest floor was not significantly impacted by N amendment at any of the four sites. Comparison of fungal community composition across the four sites showed highly significant differences with site, illustrating that the resident fungal community differed across the latitudinal transect even where the major plants and soil

conditions were similar. N amendment had a lesser, but significant, impact on fungal community structure in the forest floor and this effect differed among sites. The number and taxonomic assignment of responsive taxa to N deposition differed among the sites within this study.

Although the two forest field experiments are different in design and implementation, it is clear that chronic N deposition impacts the geochemistry and resident fungal community in both the pine and maple forest. Fungal community composition was affected by N deposition in site- specific ways. Soil metatranscriptomes have been generated from these soils, and soil enzyme assays are in progress (pine site). Together with the taxonomic information, these gene expression and enzyme activity surveys will identify shifts in C cycling patterns that are mediated by both fungal and bacterial communities.

(1) Weber CF, R Vilgalys, CR Kuske (2013) *Changes in fungal community composition in response to elevated atmospheric CO_2 and nitrogen fertilization varies with soil horizon*. Frontiers Terrest Microbiol. doi: 10.3389/fmicb.2013.00078

Funding statement: The information in this poster was supported by the U.S. Department of Energy Biological System Science Division (DOE BSSD), through a Science Focus Area Grant (2009LANLF260) to CRK, a DOE BSSD Microbial Communities grant to DRZ and CRK (maple forests), and the Los Alamos National Laboratory LDRD program.

21. Molecular tools and databases to monitor and classify soil fungal and bacterial communities

John Dunbar[1*] (dunbar@lanl.gov), Jason Gans[1], Andrea Porras-Alfaro[1], Gary Xie[1], Kuan-Liang Liu[1,3], Cedar Hesse[1], La Verne Gallegos-Graves[1], and **Cheryl R. Kuske[1]**

[1] Los Alamos National Laboratory, Los Alamos, NM; [2] Western Illinois University, Macomb; [3] Institute of Information Management, National Cheng Kung University, Taiwan

Project Goals: One goal of our program is to develop resources that broadly support the microbial ecology research community. We are conducting three activities that contribute to this goal: 1) production of hand curated fungal large subunit rRNA (LSU) and internal transcribed spacer (ITS) sequence databases, 2) validation of their classification accuracy using a naïve Bayesian fungal classifier and BLAST, and 3) production of quantitative PCR assays targeting bacterial and fungal taxa important in carbon cycling in soils from multiple biomes.

Fungal rRNA sequence databases and classifier validation. Reliable, publicly available sequence databases are critical for interpretation of ecosystem surveys and for design of molecular assays that can be used to monitor specific genes in large-scale ecosystem experiments and regional surveys. In 2011, we published a hand-curated LSU database that spanned the Ascomycota and Basidiomycota and demonstrated the classification accuracy of this database using a leave-one-out-classification scheme (1). This database was delivered to the Ribosomal Database Project (RDP) and has been publicly available. As of fall 2013, the RDP was averaging about 2,200 classifier runs and 925,000 query sequences per month against the LSU database. In the past year, we considerably improved the database in multiple ways. We increased coverage of the Glomeromycota, Chytridiomycota and other basal lineages, added representative non- fungal Eukarya, and resolved taxonomy discrepancies (cases where prior publications had called the same sequence by different names). The updated LSU database will soon be publicly available at the Ribosomal Database Project (RDP) website (http://rdp.cme.msu.edu/) and may be downloaded for local use (2).

Although the LSU gene is superior for phylogenetic placement of unknown sequences, the ITS region is widely used as a taxonomic barcode to identify fungal isolates, resolve species and strain differences, and to characterize diversity. We generated a curated ITS database (about 9,000 sequences) that spans all Fungi and demonstrated the accuracy with which the naïve Baysian classifier can correctly place sequences (3). Using parallel datasets for ITS and LSU, where both regions were sequenced from the same isolate (about 1100 sequences), we compared the classification accuracy of the two regions and found that, from phylum to genus, they gave comparable results across a range of sequence sizes and PCR anchor points. With both genes, species-level classification of sequences from environmental surveys is hampered by the lack of sufficient sequence coverage at this level in the databases.

Our soil surveys from multiple forest ecosystems show that a noteworthy proportion of the LSU sequences retrieved from soils represent substantially divergent, novel clades that are yet not classifiable to finer scale taxonomy. Clearly, both the LSU and ITS databases need to be expanded to improve coverage of environmental sequences that are not currently represented in culture collections.

Quantitative PCR assays for climate change responsive fungi and bacteria. Use of qPCR assays to detect and quantify microbial taxa and gene sequences within a complex background of microorganisms is a scalable, rapid and statistically rigorous approach to track populations of interest. However, the fragmentary nature and growing quantity of DNA-sequence data make group-specific assay design

challenging. We solved this problem by developing a software platform that enables PCR-assay design at an unprecedented scale. The platform provides a powerful capacity to address previously intractable assay design problems. The software accommodates the use of thousands of target and non-target sequences, allows degeneracies, applies sophisticated rejection criteria, and attempts to produce the minimum number of assays to detect a target group. Previously, we used the software to design qPCR assays for Acidobacteria Group 1 (4). We are currently designing assays for other bacterial groups (7 Actinomycete suborders and 10 families) and for fungal groups (11 Ascomycete suborders) of interest in terrestrial carbon cycling.

(1) Liu K-L, A Porras-Alfaro, CR Kuske, SA Eichorst, G Xie (2012) *Accurate, rapid taxonomic classification of fungal large subunit rRNA genes.* Appl Environ Microbiol 78:1523-33.

(2) Cole JR, Q Wang, JA Fish, B Chai, D McGarrell, Y Sun, CT Brown, A Porras-Alfaro, CR Kuske, JM Tiedje (2013) *Ribosomal database project: data and tools for high throughput rRNA analysis*. Nucl Acids Res, doi:10.1093/nar/gkt1244, pg 1-10.

(3) Porras-Alfaro A, K-L Liu, CR Kuske, G Xie (2013) From genus to phylum: large-subunit and internal transcribed spacer rRNA operon regions show similar classification accuracies influenced by database composition. Appl Environ Microbiol, doi:10.1128/AEM.o2894-13.

(4) Gans JD, J Dunbar, SA Eichorst, LV Gallegos-Graves, M Wolinsky, CR Kuske (2012) *A robust primer design platform applied to detection of Acidobacteria Group 1 in soil*. Nucleic Acids Res 1-11 DOI:10.1093/nar/gks238.

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 (2009LANLF260), and by the Los Alamos National Laboratory LDRD program.

22. Soil carbon cycling communities and their response to climate and land use changes in patchy arid land ecosystems

Blaire Steven[1], La Verne Gallegos-Graves[1], Jayne Belnap[2], Chris Yeager[3], John Dunbar[1], Cedar Hesse[1], Rebecca Mueller[1], and **Cheryl R. Kuske[1*]**(kuske@lanl.gov)

[1] Los Alamos National Laboratory, Los Alamos, NM, [2] USGS, Moab, UT

Project Goals: The goals of this project are to (a) explore the relative utility of multiple 'omic approaches to track soil community responses to environmental changes, (b) map the relative abundance and composition of soil bacterial and fungal communities in shallow soil strata of arid land ecosystems at local and regional scales, (c) determine community responses to combinations of climate and land use changes, (d) identify key responsive community members with relevance as indicators of community change and with utility for modeling soil processes. Achieving these goals will provide an understanding of the active and responsive components of arid land soils that contribute to carbon cycling, their collective responses to environmental change, and development of efficient molecular tools for broad-scale soil monitoring.

The majority of arid land soil biomass resides in shallow strata and this biomass contributes greatly to C and N cycling, both as plant root associates and as biological soil crusts (biocrusts) that colonize the large patches between the widely spaced plants. Targeted rRNA gene and shotgun metagenomic approaches were used to map the relative biomass and composition of soil bacterial and fungal communities at multiple scales in the landscape. Using bacterial and fungal rRNA-based quantitative PCR assays, rRNA gene sequencing, and shotgun metagenomes the communities inhabiting root zones of the dominant shrub, *Larrea tridentata* (creosote bush), and the interspace biocrusts in a Mojave desert shrubland within the Nevada Free Air CO_2 Enrichment (FACE) experiment were mapped (1). Most of the numerically abundant bacteria and fungi were present in both the biocrusts and root zones. However, the proportional abundance of those members differed significantly between root-zones or biocrusts. Functional gene abundances in metagenome sequence datasets reflected the taxonomic differences noted in the 16S rRNA datasets. For example, functional categories related to photosynthesis, circadian clock proteins, and heterocyst-associated genes were enriched in the biocrusts, where populations of Cyanobacteria were larger. Genes related to potassium metabolism were also more abundant in the biocrusts, suggesting differences in nutrient cycling between biocrusts and root zones.

To understand the influence of soil type and soil depth (0-1 cm, 2-5 cm) on community structure, we used spatially nested sampling and 16S rRNA gene sequencing to describe the soil bacterial/archaeal communities in three soils derived from different parent material (2). In all three soils, Cyanobacteria and Proteobacteria demonstrated significantly higher relative abundance in the biocrusts, while Chloroflexi and Archaea were significantly enriched in the below-crust soils. Biomass and diversity of the communities in biocrusts or below-crust soils did not differ significantly with soil type, but composition was affected by soil type. The uniformity with which small-scale vertical community differences were maintained across larger horizontal spatial scales (5 m to 10 km) is a feature of dryland ecosystems that should be considered when designing management plans and determining the response of these patchy ecosystems to environmental disturbances.

Using multiple long-term (5 to 15 yr duration) ecosystem experiments supported by the DOE, USGS, and National Park Service, we conducted replicated field surveys to determine the impacts of multiple climate change factors and land use changes on soil bacterial and fungal communities. Fungal and bacterial community responses to over 10 years of elevated CO_2 were minimal (2, 3), except for a

surprising negative impact on the photosynthetic cyanobacteria, which comprises the dominant biocrust biomass (4). Quantitative PCR, rRNA sequencing and shotgun metagenomes all supported this observation.

Multiple climate and land use changes are operating in concert in arid lands worldwide. Physical damage (e.g. foot traffic) and altered environmental conditions caused by changes in precipitation pattern and/or warming temperatures have been shown to cause dramatic changes in biocrust structure and function. To determine the potential for resilience and regrowth of biocrusts after multiple years of physical or physiological damage, we characterized the bacterial communities in biocrusts after multiple years of foot traffic, altered precipitation pattern and 2- 3°C warming. Targeted rRNA sequencing identified significant differences in community structure when biocrusts were subjected to different types of stress. Impacts of physical disturbance and altered precipitation pattern were the most noticeable and were visibly similar at the soil surface. However, the impacted soil communities were different in structure, suggesting legacy effects specific to the type of disturbance. Furthermore, combined soil warming and altered precipitation resulted in biocrust compositional changes that differed from precipitation alone, highlighting the importance of considering the combined influences of multiple disturbances. Analysis of shotgun metagenomes and soil transcriptomes is in progress.

Using combinations of targeted and shotgun metagenome approaches, we have shown the benefits and pitfalls of each approach for detecting soil microbial community shifts. Target gene approaches generally provide more sensitive detection of taxonomic differences among complex communities across the inherent variability of field-scale experiments and shotgun surveys provide clues about altered physiology that may be a consequence of community structure changes.

(1) Steven B, LV Gallegos-Graves, C Yeager, J Belnap, CR Kuske (2014) *Common and distinguishing features of the bacterial and fungal communities in biological soil crusts and shrub root zone soils.* Soil Biol Biochem 69:302-312.

(2) Steven B, LV Gallegos-Graves, J Belnap, CR Kuske (2013) *Dryland soil bacterial communities display spatial biogeographic patterns associated with soil depth and soil parent material*. FEMS Microbiol Ecol 86:101-113.

(3) Steven B, L Gallegos-Graves, SR Starkenburg, PS Chain, CR Kuske (2012) *Targeted and shotgun metagenomic approaches provide different descriptions of dryland soil microbial communities in a manipulated field study*. Environ Microbiol Rep 4:248-256

(4) Steven B, LV Gallegos-Graves, CM Yeager, J Belnap, RD Evans, CR Kuske (2012) *Dryland biological soil crust cyanobacteria show unexpected decreases in abundance under long-term elevated* CO_2. Environ Microbiol 14:3247-3528.

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 (2009LANLF260). Sequencing was made possible through the DOE Joint Genome Institute and the Los Alamos National Laboratory LDRD program.

23. Systems-level dissection of anaerobic methane cycling: single cell ecophysiology, genetic mechanisms, and microbial interactions

Shawn E McGlynn[1], Grayson Chadwick[1], Chris Kempes[2], Jennifer Glass[1], Joshua Steele[1], Hang Yu[1], Karuna Chourey[3], Chongle Pan[3], Robert L Hettich[3], Tomas J Deerinck[4], Mark H Ellisman[4], Shulei Sun[4], Mohamed Haroon[5], Gene Tyson[5], and **Victoria J. Orphan[1*]** (vorphan@gps.caltech.edu)

[1] California Institute of Technology, Pasadena, CA [2] NASA Ames Research Center, Mountain View, CA [3] Oak Ridge National Research Lab, Oak Ridge, TN [4] University of California San Diego, San Diego, CA, [5] University of Queensland, Brisbane, AU

www.gps.caltech.edu/people/vorphan/profile

Project Goals: Energy and resource limitation typify much of microbial existence, and proliferation in such regimes often requires the interaction of multiple species, each catalyzing unique sets of reactions relevant to the community at large. Multispecies microbial consortia which carry out the anaerobic oxidation of methane are premier examples of growth permitting metabolic coupling, yet very little is known about the nature or controls of these globally relevant associations. By applying a diversity of 'omics and advanced microscopy techniques complimented by geochemical measurements, we are developing a unique systems-level perspective of methane-based microbial interactions. This coordinated research effort will yield fundamental insight into the associations driving anaerobic methane oxidation, and provides direction for the investigation of structured microbiological communities at-large, be they antagonistic, competitive, mutualistic, or engineered.

We are investigating nutrient utilization in multispecies consortia at multiple levels using community targeted 'omics based approaches in tandem with targeted microscale analysis of syntrophic microbial consortia at the level of individual cells using epifluorecence microscopy (FISH), stable isotope labeling and nanoSIMS. Metaproteogenomic investigations of sediment-hosted anaerobic methane oxidizing consortia have identified a potential tungsten-containing formylmethanofuran dehydrogenase- the first evidence of a tungstoenzyme expressed by psychrophilic methanotrophic archaea[1]. Analysis of ^{15}N enriched proteins from these methanotrophic consortia recovered from stable isotope tracer experiments is yielding insight into the pool of actively synthesized proteins by these slow growing anaerobic microorganisms. High resolution fluorescence and single cell stable isotope analysis of cells within consortia provide evidence that aggregate structure and distance between microbial partners are major contributors to the distribution of single cell metabolic activity within consortia mediating the anaerobic oxidation of methane. Between individual cells, anabolic activity rates are strongly correlated between neighboring members within a population, as well as between symbiotic partners. Surprisingly, the average anabolic activity of an individual aggregate does not appear to be a function of spatial structure (i.e. layered or well mixed), suggesting that environmentally derived resources are differentially partitioned within co-associated populations based on the geometry of cell- cell interactions. Using modeling approaches, we are now investigating the metabolic consequences of variable types of physiological interaction, and thereby confirm or eliminate particular nutrient and energy exchange regimes as candidates for producing our empirical observations.

Supported by the Office of Biological and Environmental Research

[1]Glass, J. et al (2013). Geochemical, metagenomic and metaproteomic insights into trace metal utilization by methane-oxidizing microbial consortia in sulphidic marine sediments. *Environ. Microbiol.* DOI: 10.1111/1462-2920.12314

24. Discovery of a Novel Colony Invading Phenotype of *Pseudomonas stutzeri* RCH2

Peter M. Rademacher[1*](PMRademacher@lbl.gov), Richard Baran[1], Megan A. Danielewicz[1], Romy Chakraborty[1], Morgan N. Price[1], Kelly M. Wetmore[1], Adam P. Arkin[1], Adam M. Deutschbauer[1], **Trent R. Northen[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley, CA

http://enigma.lbl.gov/

Project Goals: Project Goals: The Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) program broadly seeks to understand the interactions between environmentally relevant microorganisms and their environment. Presumably one of the most significant influences on microbial fitness is the microbial environment. Our goal in this project is to develop a generalizable approach combining large-scale microbial interaction screening, chemogenomics and metabolomics to better understand fitness in the context of other organisms.

Chemogenomics, using DNA-barcoded transposon mutant libraries, has proven to be a powerful approach to link genes to function under defined environmental conditions.We hypothesized that this approach could be used to find genetic determinants of microbial interactions. To expedite the discovery of relevant microbial interactions we performed a large-scale interaction mapping experiment using acoustic printing of soilisolates. Acoustic printing uses acoustical energy generated from a piezoelectric source to eject nanoliter droplets from multi-well plates into precisely guided locations on agarose gels.Using arrays of precisely grown bacterial colonies, we rapidly identified microbial interactions that affected colony size and morphology. We observed that *Pseudomonas stutzeri* RCH2 displays an invasive, rugose-forming phenotype when cultured in rich media on colonies of *Pseudomonads* and *Bacilli* bacteria isolated from the Oak Ridge Field Research Center. A DNA-barcoded transposon mutant library of RCH2 was used for co- culture mutant fitness assays to identify regulators of RCH2 colony invasion. Mutants that were highly sensitive and associated with this phenotype were genes encoding glutamate-5-kinase, gamma-glutamyl phosphate reductase, OHCU decarboxylase and formyltetrahydrofolate deformylase. The latter two genes are involved in purine metabolism, suggesting a role for purine metabolites in the colony invasion phenotype in RCH2. To further investigate the role of nucleobases in colony invasion, we performed metabolic profiling of scraped colonies using LC- qTOF-MS. As predicted from the mutant fitness profiling experiments, purine metabolites such as adenine were found at high levels in invaded colonies and were also consumed by the addition of RCH2. We are now constructing targeted mutants to further investigate the role of nucleobases in regulating the community behavior of P. *stutzeri* RCH2.

This work conducted 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, 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.

25. Natural and Synthetic Ecology in ENIGMA: Determining the links between Microbial Community Structure and Function

Dwayne A. Elias[1] (eliasda@ornl.gov), Michael W. W. Adams[2], Romy Chakraborty[3], Matthew W. Fields[4], Terry C. Hazen[5], Jizhong Zhou[6], Trent Northern[3], Nitin Baliga[7], John-Marc Chandonia[3], **Adam P. Arkin**[3] and **Paul D. Adams**[3]

[1]Oak Ridge National Laboratory, Oak Ridge, TN; [2]University of Georgia, Athens, GA; [3]Lawrence Berkeley National Laboratory, Berkeley, CA; [4]Montana State University, Bozeman, MT; [5]University of Tennessee, Knoxville, TN; [6]University of Oklahoma, [7]Institute for Systems Biology, Seattle, WA.

http://enigma.lbl.gov

Project Goals: The overarching goal of the Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) is to understand environmentally-relevant microbial community structure and function through a series of integrated field-to- laboratory campaigns. The Natural and Synthetic Ecology campaign is designed as an interdisciplinary platform to elucidate the fundamental ties between structure and function, as well as determine the environmental influences on these ties. Our initial focus has been the development of in-field bioreactors as a method of capturing temporal fluctuations in the *in-situ* community due to either natural or induced influences. Going forward we will determine the influence of naturally-occurring carbon sources, invasive species and selected pressures on microbial activities such as nitrate- and metal- reduction.

One of the most difficult aspects of studying microbial ecology is determining and understanding the fundamental ties between microbial community structure (the organism biodiversity and their relative abundances that comprise a given microbial community) and the observed functions (the detectable biochemical activities that support survival of the observed species). Although microorganisms are important in controlling the fate of contaminants in the subsurface, information on the basis of how why microbial communities respond to contaminants is lacking. Hence, it is important to characterize microbial communities, establish linkages between biodiversity and function, and study interactions between different species. As a part of the overall ENIGMA goal to link genotypes to phenotypes, the overall objective of this campaign is to obtain deep understanding of the composition, structure, function, activity and interaction of subsurface microbial communities at DOE contaminated sites (i.e. Oak Ridge Integrated Field Research Center). We have developed a bioreactor system for manipulating and temporally monitoring the *in-situ* microbial community in the field so as to maintain the *in-situ* community structure. Community structure was measured through sequencing, PCR and qPCR for selected genes, cell counts and total protein before and after the cells entered the bioreactor system. Temporal community function was qualified by alterations in the concentration of 53 metals, 12 organic acids, 14 anions and 4 sugars, pre- and post- bioreactor exposure. Near future experimental plans include determining an adequate naturally-occurring carbon source that will allow for an increase in total biomass with minimal alteration to the relative abundances of the major phylogenetic groups within the community. Finally, in order to establish the environmental relevance of new field isolates from this site, a pilot study using the bioreactor system will be conducted in collaboration with the Metals Metabolism campaign of ENIGMA. The bioreactors will use synthetic groundwater mimicking the geochemistry of the ORNL wells and will be inoculated with groundwater supplemented with various Mo concentrations since a current hypothesis is that a lack of Mo may inhibit nitrate-reduction. The experimental duration is expected to be 30 days with temporal

measurements of metals (53 elements), metabolites, 16S rRNA to determine changes in community structure, and by qPCR of key denitrification genes. End-point samples will be used for the isolation and characterization of new denitrifying strains. This work is highly collaborative, involving several ENIGMA campaigns including the 100 Well Survey, Microparticle Mesogenomics, Microbial Isolations and Characterizations and, Metals with down the line benefits to the Predictive Biology and Printable Worlds campaigns.

This work conducted 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, 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.

26. Metal Metabolism in ENIGMA: The Environmental Role of and Regulation by Molybdenum under Denitrifying Conditions

Michael W. W. Adams[1], Farris L. Poole[1], W. Andrew Lancaster[1], Brian J. Vaccaro[1], Michael P. Thorgersen[1], Romy Chakraborty[2], Pavel Novichkov[2], Adam Deutschbauer[2], Dwayne Elias[3], Judy Wall[4*] (wallj@missouri.edu), Aindrila Mukhopadhyay[2], John-Marc Chandonia[2], **Adam P. Arkin[2]** and **Paul D. Adams[2]**

[1]University of Georgia, Athens, GA; [2]Lawrence Berkeley National Laboratory, Berkeley, CA; [3]Oak Ridge National Laboratory, Oak Ridge, TN; [4]University of Missouri, Columbia, MO

http://enigma.lbl.gov

Project Goals: The overarching goal of the Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) is to understand environmentally-relevant microbial communities and processes through an integrated field-to-laboratory approach. A critical component of microbial community interactions is assimilation and metabolism of metals. The Metal Metabolism campaign is developing an interdisciplinary platform to elucidate the fundamental mechanisms that drive metal assimilation. Our initial focus is on the role of molybdenum in denitrification. Our functional goal is to understand fully the pathways and gene networks that coordinate Mo homeostasis. Of particular interest are the identities of proteins and regulatory elements that provide competitive advantages or disadvantages to microbes in defined biogeochemical environments.

Molybdenum (Mo) is an essential component of all nitrate reductases, the enzyme that reduces nitrate to nitrite, while copper (Cu) and iron (Fe) are essential components of other enzymes in the pathway for complete denitrification of nitrate to nitrogen gas. We originally hypothesized that biological nitrate reduction in contaminated wells at Oak Ridge National Laboratory (ORNL) is limited by the environmental availability of Mo, Cu and/or Fe. We have demonstrated in the ENIGMA 100-well Global Survey that in many of the contaminated wells containing high nitrate concentrations (>10 mM), the concentrations of Mo are very low (<10 nM) and in a range that severely limits nitrate reduction by the model denitrifer *Pseudomonas stutzeri* RCH2 under laboratory conditions. In contrast, concentrations of Cu and Fe measured in these wells are sufficient for denitrification. The first goal of the new platform will be to study the effects of Mo concentrations on denitrification at the protein and genomic level. Techniques and tools that have been developed and refined to investigate Mo metabolism include a barcoded transposon mutant library of *P. stutzeri* RCH2 and computationally-reconstructed regulons. These will be used together with μscale growth experiments with environmentally-relevant concentrations of Mo (1–100 nM) and nitrate (1–500 mM) to develop draft regulatory networks of cellular processes. Fifteen *Pseudomonas* strains have been isolated from samples obtained through the ORNL 100- well survey and, while closely related by 16S rRNA sequence, these isolates vary greatly in their tolerance and ability to reduce nitrate. Barcoded transposon mutant libraries will be developed for some isolates and dentrification and Mo limitation experiments will be carried out in order to gain genetic insights into Mo uptake, Mo homeostasis and denitrification. It is anticipated that this platform will be extended to field isolates from other genera such as *Castellaniella* and *Rhodanobacter* that appear to play important metal-based roles in well communities. To establish the environmental relevance of RCH2- and the field isolates, a pilot study with environmentally-based laboratory bioreactors will be conducted in collaboration with the Natural and Synthetic Ecology campaign of ENIGMA. The bioreactors will use synthetic groundwater mimicking the geochemistry of

the ORNL wells and will be inoculated with groundwater supplemented with various Mo concentrations. The experimental duration is expected to be 30 days with temporal measurements of metals (40 elements), metabolites, 16S rRNA to determine changes in community structure, and by qPCR of key denitrification genes. End-point samples will be used for the isolation and characterization of new denitrifying strains.

This work conducted 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, 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.

27. Computation Component of ENIGMA: from data to predictive models

Pavel S. Novichkov[1*](psnovichkov@lbl.gov), Serdar Turkarslan[2], Alexey E. Kazakov[1], Sarah P. Preheim[3], Marcin P. Joachimiak[1], Max Shatsky[1], Michael S. Samoilov[4], John-Marc Chandonia[1], Inna Dubchak[1], Adam P. Arkin[1], Eric J. Alm[3], **Nitin S. Baliga[2]** and **Paul D. Adams[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Institue for Systems Biology, Seattle, WA; [3]Massachusetts Institute of Technology, Cambridge, MA; [4]Department of Bioengineering, University of California Berkeley, Berkeley, CA

http://enigma.lbl.gov

Project Goals: The overarching goal of the Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) project is to understand environmentally- relevant microbial communities and processes through an integrated field-to-laboratory approach.

The Computation Component facilitates the analysis, management, and dissemination of data generated as part of ENIGMA. These activities have a reciprocal relationship with the other Core Components – They support the scientific objectives of the cores, and in turn the scientific needs drive the development of new, innovative computational tools. ENIGMA has put a strong focus on all three aspects of computation. Analysis is a strong point of the ENIGMA team, but roughly equal emphasis has put on data management, and visualization/sharing of data.

ENIGMA has developed innovative tools to make high-quality **prediction of associations between the content of microbial community and various geochemistry parameters**. These include a novel Distribution-based OTUs algorithm for grouping of individual reads into operational taxonomic units (OTUs) and the SparCC algorithm for identifying correlations within "compositional data", in which proportions rather than actual numbers are counted.

Molecular network inference continues to be a strength of ENIGMA. We have advanced and integrated the regulatory network inference algorithms into an ensemble modeling framework. This framework takes as input just gene expression data from carefully designed experiments and genomic sequence to (1) simultaneously discover environment-dependent membership of genes within co-regulated modules, (2) predict transcriptional changes within each module in new environments, and (3) predict cis- and trans-acting regulators for each module. Further, we have developed a suite of algorithms for automated reconstruction of regulons through comparative genomics

Data standardization is key to a large scale project like ENIGMA with multiple data types. The Computation Component recently completed a survey to identify and prioritize data exchange modalities for standardization. To facilitate this effort, we released a Data Management Guide that includes standards for data types (esp when produced by several labs), links to data resources, guidance for experiment planning, sample identifiers, etc.

Data dissemination both within and outside of ENIGMA continues to be a focus of our efforts. ENIGMA maintains several widely used websites, including MicrobesOnline, Network Portal, and RegPrecise, and has begun to incorporate ENIGMA data sets and algorithms into DOE Systems Biology Knowledgebase (KBase). Currently four ENIGMA computational tools are implemented as KBase services: cMonkey, Inferelator, MAK, and BAMBI. Within ENIGMA, data dissemination is critical to engage multiple off-site investigators in meaningful scientific dialog.

This work conducted 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, 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.

28. Single Cell Genomics Applications in ENIGMA

Suzan Yilmaz[1*] (syilmaz@sandia.gov), Mark Smith[2], Eric J. Alm[2], Dwayne A. Elias[3], Terry C. Hazen[3,4], Adam P. Arkin[5], Anup K. Singh[1], and **Paul D. Adams[5]**

[1]Sandia National Laboratories, Livermore, California; [2]Massachusetts Institute of Technology, Cambridge, Massachusetts; [3]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [4]University of Tennessee, Knoxville, Tennessee; [5]Lawrence Berkeley National Lab, Berkeley, California

http://enigma.lbl.gov

Project Goals: We are developing a versatile single cell genomics pipeline that can respond to the needs of a multi-institutional program like ENIGMA. Our pipeline utilizes FISH (fluorescence *in situ* hybridization) for targeting species of interest, FACS (fluorescence activated cell sorting) for high throughput isolation of single cells, and MDA (multiple displacement amplification) for production of sufficient DNA for genome sequencing. At present, this pipeline is being used for a number of collaborative projects in ENIGMA.

Single cell sequencing is a powerful tool for the analysis of uncultivated microorganisms. Current culture-independent, population based techniques (i.e., metagenomics) relying on pooled nucleic acids from communities of microorganisms can independently measure metabolic activity and the species present, but cannot link the activity deterministically to the species. In an attempt to unravel the complex dynamics of population, gene expression, and metabolic function in mixed microbial communities, we developed a high-throughput approach to study uncultivable microorganisms one cell at a time. Our approach includes isolation of individual cells by cell sorting, followed by whole genome amplification and sequencing. This pipeline is being utilized to analyze groundwater samples from DOE bioremediation sites (e.g., Hanford 100H, Oak Ridge FRC) to identify keystone organisms and link their functions to species as well as to estimate the level of horizontal gene transfer within the community; to isolate and identify viruses in deep subsurface groundwater, and investigate their role in microbial community structure and function; to assess the composition of bioaggregates in environmental samples with the ultimate goal of verifying the stereotypical configurations of microorganisms.

This work conducted 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, 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

29. Microbial Community Structure Predicts Groundwater Geochemistry

Andrea M. Rocha[1,2*] (rochaam@ornl.gov), Mark Smith[3], Christopher Smillie[3], Julian L. Fortney[1,2], Stephen M. Techtmann[1,2], Dominique Joyner[1,2], Tonia L. Mehlhorn[1], Jennifer E. Earles[1], Kenneth A. Lowe[1], Janna Phillips[1], David B. Watson[1], James H. Campbell[1], Susan Pfiffner[1,2], Kati Ayers[2], Charlie Paradis[2], Joy D. Van Nostrand[4], Liyou Wu[4], Ping Zhang[4], Zhili He[4], Jizhong Zhou[4], Michael W. W. Adams[5], Andrew Lancaster[5], **Paul D. Adams**[6], Adam P. Arkin[6], **Eric J. Alm**[3] and **Terry C. Hazen**[1,2,6]

[1] Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, [2] University of Tennessee, Knoxville, TN, [3] Massachusetts Institute of Technology, Cambridge, MA, [4]University of Oklahoma, Norman, OK, [5]University of Georgia, [6]Lawrence Berkeley National Laboratory, Berkeley, CA,

http://enigma.lbl.gov

Project goals: ENIGMA's Field Microbiology component works with microbial communities in the environment to investigate how biological structure and function relate to critical biochemical conditions/changes, such as the immobilization of toxic metals and the flow of carbon and nitrogen. The goal is to identify key microbial populations and determine the community events and mechanisms of these populations that impact and control environmental activities of interest, ultimately predicting how perturbations of the environment may affect community structure. From these results, we can develop models that can be applied to microbial populations overlaid with geochemical parameters and engineering controls. Development of such a model and key findings from a 100-well survey at the Department of Energy's Oak Ridge Field site is described below. Specifically, we demonstrate the ability to utilize microbial community assembly from independent groundwater environments to accurately predict the geochemistry and elucidate key systems biology features.

One of the primary objectives of the ENIGMA Environmental Core Field Microbiology Component is to design an efficient field sampling study that will maximize the geochemical diversity of the study site and provide for enhanced resolution of microbial communities and geochemical associations. At the Department of Energy's Oak Ridge field research site, 243---acres of contaminated area is located within the Y---12 plant area of responsibility of the Oak Ridge Reservation (ORR). Here, over 20 years of historical and published data for more than 800 groundwater wells is available in a computer queryable database. In this study, we conducted a survey of 100 groundwater wells in order to (1) characterize key microbial populations at geochemically distinct locations and (2) identify associations between environmental gradients and microbial communities. We hypothesize that differences in geochemistry underlie differences in microbial community structure across the groundwater wells. In order to identify microbe-environment associations, a statistically informed experimental design was developed by ENIGMA's computational core. To optimize geochemical diversity and identify wells where environmental factors are uncorrelated, k-medians clustering algorithm was used to group 818 wells into 100 clusters using 14 geochemically similar measurements. At each well, *in situ* groundwater measurements were recorded and unfiltered and filtered groundwater samples were collected for both geochemical measurements and analysis of microbial communities. Nucleic acids were collected by filtering 4-L of water through a 10.0µm pre-filter and 0.2µm-membrane filter and then extracted using a Modified Miller method. Community genomic DNA yields from the 0.2µm and the 10.0µm filters range from 0.096- 8.5µg and 0.096-22µg, respectively. Cell counts for the field samples ranged from 10^3-10^6 cells/mL and were consistent with biomass estimates from phospholipid fatty acid analysis (PLFA). Initial analysis of PLFA data suggests stress indicators for Gram Negative populations, such as those associated with pollutants and nutrient limitation, are present in 15% of the wells sampled. Evaluation of

divergence of microbial communities across all the wells indicates the microbial communities are fairly distinct. Comparison of microbial communities within each well shows taxa are not as divergent compared to across all wells. However, initial analysis indicates there are some organisms unique to the 10.0µm size fraction. A total of 27 wells, which can be classified into 7 groups based on pH and contaminant concentrations, were analyzed with the newest GeoChip 5.0_180K to evaluate functional differences and potentials of microbial communities under different environmental conditions. Our results indicated that the groundwater microbial community functional structure was significantly different among seven different groups of wells, and such community differences were largely correlated with the differences in pH, U(VI), nitrate, dissolved organic carbon (DOC) and sulfate in the groundwater. The microbial functional diversity significantly decreased with Ur concentrations. Metadata correlations of all the wells show many of the geochemical parameters are independent of each other. Using the synthetic learning in microbial ecology (SLiME) algorithm and the large independent dataset, we are able to predict the geochemistry from the 16S rRNA. Additionally, using the relationship between the geochemistry and microbiology, the critical OTUs that geochemistry can be identified. Overall, results from this study indicate that with careful design and a large dataset, the groundwater microbial community structure can be used to accurately predict the geochemistry. Such predictions may provide the ability of microbial monitoring for natural attenuation at legacy sites and be enabling for ENIGMA for more specialized questions on microbial community and network structure and function.

This work conducted 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, 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.

30. Evolution of alternative adaptive strategies sustaining two-member syntrophic communities

Authors: Birte Meyer[1*] (birtemey@uw.edu), Jennifer V. Kuehl[2], Jayashree Ray[2], Morgan N. Price[2], Adam M. Deutschbauer[2], Adam P. Arkin[2], Marcin Zemla[2], Amita Gorur[2], Manfred Auer[2], **David A. Stahl[1]**, and **Paul D. Adams[2]**

Author Affiliation: [1] University of Washington, Seattle, WA; [2] Lawrence Berkeley National Laboratory, Berkeley, CA

http://enigma.lbl.gov

Project Goals: The comparative analysis of simple two-member syntrophic communities composed of different pairings of *Desulfovibrio* and methanogenic species is used to identify evolutionarily conserved features of an environmentally relevant interspecies interaction. Initial studies have shown that although *Desulfovibrio* species have evolved alternative systems of electron and metabolite transfer for syntrophic growth in environments fluctuating in electron acceptor availability, they share similar energetic strategies involving flavin-/menaquinone-based electron confurcation and electron bifurcation (1). Alternative adaptive strategies are also associated with the syntrophic partner, revealed by the influence of different methanogenic species on the metabolism of an individual *Desulfovibrio*, and by inference the fitness of the pairing (2). Thus, continued comparative analyses of the genetic and biochemical systems sustaining this fundamental trophic interaction will contribute to a knowledgebase needed for a more predictive understanding of how complex microbial communities form and persist.

Abstract: The mineralization of organic matter in anoxic environments relies on the cooperative activities of hydrogen producers and consumers obligately linked by interspecies metabolite exchange in syntrophic consortia that may include sulphate reducing species such as *Desulfovibrio*. To evaluate the metabolic flexibility of syntrophic *Desulfovibrio* to adapt to naturally fluctuating methanogenic environments, we studied *Desulfovibrio alaskensis* str. G20 grown in chemostats under respiratory, fermentative and syntrophic conditions with alternative methanogenic partners, *Methanococcus maripaludis* and *Methanospirillum hungatei*, at different growth rates on varying energy sources. Comparative analyses of whole genome transcriptional and gene fitness (tagged transposon mutant library) data, complemented by individual G20 mutant strain growth experiments, and physiological data, revealed a significant influence of (a) energy source, (b) availability of electron donor (as controlled by dilution rate), and (c) methanogenic species on the electron transfer enzyme systems, mechanisms of energy-conservation, ratios of interspecies electron carriers, coculture population dynamics, and interspecies physical association. All data indicate that *D. alaskensis* str. G20 use both flavin- and menaquinone-based electron confurcation and bifurcation processes to drive the production of metabolites (H_2 and formate) sustaining its syntrophic association with a methanogen. During syntrophic growth on lactate, a reduced thiol/disulfide redox pair (most likely DsrC) and ferredoxin (Fd) are energetically coupled to H^+/CO_2 reduction by periplasmic formate dehydrogenase and hydrogenase via a flavin-based electron confurcation process and a menaquinone (MQ) redox loop-mediated reverse electron flow involving the membrane-bound Qmo and Qrc complexes. In contrast, *D. vulgaris* str. Hildenborough uses a larger number of cytoplasmic and periplasmic proteins linked in three intertwining pathways to couple $DsrC_{red}$ and Fd_{red} reoxidation to H^+ reduction during lactate oxidation. The faster growth of strain G20 in coculture is associated with a kinetic advantage conferred by the Qmo-MQ-Qrc loop as electron transfer system that permits higher lactate utilization rates under elevated hydrogen levels (thereby enhancing methanogenic growth), and use of formate as main electron exchange mediator (>70% electron flux), as opposed to the primarily hydrogen-based

exchange by strain Hildenborough. Although the collected data support the absence of a conserved gene core in *Desulfovibrio* that would determine the ability for syntrophic lifestyle in sulfate-reducing bacteria, systems of flavin-/menaquinone-based electron confurcation or electron bifurcation are common to both species. Remarkably, only 68 genes in *D. alaskensis* str. G20 were commonly differentially expressed under syntrophic versus respiratory lifestyle which points to its high metabolic flexibility to adjust energetically to the naturally fluctuating growth conditions in methanogenic environments. Under low energy (low growth rate) conditions, strain G20 further adapts to the metabolism of its methanogenic partners as shown by the differing gene expression of enzymes involved in the direct metabolic interactions (e.g. periplasmic hydrogenases), and the ratio shift in electron carriers used for interspecies metabolite exchange (H_2/formate). A putative monomeric [Fe-Fe] hydrogenase and Hmc complex-linked reverse MQ redox loop become increasingly important for the reoxidation of the lactate oxidation derived redox pairs, $DsrC_{red}$ and Fd_{red}, relative to the Qmo-MQ- Qrc loop. The lower growth rates also promoted close physical interspecies polar associations, presumably enabling more efficient metabolite transfer and more energy-efficient energy coupling (a similar effect was observed for strain Hildenborough cocultures). Transition from lactate to pyruvate in *D. alaskensis* str. G20 cocultures resulted in a dramatic shift in the population structure and even closer interspecies cell-to-cell interactions. Lower methane production rates in coculture than predicted from pyruvate input was attributed to redirection of electron flow to fumarate reduction. Without a methanogenic partner, accumulation of H_2 and formate resulted in greater succinate production indicating that pyruvate fermentation in strain G20 involves respiration of endogenously formed fumarate using cytoplasmic and membrane-bound energy-conserving complexes, Rnf, Hdr-Flox-1, and Hmc. At the low H_2/formate levels maintained in coculture, Rnf likely functions as proton-pumping Fd_{red}:type-I-cytochrome-c_3 oxidoreductase which transitions to a proton-pumping Fd_{red}:NADH oxidoreductase at high H_2/formate levels during fermentation in monoculture. Hdr-Flox-1 is postulated to recycle Fd_{red} via a flavin-based electron bifurcation involving NADH, Fd_{ox}, and $DsrC_{ox}$. In a menaquinone-based electron confurcation reaction, the Hmc complex is proposed to then couple $DsrC_{red}$ and periplasmic H_2/formate oxidation using the menaquinone pool to fuel a membrane-bound fumarate reductase (3).

Together these data underscore the high metabolic and energetic adaptive flexibility that likely sustains *Desulfovibrio* in naturally fluctuating methanogenic environments. These laboratory-based studies provide an important mechanistic understanding of the assembly and stability of two-member model assemblies in nature that will provide a predictive understanding of microbial processes stabilizing or destabilizing critical communities of microorganisms in the environment.

References:

1. Meyer, B., Kuehl, J., Deutschbauer, A. M., Price, M. N., Arkin, A. P., and Stahl, D. A. 2013. Variation among *Desulfovibrio* species in electron transfer systems used for syntrophic growth. Journal of Bacteriology 195:990-1004.
2. **Meyer, B., Kuehl, J. V., Deutschbauer, A. M., Arkin, A. P., and Stahl, D. A.** 2013. Flexibility of syntrophic enzyme systems in *Desulfovibrio* species ensures their adaptation capability to environmental changes. Journal of Bacteriology **195:**4900-4914.
3. **Meyer, B., Kuehl, J. V., Price, M. N., Ray, J., Deutschbauer, A. M., Arkin, A. P., and Stahl, D. A.** The energy-conserving electron transfer system used by *Desulfovibrio alaskensis* strain G20 during pyruvate fermentation involves reduction of endogenously formed fumarate and cytoplasmic and membrane-bound complexes, Hdr-Flox and Rnf. Environmental Microbiology **accepted for publication**.

Funding Statement: *This work conducted 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, 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.

31. High Throughput Isolation and Environmental Isolate Characterization in ENIGMA

Romy Chakraborty* (rchakraborty@lbl.gov), Angelica Pettenato, Marcus Schicklberger, Adam Deutschbauer, Jennifer Kuehl, Jayashree Ray, **Adam P. Arkin** and **Paul D. Adams**

Lawrence Berkeley National Laboratory, Berkeley, CA

http://enigma.lbl.gov

Project Goal: The goal of ENIGMA's Environmental Isolation and Characterization Campaign is to answer outstanding questions in microbial ecology that are best addressed with a combination of microbial isolation and downstream functional characterization of those isolates. The specific goal of the Isolation project within this campaign is to develop high throughput isolation platform for cultivating and subsequently investigating microbes performing metabolic processes of interest at the ENIGMA field site, Oakridge Field Research Center.

The constraints of obtaining isolates from key environments have undermined testing and validating hypotheses that emerge from powerful gene-based technologies. Moreover, how individual microbes function, predictions about how they might function, and the identification of genes and pathways critical for fitness under a given condition can be best answered by lab- based controlled experiments with the representative microbial isolates from the field. Towards this direction, we are developing and testing different approaches to obtain isolates from Oakridge FRC, which contains high plumes of uranium, technetium, nitrate, volatile organic compounds and has a pH gradient from 3-10.

Our first approach uses microtiter plates for high throughput, and aims to recover a large number of diverse culturable aerobic and anaerobic strains across multiplexed conditions of soluble electron acceptors, pH, salinity and metals among others. As a result, several hundred strains have been isolated from selected groundwater wells across geochemical gradients, including representatives from those genera identified to be most abundant by the ENIGMA 100 Well 16S- survey Campaign. These were cultivated using a combined approach of both minimal and rich complex media. Select isolates are being characterized in details as part of the ENIGMA Printable Worlds and Environmental Characterization campaigns.

Our second approach is directed towards specific enrichments and isolation of strains that metabolize unique substrates and/or catalyze novel metabolisms. As part of this, we have identified strains capable of oxidizing dissolved organic matter (humics), and metals such as Fe coupled to nitrate reduction. Some of these isolates are denitrifiers, while others reduce nitrate to ammonia. We have also focused on isolating specific phylogenetic groups of interest. Several *Pseudomonas spp* have been isolated from wells with different nitrate content, and although closely related by 16S rRNA sequence, they vary greatly in their ability to reduce nitrate ranging from 10mM to 300mM. We are investigating the role of metals (such as Mo, Cu, Fe) as limiting factors on their ability to reduce nitrate in monocultures and co-cultures, and performing pan- genome analysis as part of ENIGMA Metals and Environmental Characterization Campaigns.

This work conducted 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, 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.

32. Microbial Characterization in ENIGMA: Flexible Tools for Annotating Genes and Pathways in Environmental Bacteria

Adam Deutschbauer[1*] (AMDeutschbauer@lbl.gov), Morgan N. Price[1], Kelly M. Wetmore[1], Jennifer V. Kuehl[1], Jayashree Ray[1], Valentine Trotter[1], Avneesh Saini[1], Jennifer He[1], Julijana Ivanisevic[2], Samuel R. Fels[3], Grant M. Zane[3], Gary Siuzdak[2], Trent Northen[1], Gareth Butland[1], Romy Chakraborty[1], Judy D. Wall[3], **Adam P. Arkin[1]** and **Paul D. Adams[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]The Scripps Research Institute, San Diego, CA; [3]University of Missouri, Columbia, MO

http://enigma.lbl.gov

Project Goals: The Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) program broadly seeks to understand the interactions between environmentally relevant microorganisms and their environment. One aim of this large interdisciplinary project is to bring environmental bacteria to model- organism status rapidly to enable systems-level investigations into microbial metabolism, regulation, stress response, and interactions under defined laboratory conditions. In addition to developing scalable, multi-scale, and high-throughput characterization techniques, the ENIGMA Microbial Characterization campaign aims to apply these techniques to answer outstanding questions in microbial ecology.

Our understanding of the roles of microbes in important natural processes is hindered by the lack of tools to study their function at the molecular level. A scarcity of functionally well-characterized microbes means that the genomes of many microorganisms, including those with important roles at the Oak Ridge FRC site, remain poorly annotated. To meet this challenge, the ENIGMA Microbial Characterization campaign is developing and applying methods for the flexible and inexpensive characterization of virtually any culturable prokaryote. Here, we present an overview of our technical developments and how we are using these approaches to address ENIGMA science.

As mutant phenotypes provide insight into gene function by providing a direct link between genes and phenotypes, we have developed a next-generation strategy, termed random barcode transposon insertion sequencing (RB-TnSeq), for assaying the mutant fitness of thousands of strains in parallel. RB-TnSeq takes advantage of two recent ENIGMA advances, transposon liquid enrichment sequencing (TnLE-seq; PMID 24077707) for the generation of mutant libraries and DNA barcode sequencing (BarSeq) for assaying mutant fitness. To demonstrate the reproducibility and scalability of RB- TnSeq, we present data from over 200 experiments in the denitrifying environmental isolate *Pseudomonas stutzeri* RCH2. As many single gene mutations do not result in a strong phenotype under laboratory conditions, we are also developing a flexible strategy for screening genetic interactions (double mutants), using the sulfate-reducing bacterium (SRB) *Desulfovibrio vulgaris* Hildenborough as a pilot. Lastly, to complement the high-throughput genetics data, we are pairing mutant fitness assays with untargeted metabolomics to link genes to specific metabolites to physiological role. Using this combined genetics and metabolomics approach, we have identified a novel metabolite (and the encoding genes) required for the SRB *Desulfovibrio alaskensis* G20 to grow under salt stress.

As part of the Microbial Characterization campaign, we are using the developed genetic and metabolomics approaches to address two hypotheses: (1) The function of many genes can only be assessed under conditions that closely mimic the natural environments from which the microorganisms are isolated, and (2) Core species are present across several Oak Ridge FRC sites, regardless of geochemistry, and the

adaptations of these ubiquitous clades to diverse environments are due to differences in gene content, protein activity, novel metabolic pathways, and gene regulation. In addition, we are investigating the molecular basis of key microbial metabolisms at the FRC site, with a current focus on the oxidation of dissolved organic matter (humics). Lastly, the data, tools, and genetic resources generated by the Microbial Characterization campaign are being used in a number of collaborative ENIGMA projects to investigate metal metabolism, microbial interactions, and gene regulatory networks.

This work conducted 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, 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.

33. Does Scale Impact Structure and Function of Microbial Biofilms?

Matthew W. Fields[1] (matthew.fields@biofilm.montana.edu), Sara Altenburg[1], Anna Zelaya[1], Suzan Yilmaz[2], A. K. Singh[2], **Adam P. Arkin[3]**, and **Paul D. Adams[3]**

[1]Montana State University, Bozeman, MT; [2]Sandia National Laboratory, Livermore, CA;
[3]Lawrence Berkeley National Laboratory, Berkeley, CA

http://enigma.lbl.gov

Project Goals: The overarching goal of the Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) is to understand environmentally-relevant microbial community structure and function through a series of integrated field-to-laboratory campaigns. The described project is designed as an interdisciplinary platform to elucidate the fundamental behavior of microbial biofilms in response to physical scale of the particles to which microbial biofilms interact. Ultimately, we propose to elucidate physical and chemical principals that contribute to the assembly, evolution, and stability of microbial communities.

Sulfate-reducing-bacteria (SRB) occur naturally in a variety of anaerobic environments where sediments are present. In order to investigate the impact of physical surface scale on microbial interactions occurring in anaerobic habitats, attempts were made to standardize the growth of *Desulfovibrio* biofilm on various particle sizes using modified biofilm reactors. The standard coupon holders were modified to contain a mass of particles with continuous access to nutrients and *Desulfovibrio* culture, thus providing a surface for biofilm formation that could be easily removed at the end of the study period. By investigating at this finer resolution, patterns in microbial community structure and composition may be more discernable. The paradigm of "Everything is everywhere" has been commonly invoked, but obviously at a small enough scale 'everything' cannot be 'everywhere'. Does scale of analysis impact observation and interpretation for microbial communities? The reactor systems have used environmental isolates, *Desulfovibrio* RCH1 (Hanford) and *Desulfovibrio* FW1012B (Oak Ridge) to characterize growth on glass beads (30 μm, 425 μm and 3,000 μm). The surface area to volume ratio decreased with increasing bead size, and ranged from 1500 to 58 to 20 cm^{-1}, respectively. The biofilm protein per surface area ($\mu g/cm^2$) was 25-fold and 50-fold greater for the intermediate and largest sized particles, respectively, compared to the smallest. A similar trend was observed for biofilm carbohydrate (17- and 40-fold increased) compared to the smallest bead size. However, the overall biofilm carbohydrate to protein ratio was similar for the tested particle sizes (0.11, 0.07, 0.09, respectively). For the particle sizes tested, the amount of biofilm per unit area decreased with the particle size even as surface area/volume increased. One possible explanation for the counter-intuitive result is that the growth of biofilm significantly alters the porous structure and consequently changes porosity, permeability and dispersivity of the substratum. Because the tested particle sizes are significantly larger than the dimension of cells, we propose that initial colonization of beads is not mass transfer limited. However, as the biofilm formation proceeds on the beads and in the inter-bead space, the porosity of the packed bed changes thereby affecting the kinetics of biofilm growth. The changes are more pronounced for smaller particles and can also lead to heterogenous distribution and/or function of microbial populations. Future analyses include per bead measurements, biomass growth kinetics for different size of particles, and impacts on local diversity in comparison to source diversity.

This work conducted 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, 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.

34. Identification of Iron Reductases Using Top-down Proteomics and Heterologous Expression

Ornella D Nelson[1*] (odn5@cornell.edu), Zhi Li[1], Anne E Otwell[1], Robert W Sherwood[1], Sheng Zhang[1], Hening Lin[1], Stephen J Callister[2], **Ruth E Richardson[1]**

[1]Cornell University, Ithaca, NY; [2]Pacific Northwest National Laboratory, Richland, WA

Project Goals: The goal of this project is to identify and analyze novel iron reductases in three microorganisms, *Geobacter sulfurreducens PCA, Desulfotomaculum reducens MI-1* and *Anaeromyxobacter dehalogenans 2CP-1,* using a combination of non-denaturing separation, proteomics-based discovery and heterologous expression to confirm and further characterize enzyme function.

Identification and analysis of enzymes involved in heavy metal reduction by microorganisms can provide better mechanistic understanding as well as improve bioremediation techniques for heavy metal contamination. In this project, novel iron reductases in three microorganisms were identified by using a combination of non-denaturing protein separation, functional screens, proteomics-based discovery and heterologous expression to confirm and further characterize enzyme function. *Geobacter sulfurreducens PCA, Desulfotomaculum reducens MI-1* and *Anaeromyxobacter dehalogenans 2CP-1* were each grown anaerobically and the cells were collected and lysed. The soluble and membrane-bound protein fractions were separated for individual analysis. The proteins were separated using strong anion exchange (SAX) chromatography, size exclusion (SEC) chromatography, and native gel electrophoresis coupled with solution-phase and in-gel iron reduction assays. Protein bands displaying iron- reduction activity in the in-gel activity assay were excised for protein digestion and peptide identification by mass spectrometry. Proteins identified were over-expressed in *E.coli* and purified by metal-affinity chromatography for characterization studies, including *in vitro* iron reduction activity. Thus far, a total of six iron reducing proteins/protein complexes have been confirmed in these three organisms. These proteins/complexes were NADPH-dependent enal/enone/nitroreductase from *G. sulfurreducens,* Oxidoreductase FAD/NAD(P)-binding subunit/Dihydroorotate dehydrogenase 1B complex and NADH: flavin oxidoreductase from *D. reducens* as well as Pyruvate flavodoxin/ferredoxin oxidoreductase homologs in each microorganism.

Funding for this project was provided by the Department of Energy's Genomic Sciences Program within the Office of Biological and Environmental Research

35. Microbial Food Web Mapping: Linking carbon cycling and community structure in soils through next generation sequencing enabled stable isotope probing

Ashley N. Campbell*, Sean T. Berthrong, Chantal Koechli, Charles Pepe-Ranney, and **Daniel H. Buckley.** Cornell University, Ithaca, NY

Project Goals: This research program explores fundamental aspects of carbon cycling in soil microbial communities. Major goals are to develop and validate methods for next generation sequencing enabled stable isotope probing (NGS-SIP) and to use this approach to dissect the microbial food web in soil. NGS-SIP offers a means to study the microorganisms that facilitate soil processes as they occur in soil, to characterize novel organisms that have escaped detection previously, and to make significant advances in our understanding of the biological principles that drive soil processes. With this approach we will examine connections between microbial community composition and soil carbon cycle dynamics. Specific objectives include 1) determine whether carbon input parameters (composition, quantity, timing of carbon additions) alter the route of carbon through the soil community, 2) determine whether these shifts interact with respect to microbial community structure, and 3) evaluate whether microbial community structure is functionally equivalent across edaphically similar soils that differ in management history.

The terrestrial biosphere contains a large fraction of global C and nearly 70% of the organic C in these systems is found in soils. Much of the organic C in soils is respired and on an annual basis soil respiration produces 10 times more CO_2 than anthropogenic emissions, but it remains difficult to predict the response of soil processes to anthropogenic changes in the environment. Our difficulty in predicting how soil processes will respond to environmental change suggests a need for a greater understanding of the biotic mechanisms that govern the soil C-cycle. It is important to examine the internal dynamics of soil microbial communities, and the manner in which they influence community function, in order to understand the how the terrestrial C-cycle responds to environmental change. The NGS-SIP approach that we are developing will allow for pulse chase style experiments that allow 13C-isotopes to be tracked through the soil community over time. The approach will involve the application of synthetic biomass containing a mix of carbon sources designed to approximate the plant biomass. The use of synthetic biomass allows substitution of 13C-labeled substrates into the mixture to track the manner in which different types of C are metabolized by different components of the community.

Experiments have explored the metabolism of cellulose (an insoluble polymer) and xylose (a soluble sugar monomer resulting from breakdown of hemicellulose) by soil communities over time. The NGS-SIP approach has provided data for more than 6,000 bacterial taxa. Soil C amendment caused change in community composition over time but only a subset of microorganisms in the community assimilated 13C from xylose or cellulose. Xylose incorporation into DNA occurred by day 1 but could no longer be detected by day 14, indicating biomass turnover. Cellulose degradation proceeded more slowly with 21% of cellulose C respired by day 7 and 60% respired by day 30. Incorporation of cellulose-C into DNA was observed only after 14 days. DNA incorporation of 13C from xylose suggest partial labeling which indicates assimilation of C from sources other than xylose. In contrast, taxa which incorporated 13C from cellulose were highly labeled indicating assimilation of C almost entirely from cellulose. Dominant taxa that assimilated C from xylose include *Arthrobacter*, *Agromyces*, *Rhizobium*, and *Paracoccus*, among others. Dominant taxa that assimilated C from cellulose include *Cellvibrio*, a novel non-cultivated lineage of *Chloroflexi*, and *Verrucomicrobia* among others. These data support the succession hypothesis of decomposition in that

sugars are degraded by fast growing opportunistic organisms and insoluble polymers are degraded more slowly by polymer degrading specialists. In addition, the results have identified novel taxa that mediate cellulose deconstruction in soil.

Several follow up experiments have been performed to explore the dynamics of C assimilation. In particular, we have performed parallel analysis of rRNA to determine if assimilation of 13C into rRNA provides additional insights not provided by DNA analysis. We have also performed parallel analysis of fungal taxa that incorporate 13C from cellulose. Finally, we have performed metagenomic analysis of NGS-SIP gradient fractions to identify genome fragments from bacteria that have assimilated 13C from 13C-cellulose. Analysis of rRNA indicates that rRNA is labeled more quickly than DNA in soil. In particular, 23% of community rRNA is labeled at day 7 while only 2% of community DNA has been similarly labeled. At day 14, 54% of community rRNA is labeled while only 6% of community DNA is labeled. The dramatic difference between the extent of rRNA and DNA labeling on a weekly timescale suggests a phenomenon other than metabolic shift up must be invoked to explain this result. We are currently evaluating the hypothesis that uncoupled growth dynamics results in 13C-cellulose degradation and incorporation of 13C into rRNA by cells that are not actively dividing and replicating DNA. This hypothesis makes the prediction that 'uncoupled taxa' will show labelling of rRNA but not DNA whereas replicating taxa will demonstrate labelling of both rRNA and DNA. If verified this hypothesis would reveal differences in ecological strategies between microbial taxa with implication on C use efficiency and C fate in soil.

In a second series of NGS-SIP experiments we have evaluated the effect of priming on cellulose decomposition in soils. The addition of labile C to soils has been shown to alter decomposition dynamics with the manner of application resulting in different responses. In particular, small additions of labile C added over time have been observed to have a larger impact on decomposition than the equivalent addition of labile C made in a single dose. This phenomenon has been hypothesized to relate to the ability of plant roots to prime organic matter mineralization by exuding labile C to stimulate microbial activity and thereby facilitate access to associated macro and micronutrients. We performed an NGS-SIP experiment with 13C-cellulose and priming doses of glucose to determine if the microbial community mediates the altered decomposition dynamics observed in response to different priming regimes. Analysis of these experiments is ongoing.

Finally, NGS-SIP results are being extended to evaluate the distribution of xylose and cellulose responsive taxa across a series of agricultural plots that vary in organic matter management history over a span of 50 years. Changes in management practice associated with crop rotation have resulted in a gradient of soil organic matter content. We have shown that soil organic matter content is the primary factor explaining differences in microbial community composition across these plots. We hypothesize that generalist taxa are favored in intensively managed low organic matter sites while cellulose specialists will be favored in sites where soil organic matter accumulates in response to plant biomass inputs.

36. Development of Quantitative Protein Biomarker Assays for Enzymes Involved in Bacterial Iron and Uranium Reduction

Ruth E Richardson [1*] (rer26@cornell.edu), Anne E Otwell[1], Alex Hernsdorf[1], Stephen J Callister[2], Sheng Zheng[1], Tom Fillmore[2])

[1]Cornell University, Ithaca, NY; [2]Pacific Northwest National Labs, Richland, Washington

Project Goals: The overall goal of this project is to develop quantitative MS/MS based assays for peptides specific to iron and uranium reducing proteins. The suite of peptides to be monitored includes representatives from proteins across a wide range of iron and uranium reducers. Synthetic, isotopically labeled peptide standards will be used for assay validation and to enable absolute quantification of target proteins in laboratory cultures as well as field samples from uranium contaminated groundwater.

Molecular biological tools hold promise for monitoring *in situ* remediation. While DNA/RNA based MBTs are becoming widespread, proteins MBTs are not as common. However, protein biomarker assays directly detect (and ideally quantify) the catalysts and therefore have a more direct relationship to instantaneous *in situ* capabilities. Our objectives in this study are to develop quantitative, mass spectrometry-based assays that will allow absolute and specific detection of a suite of iron and uranium reductases in pure culture proteomes and in metaproteomes from environmental samples.

In this work, we combine comparative genomics and high-throughput proteomic data to identify a collection of proteotypic peptides (PTPs) for proteins involved in iron and uranium reduction in six phylogenetically diverse bacteria: *Geobacter sulfurreducens, Geobacter bemidjiensis, Shewanella oneidensis, Anaeromyxobacter dehalogenans, Desulfovibrio desulfuricans,* and *Desulfotomaculum reducens*. Proteins include those previously identified in the literature as well as those discovered by our project. The list of proteins includes several cytochromes, pili proteins, membrane-bound oxidoreductases, and cytoplasmic iron reductases.

The selected method for quantification of proteins of interest is by multiple reaction monitoring (MRM) of PTPs via tandem mass spectrometry. Two to three PTPs per protein of interest are being selected for MRM based upon several criteria: previous detection in shotgun proteomic surveys of trypsin-digested proteomes, conservation across orthologs of the protein in other species, uniqueness to the protein of interest or its orthologs, and absence of methionine residues. Additionally, preference is given to PTPs that are located at the active site (if known) of the protein and have relatively large parent ion intensities during shotgun proteome characterizations relative to other PTPs in the proteins. Isotopically labeled synthetic peptides will be used to validate the MRM assays and to enable development of quantitative standard curves for multiple transitions (parent ion/fragment ion pairs) for each peptide. Assays will be used on proteomes from laboratory cultures as well as metaproteomes recovered from aquifers at uranium contaminated field sites.

Funding for this project was provided by the Department of Energy's Genomic Sciences Program within the Office of Biological and Environmental Research

37. Investigations into the Proteome of Gram-positive Metal-reducing Bacterium *Desulfotomaculum reducens* MI-1

Anne E Otwell[1*] (aeo28@cornell.edu), Robert W Sherwood[1], Sheng Zhang[1], Ornella D Nelson[1], Hening Lin[1], Stephen J Callister[2], **Ruth E Richardson[1]**

[1]Cornell University, Ithaca, NY; [2]Pacific Northwest National Laboratory, Richland, WA

Project Goals: The major goal of this project is to improve characterization of iron and sulfate respiration in *Desulfotomaculum reducens* MI-1. Towards this goal, other aims of the project included creating and optimizing a top-down proteomic workflow involving nondenaturing separations in order to screen for iron reduction active proteins and then heterologously expressing and characterizing identified proteins. Additional aims included selecting proteins of interest for roles in sulfate or iron respiration in *D. reducens* based on comparative proteomic analyses across different growth conditions.

Microbial respirations based on iron and sulfate occur readily in subsurface environments and have important impacts on heavy metal and radionuclide contaminants. *Desulfotomaculum reducens* MI-1 provides a unique study system as a Gram-positive sulfate reducing bacteria (SRB) that is also capable of respiring a variety of metals. Predictions of sulfate reduction pathways in *D. reducens* currently rely on homology to model Gram-negative SRB. Furthermore, no metal reductases have been described in this organism, and no orthologs to characterized metal reductases from other organisms exist in the *D. reducens* genome. In this project, we utilized proteomic-based techniques in order to investigate iron and sulfate reduction in *D. reducens*. We optimized a top-down proteomic approach based on a workflow of multidimensional protein complex separation followed by iron reduction activity assays and implemented it in order to identify proteins capable of iron reduction from the proteome of *D. reducens*. These proteins have been heterologously expressed and validated as iron reductases in purified form. In addition, we have performed comparative proteomic analyses on the proteome of *D. reducens* across different growth conditions. This bottom-up proteomic technique has highlighted proteins of interest based on differential abundance and has led to predictions of proteins involved in the respiration of iron and sulfate in *D. reducens*. Targets for heterologous expression and characterization have also been selected based on these comparative proteomic analyses.

Funding for this project was provided by the Department of Energy's Genomic Sciences Program within the Office of Biological and Environmental Research

38. Cross-system analysis of carbon assimilation dynamics in soil microbial communities: Documenting the function of non-cultivated microorganisms in terrestrial ecosystems

Daniel H. Buckley*, Sean T. Berthrong, Ashley N. Campbell, Chantal Koechli, Charles Pepe- Ranney, and Nick Youngblut**.** Cornell University, Ithaca, NY

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

Global changes in atmospheric CO_2, temperature, precipitation, and ecosystem N inputs, are expected to impact primary production and carbon inputs to soils, but it remains difficult to predict the response of soil processes to anthropogenic change. Models that predict soil C- cycling as a function of ecosystem properties do not explain well variation in soil processes. Our difficulty in predicting the response of soil processes to environmental change suggests a need for a greater understanding of the biotic mechanisms that govern the soil C-cycle. Changes in microbial community structure and function have been proposed to impact soil C-cycling both qualitatively and quantitatively. However, our ability to predict the impacts of these changes on terrestrial ecosystems is constrained by our limited understanding of mechanisms that drive microbial processes in soil systems.

A fundamental limitation in understanding microbial C-cycling in soils is that we have a glaring lack of information about which microorganisms actually mediate critical soil processes and whether their activities vary across ecosystems. This research project will address this fundamental problem by using a suite of experiments to map C-assimilation dynamics for thousands of discrete microbial taxa across dimensions of both seasonal and edaphic variation.

First, we plan to deploy a full cycle food web mapping experiment in contrasting soils to simultaneously track the assimilation of eleven ^{13}C-labeled compounds into soil microorganisms. These compounds have been selected to represent dominant components of plant biomass (cellulose, hemicellulose, lignin) and intermediate products of plant biomass decomposition (glucose, xylose, benzoic acid, glycerol, palmitate, amino acids, lactate, and oxalate). By tracking the assimilation of these isotopically labeled compounds in parallel and in relation to appropriate controls it will be possible to map routes of plant biomass C metabolism by soil communities and identify the specific contribution of thousands of microorganisms to this process. This approach will be employed in contrasting ecosystems including a cultivated agricultural site and a primary forest site. These sites are located in the same geographic region with soils of the same soil order but have very different soil C content representative of their ecosystem types. This effort will reveal the contribution of thousands of uncultivated microorganisms in the soil C-cycle. We will identify the metabolic

capabilities of non-cultivated microorganisms in soil and explore how their activity varies across contrasting ecosystems. Furthermore, we will explore differences in how microbial communities process plant biomass C across these ecosystem types.

Second, we plan to deploy a series of streamlined food web mapping experiments to track cellulose and xylose (the major product of hemicellulose degradation) assimilation by soil communities as a function of soil C content and pH. While both soil C and pH are major drivers of soil microbial community structure it remains unknown whether and how these factors alter soil microbial C-cycling. We will evaluate how these factors impact the C-assimilation dynamics of soil communities across a series of sites. Sites will include conventionally managed cultivated fields, fields managed organically for 10 or 20 years, non-cultivated pasture, and secondary and primary forest sites. These 6 sites represent a gradient of C content and represent the range of soil C commonly found in much of the Northeastern USA. In addition, the impact of pH will be assessed in each of three sites (acidic, slightly acid, neutral to alkaline soils) representing two types of systems: meadow and primary forest. We will determine how these critical soil parameters alter the route that C takes through the soil community and we will identify microorganisms whose activity varies with respect to soil C status, pH, and ecosystem type.

Finally, we will characterize the seasonal dynamics of soil microbial communities for replicate plots at all field sites. SSU rRNA (RNA) and SSU rRNA gene (DNA) sequencing will be used to assess changes in community composition over time and across sites. Food web mapping experiments from objectives 1 and 2 will yield information on C metabolism for discrete taxa across different soils and ecosystem types, but this approach is not amenable to the sampling density required to analyze microbial communities over time. Hence, C assimilation dynamics of discrete taxa identified in objectives 1 and 2 will be evaluated as a function their SSU rRNA and SSU rRNA gene relative abundance over space and time for all field sites. This combined approach will make it possible to build and test hypotheses relating C-assimilation by discrete microbial taxa to their ecological distribution in space and time. We will test species-time relationships for microbial taxa as a function of correspondence in their C-assimilation profiles.

The genetic capacity of microbial communities can be studied through ‘omic approaches but it remains difficult to make direct links between the genetic capacity of microorganisms and their function in the soil C-cycle. Microbial Food Web Mapping through next generation sequencing enabled stable isotope probing (NGS-SIP) makes it possible to link gene sequences to soil C- cycle processes as they occur in soil. This approach allows us to characterize the activity of non- cultivated microorganisms in a range of terrestrial systems. This data will be used to build a base of information about the role of non-cultivated organisms in critical C-cycle processes in terrestrial ecosystems and will provide insight on the manner in which soil communities metabolize soil organic matter. This research will yield fundamentally new insights into the role of soil communities as a driver of soil C-cycling in soils, revealing information about how individual microbial taxa contribute to variation in community function.

39. Microbial Communities and Carbon Cycling in California Coastal Wetlands

Susannah G. Tringe[1*] (**sgtringe@lbl.gov**), Shaomei He,[1] Susanna Theroux,[1] Lisamarie Windham-Myers[2] and Wyatt Hartman[1]

[1]DOE Joint Genome Institute, Walnut Creek, California; [2]US Geological Survey, Menlo Park, California

Project goals: To understand the role of belowground microbial communities in plant growth and nutrient cycling in wetlands, particularly their potential to enhance or inhibit effective carbon retention. This includes characterizing the organisms, genes and pathways present and active in wetland soils via nucleic acid sequencing and correlating them with biogeochemical features and carbon cycling metrics.

Wetlands are an important global carbon reservoir and carbon sink. While wetland carbon uptake is largely due to plant primary productivity, the overall greenhouse gas budget is heavily dependent on microbial activities such as plant biomass degradation and methanogenesis.

Wetland restoration efforts on the deeply subsided peat islands in the Sacramento / San Joaquin river delta, currently used primarily for agriculture, have been demonstrated to increase land elevation and sequester atmospheric carbon. We are using DNA and RNA sequencing to study belowground microbial communities in these restored wetlands to assess their composition, functional potential and impact on carbon cycling. The wetland soils harbor diverse communities of bacteria, archaea and fungi whose membership varies with sampling location, proximity to plant roots, and sampling depth in patterns that closely correlate with gradients of electron acceptor availability and methane production. Shotgun metagenome sequencing revealed complementary patterns in functional gene distribution, including apparent competition between methanogens and other anaerobic guilds. Expanded studies throughout the San Francisco Bay/Delta region are enabling us to correlate microbial community composition and greenhouse gas cycling with key environmental variables including salinity, soil carbon and duration of flooding.

This project was funded by DOE Early Career Research Program, grant number KP/CH57/1, and was also supported by DOE JGI Community Sequencing Program. The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

40. Composition and Distribution of Core Carbohydrate-Active Microbial Genes in Biofuel Soils

K. Hofmockel[1]* (khof@iastate.edu), A. Howe,[2,3] R. Williams,[1] F. Yang,[1] E. Bach,[1] F. Meyer[2] [1] Iowa State University, Ames, Iowa; [2] Mathematics and Computer Science Division and Institute for Genomics and Systems Biology, Argonne National Laboratory; Michigan State University, East Lansing, MI.

Project Goals: To identify the key organisms and functions involved in the metabolism of cellulose within soil communities and determine if the active carbon cycling microbial communities differ among soil aggregate fractions and bioenergy cropping systems.

Understanding how plant communities and management systems influence the composition and function of soil microbial communities is paramount for reducing environmental impacts associated with bioenergy agriculture. Because soil microbial communities are important drivers of carbon and nitrogen cycling, the integration of microbial metabolism into biogeochemical models is an important next step for accurate prediction of ecosystem responses to land use change and climate feedbacks. A key challenge for integrating microbial ecology into models is that soil contains highly complex and diverse microbial communities in a spatially heterogeneous structure. Soil fractionation techniques provide an opportunity to examine intact microbial communities in a context that is relevant to both microbial community metabolism and ecosystem processes. The dynamic hierarchy of aggregates of different sizes creates intra-aggregate pore spaces that are the habitats in which microbes live. By considering soil aggregates as the fundamental units of microbial assemblages, soil heterogeneity can be reduced, and microbial communities investigated in systems that reflect natural communities and at scales consistent with micron-scale processes. Our research aims to use soil aggregates to develop laboratory and field approaches that target metabolically active microorganisms and functions that drive carbon cycling in soils from bioenergy cropping systems.

As a first step we have identified core carbohydrate-active genes in multiple localized samples (n=4) of a fertilized prairie used for biomass feedstock production. The core metagenome from whole soil samples were compared to soil aggregate fractions (n=20) from the same site, as well as in soils from adjacent plots of both unfertilized prairie and corn. Among a total of 226,998 genes with similarity to known carbohydrate genes, 911 genes were common to all samples, encompassing our defined fertilized prairie carbohydrate-active core metagenome.

Within the carbohydrate-active core, the distribution of enzyme classes and their taxonomic origin were characterized. Abundant glycosyltransferases were found in Proteobacteria, Bacterioidetes, and Firmicutes; glycoside hydrolases were found in Fungi and Proteobacteria; and carbohydrate esterases in Proteobacteria, Chloroflexi, and Fungi. The core genes were present in the metagenome at a broad range of abundances, estimated from 1 to 6 copies of a carbohydrate gene per 100 cells. Many core fertilized prairie genes were also found to be present in multiple other agricultural and grassland metagenomes, with decreasing presence in forest, desert, and tundra metagenomes. Comparing core carbohydrate-actives with a database of known genomes from sequenced soil isolates revealed that current references lack information on several core phyla, especially Chlorobi, Spirochaetes, and Korachaeota. In soil ecosystems, where high diversity remains to be a key challenge for metagenomic investigations, these core genes represent a subset of critical functions necessary for carbohydrate metabolism, which can be targeted to compare and model carbon fluxes of varying soils.

This research was supported by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research under Contract No. DESC0010775 and by the Carbon Cycling Science Program, USDA-NASA grant 20116700330364.

41. Genome-scale Reconstruction of Metabolic Networks from Microbial Communities at Sites Undergoing Natural Attenuation of Uranium

Jiao Zhao[1*] (jiao.zhao@utoronto.ca), Kelvin Li[2], Monika Bihan[2], Radhakrishnan Mahadevan[1], and **Barbara A. Methé [2]**

[1]Department of Chemical Engineering and Applied Chemistry, University of Toronto, Ontario, Canada.
[2]J. Craig Venter Institute, Rockville, Maryland, USA.

Project Goals: As part of a larger research project using 'metaomics' approaches to study subsurface microbial processes and their role in cycling of carbon, nitrogen and metals (see also "The Study of Microbial Communities using 'Metaomic' Approaches at Naturally Reducing Subsurface Sites") the objectives of this computationally focused work are: 1) to use data from 'metaomic' approaches to reconstruct metabolic networks of microbial communities at the Rifle Integrated Field Research Challenge (RIFRC) site undergoing natural attenuation of uranium, 2) provide insights into the potential mechanisms that lead to numerical dominance of specific bacterial taxa under different environmental conditions, 3) study the interaction and collaboration among members of the microbial community and 4) integrate gene expression data characterizing metabolic activity into genome-scale models to describe microbial activity in sub-optimal states.

Reductive immobilization of soluble U(VI) to the insoluble U(IV) is an efficient remediation strategy of subsurface groundwater contaminated with uranium. While U(VI) can be reduced to U(IV) by stimulation of indigenous bacteria with amendments of organic carbon, there are zones of natural attenuation of U(VI) at the RIFRC site, where U(VI) can be both reduced and remobilized in the absence of biostimulation. The elucidation of the potential mechanisms of microbial community structure and metabolism under these conditions of natural attenuation can facilitate more efficient remediation design and management strategies.

In order to undertake computational modeling approaches of microbial metabolic potential we first investigated the microbial community structure at a site previously identified as undergoing natural attenuation of uranium at the RIFRC ('JB01-05' site) using metagenomic data. This examination showed that in terms of relative abundance, β-proteobacteria dominate the community (45.2%), followed by Actinobacteria (17.9%), α-proteobacteria (14.8%) and γ- proteobacteria (13.3%). This result is in sharp contrast to the community structure seen in sites undergoing biostimulation through amendment with acetate in which δ-proteobacteria (especially members of the genus *Geobacter*) dominate with relative abundances as high as 99%. Due to a similarity in genome content between subsurface microbial community members assigned to the same taxonomic class at our study site, we pursued a pan-genome-scale approach to subsequently analyze metabolic potential at the class-level. A statistical analysis of the functional profiles from the JB site indicated that within the numerically dominant taxonomic classes there is an abundance of enzymes related to CO_2 fixation (e.g. Rubisco in the α- β-, and γ- proteobacteria, and PEPCase in the Actinobacteria). In contrast, in acetate amended sites within the numerically dominant δ-proteobacteria, there is a high abundance of enzymes related to N_2 fixation (e.g. Nitrogenase). Collectively, these results reveal different community structures and metabolic functions mediating C and N cycling under these contrasting environmental conditions.

Using metagenomic and reference genome datasets, pan-genome-scale metabolic networks were reconstructed for α- , β-, γ- and δ-proteobacteria, and Actinobacteria, respectively. The models were optimized and gaps filled to ensure that they are capable of growth in geochemical conditions similar to that of the RIFRC site. These class-level models were then integrated into a Dynamic Multi-species

Metabolic Modeling (DMMM) framework for investigating the interaction and collaboration among community members. The model analysis indicates that *Thiobacillus denitrificans* may dominate the community at the JB site due to its ability to use inorganic electron donors for energy and fix CO_2 as its major carbon source. Through electron transport with cytochrome bc1 complexes and NADH-Q oxidoreductase, a tight coupling between Fe(II) oxidation and NO_3 reduction can be established to support use of CO_2 as the main source of carbon. Similarly, reduced inorganic sulfur compounds may be oxidized to sulfate by ferricytochrome c with reduction of nitrate as a terminal electron acceptor and fixation of CO_2 as the major carbon source.

Interaction and collaboration of microbial community members were quantitatively estimated through the DMMM approach. While competitive interactions mainly occur in the community for electron donors and acceptors, and carbon sources, the simulations indicate that there are potential syntrophic interactions between β-proteobacteria (e.g. *Thiobacillus denitrificans*) and Actinobacteria (e.g. *Streptomyces*). *Streptomyces* may use the products of sulfur oxidation (e.g., sulfate) from *Thiobacillus denitrificans* as the final electron acceptor for CO_2 fixation under anoxic conditions.

Subsurface microorganisms may grow within either optimal or sub-optimal states depending on ever changing environmental conditions. Hence, application of a flux balance analysis (FBA) that seeks to maximize or minimize an objective function may not be always appropriate for describing and predicting microbial activities in the subsurface. Therefore, we are currently developing methods to integrate metatranscriptomic (gene expression) data into the genome- scale models to better identify the metabolic states of various community members, thereby elucidating functional mechanisms of the community indicative of, and relevant to, sub-optimal states. These combined models will be incorporated into the DMMM framework to improve the predictive capability of the genome-scale models. Further, the metatranscriptomic data will also be used for reconstructing the metabolic network of specific individual organisms of interest in the microbial community.

This project is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

42. The Study of Microbial Communities using 'Metaomic' Approaches at Naturally Reducing Subsurface Sites

Kelvin Li[1], Monika Bihan[1], Mary Lipton[2], Brian Anderson[2], Krishna Mahadevan[3], Kenneth H. Williams[4], Jiao Zhao[3], **Barbara A. Methé[1*]** (bmethe@jcvi.org)

[1]J. Craig Venter Institute, Rockville MD, [2]Pacific Northwest National Laboratory, Richland WA, [3]University of Toronto, Canada, [4]Lawrence Berkeley National Laboratory, Berkeley CA

Project Goals: Diverse microbial communities exist in subsurface environments that possess significant metabolic potential to effect global carbon, nitrogen and metal cycles including the transformation of radionuclides. The objectives of this ongoing project are: 1) to apply systems-level biology through the application of 'metaomics' approaches (collective analyses of whole microbial community DNA, RNA and protein) to the study of microbial environmental processes and their relationship to carbon, nitrogen and metals including the influence of microbial communities on uranium contaminant mobility in subsurface settings undergoing natural attenuation, 2) improve methodologies for data generation using metaomics technologies and the analysis and interpretation of that data and 3) use the data generated from these studies towards microbial community-scale metabolic modeling (see also "Genome-scale Reconstruction of Metabolic Networks from Microbial Communities at Sites Undergoing Natural Attenuation of Uranium").

To meet the goals of this project, two subsurface sites from the Department of Energy (DOE) Rifle Integrated Field Research Challenge Site (RIFRC) are being interrogated using a suite of metaomic approaches. The first site consists of sediments from the Winchester 2007 gallery, 'JB" well locations and was chosen due to the occurrence of natural attenuation of uranium (uranium reduction in the absence of biostimulation or other remedial interventions) which has critical implications towards the design and implementation of remedial strategies for uranium removal from groundwater. Although biostimulation experiments via the addition of acetate to groundwater have not taken place here, uranium reduction has been verified by absorption spectroscopy. Monitored natural attenuation from a practical standpoint is an important and likely necessary, complement to any bioremediation strategy that might be employed in such settings due to the size and scope of the remediation required. The second and more recent sites of study within this project have been collected from Colorado River floodplain sediments representing recent sediment depositions. Overbank deposits in the floodplain have become enriched in C, Fe and S minerals. Aggradation processes have led to the subsequent burial of these enriched sediments creating "hotspots" of biogeochemical activity which serve as analogs to the buried naturally reduced sediments at the JB sites.

From the 'JB' sediments (including the JB01-05' at 4m depth) metagenomic (DNA) and metatranscriptomic (RNA) sequence has been generated using the Illumina HiSeq and MiSeq platforms While uranium reduction has been confirmed at these sites, complete immobilization of uranium has not been noted and results from this analysis provides insights into this finding. Taxonomic profiles generated from both assemblies and high quality alignments (>60bp quality trimmed read length at >80% composite identity) to the NCBI NT database revealed that for both the metagenomic and metatranscriptomic data sets, the most abundant species based on best matches (~28% DNA, ~14% RNA) are to relatives of the facultative anaerobic chemolithotroph, *Thiobacillus denitrificans* which is capable of coupling the oxidation of inorganic sulfur compounds to the reduction of oxidized nitrogen compounds. While there is interest in the use of *T. denitrificans* for use in the removal of nitrate and sulfide in environmental settings, this organism has also been shown

to be capable of the oxidization of U(IV) to soluble U(VI) in the presence of nitrate potentially accounting for the lack of complete uranium immobilization at this site. Evidence for the presence of metal reducing bacterial relatives (although not necessarily demonstrated to reduce uranium) such as *Rhodoferax ferrireducens*, were determined. Intriguingly, evidence for other biological mechanisms that could serve as potential contributors to uranium immobilization were also determined including homologs to the genes responsible for the synthesis of delftibactin, a secondary metabolite responsible for the biomineralization of gold as part of a protective mechanism against gold toxicity.

The relatively high abundances of *T. denitrificans* and *R. ferrireducens* in the JB data has served as motivation to develop a single combined genome-scale model based on a constraint-based metabolic reconstruction of the two organisms to gain an understanding of organism interaction. While a manually curated metabolic model for *R. ferrireducens* was already available, only an automatically generated model (through Model SEED) was available for *T. denitrificans* which has required additional manual curation to facilitate its use in this study. In order to match the biomass objective functions of the two organisms, the *T. denitrificans* model was augmented to share the same reaction stoichiometry as *R. ferrireducens* for DNA, RNA, Protein, phospholipids and lipopolysaccharides synthesis and missing metabolic capabilities such as sulfur oxidation reactions associated with known SOX genes have been added. Once both of the individual models were capable of producing flux through their respective biomass reactions, the stoichiometry matrices underlying each organism's model was combined. To generate a combined biomass reaction function, the objective function was chosed to be a weighted combination of *R. ferrireducens* and *T. denitrificans* equations producing 21 combined models based on the spectrum of biomass stoichiometry reweightings. Preliminary flux balance analysis and flux variability analysis are underway as well as additional work to include modifying the bounds of intracellular reaction fluxes based on metatranscriptome data.

The recently acquired floodplain samples represent an important opportunity to contrast microbial community diversity and function especially the coupling of carbon, nutrient and metal cycles with the results obtained from the JB sites. To that end, biological and technical replicates of metaomic data are currently being produced. To improve the contiguity of metagenomic assemblies (which subsequently form a framework for metatranscriptomic and metaproteomic analysis and computational modeling) in addition to the use of the Illumina sequencing platform, the use of PacBio sequencing is being investigated. PacBio sequencing has the ability to produce significantly longer read lengths (3kb or longer) when compared to Illumina platforms (<300bp) although with reduced accuracy (~90%). Therefore, new approaches for hybrid assembly including steps which incorporate mapping Illumina assemblies to PacBio reads in an iterative fashion are being tested. These approaches are broadly applicable and can be applied to other environments dominated by microbially mediated elemental cycling processes.

This project is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

43. Opportunities for Cheating? Transcriptomic Analysis of Model Organisms Exposed to Shifting Communities and Substrates

Shifang Hsu, [1] Yemin Lan, [2] Nivedita Clark,[1] Y. Devinda Hiripitiyage,[1] Laura G. Leff, [1] Gail Rosen, [3] and **Christopher B. Blackwood[1*]** (cblackwo@kent.edu)

[1]Dept. of Biological Sciences, Kent State University, Kent OH 44242
[2]Dept. of Biomedical Engineering, Drexel University, Philadelphia PA 19104
[3]Dept. of Electrical and Computer Engineering, Drexel University, Philadelphia PA 19104

http://drosophila.biology.kent.edu/users/cblackwo/overview.htm

Project Goals: Plant cell wall polymers, such as cellulose, xylan, pectin (PGA), and lignin, must be degraded into monomers before being taken up and metabolized by microorganisms. Polymer degradation is catalyzed by extracellular enzymes produced by certain microorganisms (investors). However, production of extracellular enzymes is an energy-intensive process. Thus, microorganisms shift expression of extracellular enzyme genes depending on the availability of monomers that can be taken up. During plant decomposition, cheaters are microbes that do not contribute to extracellular enzyme pools but take up polymer degradation products released by the activity of other organisms. In this experiment our goal was to examine the transcriptional and physiological responses of model fungi growing on polymers to amendment with monomers. Then, we tested predictions about enzyme expression and monomer uptake (and therefore cheating) when grown together in a mixed culture.

The model fungi *Talaromyces stipitatus* and *Schizophyllum commune* were grown in microcosms containing cellulose as a carbon substrate, requiring investment in extracellular cellulolytic enzymes. After respiration stabilized, monomers of differing types and amounts were added to replicate microcosms to determine effects on organism transcriptomes and physiology.

Respiration and enzyme activity of *S. commune* shifted slowly but for a sustained period in response to monomer amendment. There was a stronger response to glucose and galacturonic acid amendment than xylose amendment. In contrast, *T. stipitatus* responded rapidly and strongly to monomer amendment, and responded more strongly to xylose than glucose or galacturonic acid.

We also observed dramatic shifts in *T. stipitatus* gene expression 7 and 53 hr after monomer amendment. RNA was isolated from microcosms using the MoBio PowerSoil RNA Isolation kit, treated with DNase, and then mRNA transcripts were sequenced using an Illumina mRNA- seq protocol at the Yale Center for Genome Analysis. Analysis of 40 samples yielded ~190 million paired end raw sequencing reads (total 2.6 billion bases) which were analyzed using TopHat and Cuffdiff. Over half of the glycosyl hydrolases (GHs) for which transcription was detected were significantly up- or down-regulated in response to increasing xylose amendment (33 of 61 genes; $P<0.05$). Up-regulated genes include both of the beta-xylosidases for which transcription was detected, whereas down-regulated genes include several glycosyl hydrolases involved in cellulose degradation (cellobiohydrolases, beta-glucosidases, and endoglucanases). Similarly, expression of over half of the major facilitator superfamily (MFS) transporter proteins were significantly up- or down-regulated (31 of 55 transcribed genes; $P<0.05$), most of them having been up-regulated. Interestingly, genes that were up-regulated (beta-xylosidases and most MFS transporters) tended to respond to the lowest concentration xylose amendment (0.77 mM); whereas genes that were down-regulated required a higher concentration of xylose before the significant down-regulation was observed (most of them required a concentration of 4.6 mM or higher).

The data described above indicate that *T. stipitatus* is adapted to rapidly responding to small changes in available monomer concentrations by increasing uptake and altering gene expression. *T. stipitatus* appears to respond more rapidly and at a lower monomer concentration than S. commune. Furthermore, *T. stipitatus* rapidly takes up and utilizes xylose at all concentrations, but xylan-degrading enzyme activity is only induced at the higher xylose concentrations. These traits should provide *T. stipitatus* the ability to opportunistically cheat when grown with S. commune by taking advantage of transient monomers. Next, we describe results from 13C- labeling to determine which of these organisms consumes monomers when growing in mixed culture.

We then tested the prediction that *T. stipitatus* should be a cheater when grown in co-culture with *S. commune*. We conducted mixed culture experiments and amended with 13C-labeled glucose and xylose to determine patterns in C consumption and investment in extracellular enzymes. DNA stable isotope probing (SIP) was performed to detect the 13C labeled genomes. Preliminary SIP data show that 7 hr after 13C-monomer amendment, the unlabeled *T. stipitatus* DNA decreased, while DNA corresponding with labeled *T. stipitatus* and unlabeled *S. commune* DNA slightly increased. As monomer concentration or incubation time increased, the microbial DNA became more strongly labeled by 13C. However, *S. commune* DNA never became fully labeled, indicating that growth was maintained on cellulose degradation products. Labeling of *T. stipitatus* DNA may be stronger, and will be confirmed by quantitative PCR of DNA fractions. Rapid uptake of *T. stipitatus* monomers resulting in 13C labeled genomes, coupled with our previously documented stronger reductions in expression of extracellular enzyme genes in response to monomer amendments, indicate cheating of *T. stipitatus* under the growth conditions we simulated.

This work was funded by Office of Science (BER), US Department of Energy (DE-SC0004335)

44. Effects of Competitors or Cheaters and Temperature on Physiological Performance and Gene Transcription of Model Fungi

Y. Devinda Hiripitiyage,[1] Nivedita Clark,[1] Shifang Hsu, Y. Lan, [2] **L. Leff**, [1] **G. Rosen**, [3] and **Christopher B. Blackwood**[1*] (cblackwo@kent.edu)

[1]Dept. of Biological Sciences, Kent State University, Kent OH 44242
[2]Dept. of Biomedical Engineering, Drexel University, Philadelphia PA 19104
[3]Dept. of Electrical and Computer Engineering, Drexel University, Philadelphia PA 19104

http://drosophila.biology.kent.edu/users/cblackwo/overview.htm

Project Goals: During the decomposition of plant materials, microorganisms take up monomers for use in respiration and biosynthesis. Plant cell wall polymers, such as cellulose, pectin and lignin, are degraded into monomers by microorganisms capable of producing extracellular enzymes. However, production of extracellular enzymes is an energy-intensive process, and some organisms may be adapted to take advantage of the activity of enzymes produced by other microbes (i.e., "cheating"). Our overall project goal is to describe in detail how substrate type and growth in a mixed community (with potential for cheating) affects gene transcription, microbial physiology, and plant litter decomposition. Here, we specifically test the hypothesis that temperature may shift the relationships among microbial species. Increased temperature is expected to increase activity of extracellular enzymes as well as aqueous diffusion of both extracellular enzymes and degradation products. Hence, we hypothesize that cheating by microbes will increase at higher temperatures due to higher availability of polymer degradation products.

Pure cultures of *Trichoderma reesei* QM6a, *Phanerochaete chrysosporium* RP-78, and *Rhodotorula sp.* were grown as monocultures and co-cultures in sand microcosms containing minimal nutrients and ground beech leaves or cellulose as the sole carbon source. The microcosms were incubated at 20°C or 30°C. Shifts in gene expression were determined by sequencing transcriptomes in replicate microcosms by Illumina HiSeq2500, resulting in ~10 million sequences per sample. Physiological measurements of microbial performance included biomass, respiration, polymer and monomer concentrations, and extracellular enzyme activity including beta-glucosidase, alpha-glucosidase, beta-xylosidase, cellobiohydrolase, nagase, and phosphatase. During growth on cellulose, *T. reesei* accumulated the most biomass, but biomass was further increased by growth in co-culture with either *P. chrysosporium* or *Rhodotorula sp.* ($P<0.05$). Respiration from cellulose microcosms reflected respiration of the organism with the highest biomass in monoculture; in particular, *T. reesei* respiration did not appear to be affected by growth in co-culture even though biomass was increased ($P>0.05$). In contrast, in beech microcosms, biomass accumulation and respiration were highest for *P. chrysosporium.* Biomass in beech co-cultures reflected the biomass of the organism with the highest biomass in monoculture ($P>0.05$), but respiration in *P. chrysosporium* microcosms was increased due to growth in co-culture ($P<0.05$). Thus, carbon use efficiency of *T. reesei* growing on cellulose appeared to increase due to co-culture, whereas carbon use efficiency of *P. chrysosporium* growing on beech leaves appeared to decrease due to co-culture. Analysis of transcript abundance and relative biomass of each organism will provide information needed to determine rates of carbon consumption relative to investment in extracellular enzymes.

This work was funded by Office of Science (BER), US Department of Energy (DE-SC0004335)

45. A Systems Biology Characterization of the Biotechnological Potential Stored in the Wood-Feeding Beetle Odontotaenius disjunctus

Javier A. Ceja-Navarro[1], Ulas Karaoz[1], Mari Nyyssonen[1], Mary S. Lipton[2], Joshua N. Adkins[2], Kristin E. Burnum-Johnson[2], Markus Bill[1], Timothy Filley[3], Mark Conrad[1], **Xavier Mayali[3], Peter K. Weber[3], Meredith Blackwell[4], Jennifer Pett-Ridge[3], Eoin L. Brodie[1*]**(ELBrodie@lbl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Pacific Northwest National Laboratory, Richland, WA; [3]Purdue University, West Lafayette, IN; [3]Lawrence Livermore National Laboratory, Livermore, CA; [4]Louisiana State University, Baton Rouge, LA.

Project Goals: We seek to develop an integrated analysis of energy flow in complex microbial communities by combining multi-scale approaches including biogeochemical, stable isotope probing, metagenomic/transcriptomic, proteomic/metabolomic and computational analyses, to understand nutrient cycling and biofuel production. A comprehensive understanding of such communities may help in the development of efficient, industrial-scale processes for microbial H_2 production and lignocellulose degradation. Our ultimate goal is the development of multi-scale models that can predict ecological and biochemical relationships within multi-trophic microbial systems.

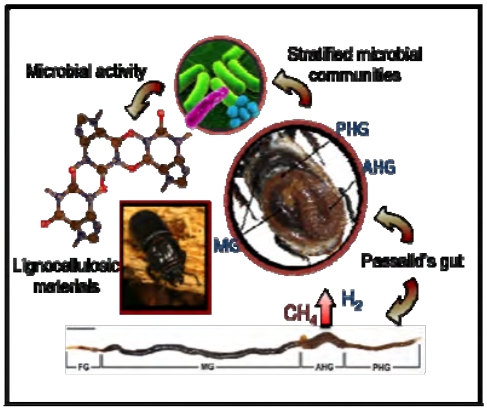


Fig 1. The passalid beetle and its gut. The associated microbial communities aid in the transformation of complex polymers.

The passalid beetle (*Odontotaenius disjunctus*) is a subsocial insect that survives on a low-nutrient diet by feeding on large amounts of decaying wood. The morphologically differentiated gut regions of these insects represent a complex of subunits with stratified microbial communities that degrade lignocellulosic materials. Our goal is to characterize the potential stored in the microbiome of the passalid beetle for the optimization of lignocellulosic-dependent energy production processes (Fig. 1).

We tested the ability of the passalid beetle to transform lignocellulosic materials by measuring the fermentation products of plant polymer decomposition (H_2 and CH_4) using microelectrodes and gas chromatography- isotope ratio mass spectrometry (GC-IRMS). Transformations of lignin after its passage through the gut were determined by ^{13}C-labeled tetramethylammonium hydroxide thermochemolysis. Using fosmid libraries constructed from DNA from different beetle gut regions we have done high-throughput screening for lignin, cellulose, and hemicellulose degrading activity. Illumina-baed metagenomic libraries were also prepared, sequenced and annotated along with the positive clones from our fosmid libraries. Proteomics and metabolomics have been used to detect expressed proteins and produced metabolites in each gut region. Finally, we have screened beetle gut microbiota for assimilation of ^{13}C-labeled cellulose using Chip-SIP isotope arrays and NanoSIMS imaging.

Lignin side chain oxidation was confirmed by thermochemolysis which show acid/aldehyde ratios increasing in the beetle frass. Hydrogen gradients, were measured using microelectrodes, and indicate concentrations as high as 140 µmol/L in the anterior hindgut (AHG). GC-IRMS analyses of C and H

stable isotope fractionation indicated that the produced CH_4 was primarily hydrogenotrophic. Fosmid library screening yielded a high number of clones with activity for the decomposition of cellulose, hemicellulose, and lignin – with the highest potential detected in the AHG. The annotation of metagenomic libraries has allowed us to identify the likely contributors to cellulose, xylose and lignin modification. Notably lignocellulosic organisms related to the Clostridiales, Bacillales, Actinobacteria were abundant. We also identified the presence of fungal laccases and peroxidases in addition to bacterial peroxidases that may be involved in the process of lignin degradation. Sequences from hydrogenotrophic methanogenic archaea were more abundant in the anterior hindgut, confirming our previous phylogenetic studies of compartmentalization in the passalid beetle gut. A filtered isolate database and predicted protein sequences from the metagenomes were used to search peptide spectra for proteome reconstruction – preliminary results indicate a variation in the protein expression patterns among the different gut segments potentially indicating a compartmentalization of function. ChipSIP analyses to identify consumers of ^{13}C-cellulose are ongoing.

Our multi-scale approach demonstrates that the passalid beetle harbors and expresses the functional potential to deconstruct lignocellulosic materials and produce H_2, CH_4 and potentially other biofuels. Identifying the microbial contributors to polymer deconstruction and fermentation, and determining their spatial arrangement in the beetle gut will improve our understanding of the ecology of these beetles and inform the design of lignocellulosic fuel production processes.

46. Understanding the carbon and hydrogen flow in constructed H_2-producing co-cultures

Yongqin Jiao[*1], R. Craig Everroad[*2], Ali Navid[1], Jackson Z. Lee[1,2], Angela M. Detweiler [1,2], Benjamin J. Stewart[1], James B. McKinlay[4], Leslie Prufert-Bebout[2], **Michael P. Thelen[1], Brad M. Bebout[2], Jennifer Pett-Ridge[1]**

[1]Lawrence Livermore National Laboratory, Livermore, CA; [2]NASA Ames Research Center, Moffett Field, CA; [3]Bay Area Environmental Research Institute, Sonoma, CA; [4]Indiana University, Bloomington, IN.

Project goals: Microbial communities have multiple pathways for energy conversion and nutrient transfer. Which pathways dominate and how these pathways are interconnected greatly affects community function. Microbiologists have also started to appreciate the fact that the whole may be much more than the simple sum of its parts. The interactions between different components of a system result in many new physiological functions that cannot be observed with individual components. Our goal here is to develop genome-level models of carbon and energy flow within simplified microbial communities (co-cultures), with H_2 production as a metric for monitoring the overall state of the system.

In order to develop a defined ecosystem and to study the community metabolic interactions in detail, we have isolated and characterized a broad suite of microbial mat community members from the intertidal mats of Elkhorn Slough, CA. We have successfully isolated (as true isolates or stable consortia) all the major functional groups of the natural intertidal mat ecosystem including sulfur-oxidizing and -reducing bacteria, oxygenic and anoxygenic phototrophs, respiring and fermentative heterotrophs, and nitrogen-fixing and denitrifying bacteria, as well as the vast majority of the 15 most abundant taxa (as determined by metagenomics analysis of the mat community). With isolation efforts largely finished, we are now shifting our effort to examine interactions in mixed cultures containing 2 or more different functional groups. Preliminary successes include the construction of a stable light-driven sulfur-cycle between phototrophic purple bacteria and a sulfate-reducing bacterium; and the construction of cyanobacterial/fermentative co-cultures that considerably stimulate net H_2-production. Our objective is to gain an emergent vision of what physiology and interactions facilitate stable and successful consortia in general ecological terms (*i.e.*, higher overall fitness, broader range of optimal growth conditions or high stability) as well as for practical applications (*i.e.*, high net H_2 productivity).

In parallel with constructing simplified microbial communities from natural systems, we have constructed an artificial co-culture containing *Clostridium cellulolyticum* H10 (CC) and purple bacterium *Rhodopseudomonas palustris* CGA676 (RP) for H_2 production based on cellulose degradation. Cellulose is the sole carbon and energy source for CC. RP utilizes fermentation products secreted by CC, and H_2 is produced by both organisms. To understand metabolism in this syntrophic system, we examined and compared the kinetics of cellular growth, H_2 production, and metabolite production/consumption in both mono- and co-cultures of these two taxa. Because both organisms have been fully sequenced, their co-culture is an ideal way to begin to understand the genomic underpinnings of syntrophy. Our results show that the presence of RP in co-culture greatly stimulates cellulose degradation and H_2 production, which likely results from accelerated metabolism of CC and consumption of acetate and pyruvate by RP. Using a high density microarray, we also investigated changes in CC's gene expression when co- cultured with RP. 291 genes had significantly different gene expression in co-culture, with 179 genes up-regulated and 112 down-regulated. Changes

in expression were calculated as an intensity ratio of CC co-culture versus CC monoculture. Many of the up- regulated genes are involved in translation, replication and cellulose breakdown, consistent with our observation of higher growth rates of CC when in co-culture. We also observed a significant increase in expression of a pyruvate kinase (16 fold) and a lactate dehydrogenase (8 fold), suggesting that CC produces more pyruvate and lactate when in co-culture.

In order to study system-wide carbon and energy flow within this co-culture, we are developing a genome-scale reconstruction of metabolism in CC and RP and are using Flux Balance Analysis (FBA) to study the metabolic capabilities of these organisms under different genetic and environmental conditions. To date, we have developed a system-level model of metabolism in RP and have used this model to examine the modes of metabolism most conducive to H_2 production. The model's predicted behaviors have been compared with experimental observations to ensure accuracy. As expected, our *in silico* analysis showed the RP does not require light to grow in nutritionally rich environments. Furthermore, we tested its ability to grow on organic acid byproducts of CC metabolism (acetate, ethanol, lactate and pyruvate). Consistent with results of fluxomic analysis, our results indicate that H_2 production is closely linked to CO_2 production. Carbon fixation results in reduced production of H_2. Consumption of carbon sources like ethanol and acetate that are more reduced than the cellular biomass results in greater production of H_2 in comparison to more oxidized compounds like pyruvate. Additionally, our *in silico* analyses indicate that RP has a very robust mechanism for autotrophic growth and is not dependent on the Calvin cycle for photoheterotrophic growth. This result has not been experimentally verified and is in disagreement with observed essentiality of Rubisco for photoheterotrophic growth in other purple- non- sulfur bacteria. Overall, since H_2 production is inversely linked to carbon fixation/conservation, increased H_2 yield adversely effects cellular growth.

Through the systems biology studies of these simplified co-cultures for hydrogen production, we have gained insights into key metabolic interactions and population dynamics in such co-cultures that are critical for improvement of syntrophic efficiency and H_2 productivity. Continued improvements in our ability to track metabolite fluxes will provide a more comprehensive picture of intermediary metabolism and the factors that control the flux of organic matter that result in H_2 production.

Funding was provided by the U.S. Department of Energy (DOE) Genomic Science Program under contract SCW1039. RCE acknowledges the support of the NASA Postdoctoral Program, administered by Oak Ridge Associated Universities through a contract with NASA. Work at LLNL was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under contract DE-AC52- 07NA27344.

47. Bioenergy and Biogeochemical Cycling in Elkhorn Slough Hypersaline Microbial mats

Rhona K.Stuart[1*] (stuart25@llnl.gov), Jackson Z. Lee[2,3*] (jackson.z.lee@nasa.gov), R. Craig Everroad[2], Luke Burrow[2,4], Dagmar Woebken[2,4,5], Angela M. Detweiler[2,3], Mike D. Kubo[2,6], Tori M. Hoehler[2], Whitney Stannard[1], Leslie E. Prufert-Bebout[2], **Xavier, Mayali[1], Jennifer Pett-Ridge[1], Michael Thelen[1], Peter K. Weber[1], Brad M. Bebout[2]**.

[1]Lawrence Livermore National Laboratory, Livermore, CA, [2]Exobiology Branch, NASA Ames Research Center, Moffett Field, CA, [3]Bay Area Environmental Research Institute, Sonoma, CA, [4]Department of Civil and Environmental Engineering and Chemical Engineering, Stanford University, CA [5]University of Vienna, Austria [6]The SETI Institute, Mountain View, CA

Project Goals: The LLNL Biofuel SFA investigates systems biology of complex microbial communities. Specific goals in this work are to develop an integrated analysis of energy flow in complex microbial communities by combining multi-scale approaches including biogeochemical, stable isotope probing, metagenomic/transcriptomic, proteomic/metabolomic and computational analyses, to understand nutrient cycling and potential biofuel production in complex microbial communities. Our ultimate goal is the development of multi-scale models that can predict ecological and biochemical relationships within multi-trophic microbial systems.

Marine hypersaline cyanobacterial mats are diverse laminated microbial assemblages, thought to represent life on the early earth. The mats at Elkhorn Slough, CA and Guerrero Negro, Mexico have intrinsic relevance to biofuels systems biology because they produce significant nighttime fluxes of hydrogen gas and other potential biofuels as fermentation byproducts (Figure 1), and they cycle carbon, oxygen, sulfur, and nitrogen on a millimeter scale. The overall goal of these studies is a systems-level understanding of the partitioning of light and geochemical energy into biomass and potential biofuels in these complex microbial communities. Past studies in Elkhorn Slough have shown production largely from *Cyanobacteria* and consumption from sulfate-reducing bacteria (SRB).

However, the mechanisms and magnitude of hydrogen cycling are not well understood. The overall goal of the present work is a systems-level understanding of the partitioning of light and geochemical energy into biomass in microbial mat communities.

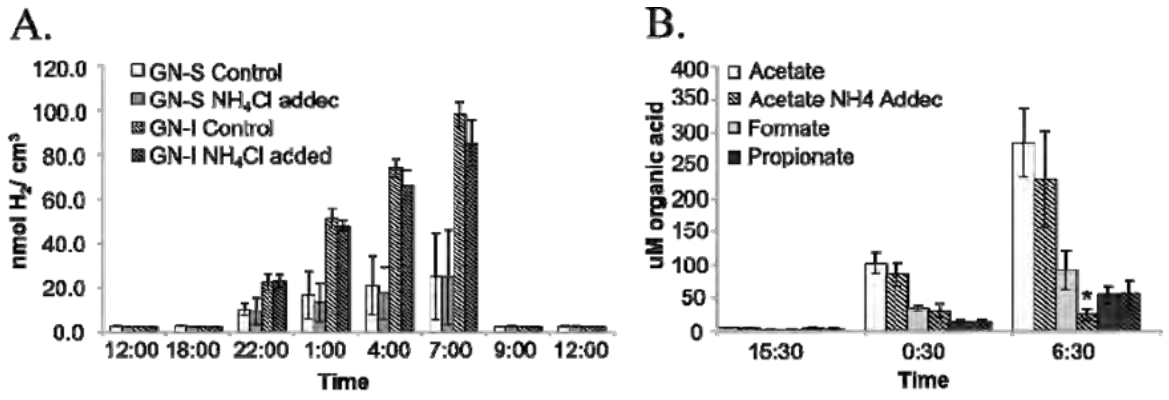


Figure 1. Hydrogen net flux over a diel (A), and net flux of common organic acids (B) in GN-I and GN-S mats.

Mats from Guerrero Negro, Mexico - permanently submerged *Microcoleus* microbial mats (GN-S), *Lyngbya* intertidal microbial mats (GN-I) - were used in microcosm diel manipulation experiments (with ammonium chloride) to determine mechanisms responsible for hydrogen cycling between mat microbes. H_2 production occurred under dark anoxic conditions with simultaneous production of a suite of organic acids (Figure 1). H_2 production appears to result from constitutive fermentation of photosynthetic storage products (glycogen) over the day night cycle. Comparisons of accumulated glycogen and CO_2 flux indicated that in the GN-I mat, fermentation released a majority (~90%) of the carbon fixed via photosynthesis during the preceding day, primarily as organic acids. In parallel, incubations with ^{13}C-acetate and nanoSIMS isotopic imaging indicated higher uptake in both *Chloroflexi* and SRBs relative to other filamentous bacteria. These manipulations and diel incubations confirm that *Cyanobacteria* were the main fermenters in Guerrero Negro mats and that the net flux of nighttime fermentation byproducts (mostly acetate) was largely regulated by the interplay between *Cyanobacteria*, SRBs, and *Chloroflexi*. These data suggest that light energy partitions primarily into the *Cyanobacteria* (stored glycogen photosynthate) and is then released at night into EPS (extracellular polymeric substances), perhaps as organic acids. Ongoing experiments will examine how this photosynthate is partitioned and the fate of carbon released into the extracellular matrix.

Preliminary analyses of the genome of ESFC-1has also been conducted. This taxa is an important member of the Elkhorn Slough system and a newly described lineage of filamentous diazotrophic cyanobacteria. We have identified EPS proteins and characterized EPS composition in both ESFC-1 and Elkhorn Slough mats. The most abundant of the extracellular proteins are predicted to be involved in protein and sugar degradation and putative structural components. This suggests that cyanobacteria may facilitate carbon transfer to other groups through degradation of their EPS components.

To determine which organisms were responsible for glycogen formation and storage and which organisms were involved in organic acid uptake, a combined metagenomic / metatranscriptomic study was completed to track the transcript response of mat-associated organisms over the course of a day as energy passed from sunlight into fixed carbon and nitrogen and subsequently into nighttime fermentation products. A total of 4 metagenomes and 9 metatranscriptomes (over 9 time points of the day) were sequenced. Future experiments also include metaproteome analyses over a diel cycle and tracing of ^{13}C – labeled EPS.

In summary, this work contributes to our understanding of the partitioning of major pathways of energy from sunlight into various hydrogen and carbon utilizing microbes in complex multi-trophic microbial systems.

Funding was provided by the U.S. Department of Energy (DOE) Genomic Science Program under contract SCW1039. Work at LLNL was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344.

48. Uncovering Uranium Resistance Mechanisms in *Caulobacter crescentus*

Mimi Yung,[1] Dan Park,[1] Cindy Hoover,[2] Matthew Blow,[2] Michelle Salemi,[3] Brett S. Phinney,[3] Grant Bowman,[4] and **Yongqin Jiao**[1*] (Jiao1@llnl.gov)

[1]Lawrence Livermore National Laboratory, Livermore, CA; [2]DOE Joint Genome Institute, Walnut Creek, CA; [3] Genome Center Proteomics Core Facility, University of California Davis, Davis; [4]University of Wyoming, Laramie.

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, as well as its ubiquitous presence in fresh water lakes and streams, including those contaminated with heavy metals. Our goal is to decipher the physiological basis for U tolerance and elucidate the U biomineralization pathways in *C. crescentus.* Our end goal is to provide insight into the effect of aerobic bacteria on U biogeochemistry and to understand whether aerobic biomineralization could be used in bioremediation applications.

We observed that U (VI) caused a temporary growth arrest in *C. crescentus* and three other bacterial species, although the duration of growth arrest was significantly shorter for *C. crescentus*, which may provide a competitive advantage in the environment.[1] We found that growth recovery was not due to a decrease in U solubility, a common detoxification strategy employed by other microorganisms. Through functional reporter assays, we discovered that *C. crescentus* is able to reduce U bioavailability through the secretion of an unknown, heat-stable metabolite(s). To the best of our knowledge, this represents a unique U detoxification strategy and has important environmental implications for how aerobic bacteria affect U biogeochemistry. Furthermore, our findings provide insight into how microbes cope with (metal) stress under non-growing conditions, a metabolic state that is under-studied but prevalent in the natural environment.

Upon recovery from growth arrest, *C. crescentus* proliferated with normal growth kinetics, during which active U biomineralization occurs. We found that phosphate metabolism facilitated U-P precipitation when organic phosphate (*e.g.*, glycerol-2- phosphate) was provided. Electron microscopic and spectroscopic analyses indicated that microbe-assisted U precipitates were distinct from their abiotic counterparts in both morphology and composition. In particular, we observed cell-surface-bound U-P minerals, indicating a biological assisted process. In addition, we found that a predicted extra-cytoplasmic alkaline phosphatase (CCNA_02545) was responsible for formation of these U-P precipitates; deletion of this gene abolished the formation of the precipitates.

Furthermore, the activity of this enzyme facilitated cell survival in a whole-cell assay under U treatment as well as during growth in minimal medium supplemented with U and glycerol-2-phosphate as the sole phosphate source. Overall we demonstrated that *C. crescentus*-facilitated U biomineralization occurs during active growth and may play an important role in cell persistence in contaminated environments.

To identify the genetic basis for U tolerance in *C. crescentus* on the genome level, we took two independent approaches: proteomic profiling and Tn-seq. We performed a label-free shotgun proteomics study of *C. crescentus* under U, Cr, or Cd exposure.[2] The goal was to identify proteins differentially expressed under heavy metal stresses, and to compare the proteomic results with the

already available whole genome transcriptional data. Under U exposure, a phytase enzyme and an ABC transporter were up-regulated. Heat shock and outer membrane responses were found associated with Cr, while efflux pumps and oxidative stress proteins were up-regulated with Cd. Experimental validations indicated that the phytase plays a role in U and Cr resistance and detoxification, and a Cd-specific transporter confers Cd resistance. Interestingly, analysis of promoter regions in genes associated with differentially expressed proteins suggests that U exposure affects cell cycle progression. The results of this study not only broaden our understanding of the fundamental aspects of metal stress response, but also provide insight into the roles of specific proteins in metal detoxification.

In collaboration with JGI, we employed a Tn-seq approach to identify the essential genomic elements that specifically confer U resistance in *C. crescentus.* The specific steps of our method included: 1) performing ultrahigh-resolution transposon mutagenesis; 2) selecting for mutants that can grow in the presence of U; 3) amplifying genomic regions adjacent to the transposon insertion sites from the pooled mutants; 4) performing high-throughput DNA sequencing to obtain genomic sequence information; and 5) mapping the transposon insertion sites onto the genome of *C. crescentus*. Genomic areas that accumulate transposon insertions under normal growth conditions but tolerate fewer insertions under U selective pressure contain genes that are specifically required for growth under U. Using this approach, we identified 18 genes with significantly lower Tn insertion frequencies under U exposure compared to the no stress control. Through subsequent mutation analyses, we confirmed 15 genes were involved in U tolerance, including those that partake in type-I secretion, flagella assembly, stationary phase response, and general stress response, many of which are also involved in cell cycle progression. Together, we are starting to gain a basic understanding of strategies employed by *C. cresecentus* for coping with U toxicity at the physiological and molecular levels.

References

1. Park, D.M. and Y. Jiao. "Metabolic secretion facilitates recovery from uranium induced growth arrest in *Caulobacter crescentus*". Applied and Environmental Microbiology. *Under review.*
2. Yung, M.C., J. Ma, M. Salemi, B.S. Phinney, G.R. Bowman, and Y. Jiao. "Shotgun Proteomic analysis unveils survival and detoxification strategies by *Caulobacter crescentus* during heavy metal exposure". Journal of Proteome Research. *Under review.*

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-648058).

49. Halophilic Communities as a Source for Novel Lignocellulolytic Enzymes

Janet K. Jansson[1,2*] (jrjansson@lbl.gov), Jeffrey Kimbrel[2,3], Nicholas Ballor[2,3], Hannah Woo[2,3], Thomas Ruegg[2,4], Terry C. Hazen[2,5], Michael P. Thelen[2,4], Blake A. Simmons[2,6], Steven W. Singer[1,2]

[1]Earth Sciences Division. Lawrence Berkeley National Laboratory, Berkeley, CA; [2] Deconstruction Division, Joint BioEnergy Institute, LBNL, Emeryville, CA; [3]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [4]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA; [5]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [6]Biological and Materials Sciences Center, Sandia National Laboratories, Livermore, California

www.jbei.org/research/divisions/deconstruction/microbial-communities

Project Goals:
Characterize halophilic communities from saline environments as sources of novel halophilic microorganisms, genes, and enzymes for biofuel feedstock deconstruction.

Selectively enrich microbial populations from complex microbial communities from saline environments on biofuel feedstocks (Miscanthus, Pine and Eucalyptus) to obtain candidates potentially capable of deconstruction of feedstocks under high salinity conditions.

Describe the metabolic potential and gene expression patterns in both natural saline communities and feedstock enrichments by sequencing and screening of metagenomes, metatranscriptomes and metaproteomes.

Use functional metagenomics to express a library of genes that potentially represent novel mechanisms for deconstruction of biomass that are currently underrepresented in gene catalogues.

Formulate (by synthesis, cloning and expression of genes characterized above) and verify activity of a cocktail of halophilic enzymes for deconstruction of biomass in the presence of ionic liquids.

Lignocellulose presents a challenge to next generation biorefineries due to its recalcitrance to microbial degradation. Ionic liquid (IL)-based pretreatment has been successful in preparing biomass for enzyme saccharification, but the most common ILs used for pretreatment inhibit many downstream enzymatic and microbial processes mediated by mesophilic enzymes. Halophiles, by definition, are adapted to high-salt environments and are thus a potential source for IL tolerant enzymes. Here we sought to discover & recover novel lignocellulolytic enzymes from environmental and feedstock-enriched halophilic bacterial communities. We collected both liquid and sediment samples from different saline environments in Puerto Rico and San Francisco including salt flats, saltern ponds and turtle grass beds. For each of the environmental samples we obtained 16S rRNA gene sequences, metagenomes and metaproteomes. The data revealed an increase in relative abundance of haloarchaea and genes and proteins implicated in a hypersaline lifestyle with increasing salinity. In addition, a fosmid library was constructed in an expression vector for high throughput functional metagenomics screening using the robotics platform at JBEI.

Samples from a turtle grass bed (3.5% salinity) and a high salinity saltern pond (33.2% salinity) were selected for enrichment on the potential biofuels feedstocks: miscanthus (M), eucalyptus (E) or pine (P) under aerobic and anaerobic conditions and followed through three 2---week passages. At the end of each passage cells were harvested, specific enzyme activities were measured and DNA, RNA and proteins were extracted. Data collected include enzyme activities for B glucosidase, cellobiohydrolase and xylanase, 16S rRNA gene sequences, metagenomes metatranscriptomes and metaproteomes. We found that enzyme activity was typically highest after the first passage, with the aerobic turtle grass enrichments having consistent activity on each feedstock. After the third passage, metagenomes were constructed and binned using MaxBin, a binning algorithm developed at JBEI. Bins were subsequently searched against the CAZY/dbCAN HMMs. In addition, expressed transcripts from 11 metatranscriptomes were identified by either alignment to the reference metagenomes or *de novo* assembled. To date we have identified over 1000 expressed candidate carbohydrate active enzymes from the enrichments and obtained reconstructed genomes for >100 feedstock-enriched archaea/bacteria. Heterologous expression of a diverse collection of 29 putative glycoside hydrolases is ongoing. The next step will be to validate and incorporate these candidate enzymes into a halophilic deconstruction enzyme mixture with high activity and IL tolerance.

This work conducted by the Joint Bio Energy 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.

50. Examining the post-transcriptional program governing the metabolic proteome of Micromonas pusilla

Peter Waltman[1*] (peter.waltman@gmail.com), Marijke J. van Baren[2], Emily Reistetter[2], Samuel Purvine[3], Charles Ansong[3], Stephen Callister[3], **Alexandra Z. Worden[2]**, and **Joshua Stuart[1]**

Institutions:
[1]Biomolecular Engineering Department, University of California Santa Cruz, Santa Cruz, CA
[2]Monterey Bay Aquarium Research Institute, Moss Landing, CA
[3]Pacific Northwest National Labs, Richland, WA

https://sysbiowiki.soe.ucsc.edu/ and http://www.mbari.org/staff/azworden/

PI: Alexandra Z. Worden, Monterey Bay Aquarium Research Institute, azworden@mbari.org
Co-PIs: Stephen Callister, Richard Smith, Joshua Stuart

Project Goals: To develop a systems biology approach to the study of the widespread marine alga *Micromonas* and use it to investigate gene function, pathways and consequences of environmental perturbations on primary production.

We are developing a model system for understanding land plant evolution and ecologically important marine primary producers. Thus far there are no such model systems for the Viridiplantae - the eukaryotic lineage containing all land plants and green algae. The primary drivers for developing our *Micromonas* system are: *i*) approximately half of global photosynthetic CO_2 uptake is performed by marine algae yet there is little understanding of the physiological consequences of current global change scenarios and *ii*) green algae provide insights to eukaryotic cellular processes and the ancestor of land plants.

Prasinophytes are a group of unicellular marine green algae that are evolutionarily distinct from the model green alga *Chlamydomonas*, but are related to both the latter and land plants. *Micromonas* is a widespread prasinophyte that is exceptional in its size (<2 micrometer diameter) and has a small genome (21 Mb). For our project, we focus upon the analysis of the metabolic proteome of *Micromonas pusilla*, a *Micromonas* strain that, notably, lacks known machinery for miRNA-based translational regulation. To examine this, we utilized a strategy that performed whole transcriptome and proteome profiling over the course of a triplicated diel experiment. Comparisons of the matched RNAseq and MS/LC proteomics samples indicated that considerable differences exist between dynamics of the transcriptomic and proteomics expression programs. Less than 10% of the genes considered for this analysis had correlated transcriptomic and proteomics expression profiles.

Despite these differences in the expression dynamics, transcriptomic and proteomic expression still exhibited considerable correlation between genes belonging to the same pathways. To examine these differences in the expression dynamics, we utilized an integrative regression framework which incorporated the matched RNAseq and MS/LC proteomics, along with other measures of various mechanisms of post-transcriptional control to generate several high- accuracy global models of protein expression. In addition, we also identified 22 co-expressed gene groups (modules) containing genes that share similar expression profiles in both the transcriptomic and proteomic data. Of these, we focused on three gene modules that are each enriched with genes in the oxygenic photosynthesis pathway. While all three gene clusters share similar mRNA expression profiles, they also exhibit highly dissimilar protein expression profiles. As a final step, we identify several potential mechanisms of post-transcriptional control which may explain the different proteomic expression programs of these gene modules.

This research is supported by U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Biological Systems Research on the Role of Microbial Communities in Carbon Cycling Program Award No. DE-SC0004765 made in July 2010 with input from other funding sources.

51. Phytochromes in widespread photosynthetic algae reveal origins of plant signaling proteins

Charles Bachy[1*] (cbachy@mbari.org), Deqiang Duanmu[2], Sebastian Sudek[1], Valeria Jimenez[1], Marijke J. van Baren[1], Chee-Hong Wong[3], Nathan C. Rockwell[2], Shelley S. Martin[2], Chew Yee Ngan[3], Emily Reistetter[1], Chia-Lin Wei[3], Adrian Reyes-Prieto[4,5], **Stephen Callister[6]**, J. Clark Lagarias[2] and **Alexandra Z. Worden[1]**

[1]Monterey Bay Aquarium Research Institute, Moss Landing, CA 95060, USA; [2]Department of Molecular and Cellular Biology, University of California, Davis, CA 95616; [3]Sequencing Technology Group, Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, CA 94598, USA; [4]Biology Department, University of New Brunswick, Fredericton, New Brunswick E3B5A3, Canada; [5]Integrated Microbial Biodiversity Program, Canadian Institute for Advanced Research, Toronto, M5G 1Z8, Canada; and [6]Pacific Northwest National Labs, Richland, WA

http://www.mbari.org/staff/azworden/

Project Goals: To develop a systems biology approach to the study of the widespread marine alga *Micromonas* and use it to investigate gene function, pathways and consequences of environmental perturbations on primary production.

PI: Alexandra Z. Worden, Monterey Callister Bay Aquarium Research Institute, azworden@mbari.org
Co-PIs: Stephen, Richard Smith, Joshua Stuart

Phytochromes are photoswitches that play master regulatory roles in land plants, fungi, and bacteria. In land plants, phytochromes have undergone gene family expansion, and now perform distinct and overlapping photoregulatory functions to optimize photosynthesis or to initiate flowering and seed dispersal. Light sensing by phytochrome relies on a covalently bound linear tetrapyrrole (bilin) chromophore, whose photoisomerization triggers a reversible photoconversion between red and far-red absorbing states that modulates downstream signaling events. Phytochrome involves a vast and complicated network of genes to control developmental transitions in land plants. Although widespread in bacteria, the limited distribution of eukaryotic phytochromes is an obstacle to creating plausible evolutionary scenarios and understanding of early functional roles. Fungi and some heterokont algae possess phytochromes but other unicellular eukaryotes with sequenced genomes do not, such as most Archaeplastida lineages, including model green algal species (e.g. the chlorophytes *Chlamydomonas reinhardtii* and *Chlorella vulgaris*).

We identified phytochrome-related sequences in the genome of the prasinophyte alga *Micromonas pusilla* CCMP1545. The *M. pusilla* phytochrome gene model was confirmed using multiple lines of evidence, including immunoblot and proteomics analyses. Phylogenetic analyses and the common protein domain architecture in prasinophyte and plant phytochromes (except for a C-terminal response regulator domain in prasinophytes), support the idea of a shared ancestry as a green phytochrome lineage. We then investigated the expression of phytochrome and genes in pathways known to be influenced by phytochrome activity over the day:night cycle in *M. pusilla* by using directional pair end RNAseq (Illumina). The data show that key bilin biosynthesis genes are coordinated with *Micromonas* phytochrome with a significant predawn peak, preceding the expression of photosynthesis-related genes. The expression of phytochrome protein and its subcellular localization under a diel showed redistribution from the cytosol to the nucleus throughout the day. These results are consistent with what we know for land plants, but also indicated that the light-dependent nuclear translocation of phytochrome

predates the divergence of streptophytes and prasinophytes within the green lineage. Moreover, the *Micromonas* phytochrome displayed a previously undescribed light sensitivity, shifted to shorter wavelength relative to their land plant counterparts. This suggests that *Micromonas* phytochrome functions as a low light sensor better suited to aquatic environments where wavelengths are differently transmitted through water compared to terrestrial environments.

This research has important implications for understanding ancestral roles of phytochrome in plant development and influences adaptive on the circadian clock. Moreover, *Micromonas* provides a simplified model system to address the role played by phytochromes to light variations in changing oceans.

This research is supported by U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Biological Systems Research on the Role of Microbial Communities in Carbon Cycling Program Award No. DE-SC0004765 made in July 2010 with input from other funding sources.

52. Responses of the Widespread Green Alga *Micromonas* to Elevated CO_2 at Different Levels of Phosphorus Availability

Jian Guo1, Susanne Wilken[1*] (swilken@mbari.org), Lisa Sudek[1], Emily Reistetter[1], Samuel Purvine[2], Charles Ansong[2], Stephen Callister[2], Alexander J Limardo[1], Denis Klimov[1], Zbigniew Kolber[3] and **Alexandra Z. Worden[1,3]**

[1]Monterey Bay Aquarium Research Institute, Moss Landing, CA; [2]Pacific Northwest National Labs, Richland, WA; [3]Ocean Sciences Department, University of California Santa Cruz, Santa Cruz, CA

http://www.mbari.org/staff/azworden/

Project Goals: To develop a system biology approach to the study of the widespread marine alga *Micromonas* and use it to investigate gene function, pathways and consequences of environmental perturbations on primary production.

Marine phytoplankton are responsible for about 50% of global carbon fixation. Changing environmental conditions such as ocean acidification, warming, increased stratification of the water column and reduced supply of nutrients to the upper mixed layer will likely alter phytoplankton physiology and carbon fixation. While increased carbon availability can cause increased rates of carbon fixation, such a response might be counteracted by intensified nutrient limitation. In order to predict changes in biogeochemical cycles a mechanistic understanding of the physiological response of phytoplankton to both individual and combined environmental perturbations is required.

Here, we focus on the response of the photosynthetic picoeukaryote *Micromonas* to the synergistic effects of high CO_2 and diminished phosphate, as would be expected from reduced surface-ocean mixing. *Micromonas*, a broadly distributed unicellular green alga, has already increased in some arctic regions related to climate change. Moreover, it is related to land plants and provides a simplified model system for understanding plant evolution. Here, we performed continuous culturing experiments with *Micromonas* sp. RCC299 under phosphate replete (control), phosphate limited and phosphate re-fed conditions by utilizing a novel culturing system developed at MBARI. Because multi-factorial stresses are closer to real-life' scenarios, the combined effect of phosphate availabilities and elevated CO_2 are being explored. Real-time dissolved oxygen, pH, cell size, photosynthetic parameters and growth rate were monitored. Besides these physiological responses, transcriptomic and proteomic responses to different phosphorus and carbon availabilities are being investigated by using stranded RNAseq and high- throughput proteomics. The preliminary results indicate that sustained phosphate limitation decreases growth rates ($p < 0.01$) and results in increased cell size ($p < 0.05$) of *Micromonas*. While CO_2 availability had only minor effects on its physiology, these effects were dependent on the phosphorus availability. Elevated CO_2 caused increasing growth rates only at high phosphorus availability and an increase in cell size only at low phosphorus availability.

The global proteomic approach generates an extensive list of proteins that directly linked to phosphate deprivation responses and phosphate resupply responses. 434 differentially expressed proteins (absolute z-score difference 2) were identified. Among those, 262 proteins were more abundant in P-replete condition and 172 proteins were more abundant in the P-deficient condition. Under phosphorus limited condition, proteins involved in starch metabolism, glycolysis and major CHO metabolism, as well as several other pathways were up-regulated, while proteins involved in protein synthesis, protein modification and targeting, and protein degradation were down-regulated. This indicates a sophisticated response to

phosphorus deficiency that strongly affects cellular carbon metabolism. These joint analyses will allow us understanding the phosphorus and carbon physiology of ecologically important microbial species and permit a better prediction of how marine ecosystems may respond to environmental change.

This research is supported by U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Biological Systems Research on the Role of Microbial Communities in Carbon Cycling Program Award No. DE-SC0004765 made in July 2010 with input from other funding sources.

53. Multiple-Element Isotope Probes, NanoSIMS, and the Functional Genomics of Microbial Carbon Cycling in Soils in Response to Chronic Climatic Change

Theresa McHugh[1*] (Theresa.McHugh@nau.edu), Egbert Schwartz,[2] George Koch,[1] Jennifer Pett-Ridge,[3] Peter Weber,[3] Xavier Mayali,[3] **Bruce Hungate**[1]

[1]Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff; [2]Department of Biological Sciences, Northern Arizona University, Flagstaff; [3]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA

Project Goals: Develop a new omics-driven technical approach that couples multiple- element stable isotope probing with nanoscale secondary ion mass spectrometry (NanoSIMS) and phylogenetic microarrays to investigate the functional processes of the microbial community involved in soil carbon cycling. This technique will be used to identify soil bacteria and fungi involved in organic carbon degradation, examine the impacts of shifting environmental variables on their functional processes, and determine if there is a "phylogenetic imprint" on the soil carbon cycle. The capability to quantify *in situ* microbial growth rates both at the community scale and for specific taxa will be the primary advantage of this new methodology.

For the past several decades, connecting biogeochemistry and microbial genomics has been a high priority in microbial ecology. Yet, techniques that actually link element flow and genomic information are scarce. In this project, we are using the Chip-SIP method to measure isotopic composition of major elements (C, N, H, and O) of nucleic acid sequences representing individual microbial taxa. RNA is extracted from an environmental sample after exposure to isotopically labeled substrates. The nucleic acids from the entire microbial community are then exposed to a microarray containing small probes that target the 16S rRNA genes of a large variety of microorganisms so that nucleic acids extracted from the environmental sample bind to matching probes. Then, the entire microarray is placed under a nanoscale secondary ion mass spectrometer, which sequentially analyzes the RNA bound to each probe for isotopic composition. In this way, element flow in the natural environment into individual microbial taxa can be determined.

Our work relates to key unknowns about soil carbon cycling and stability. The methods we are developing will allow us to examine relationships between soil microbial diversity and the processing of soil C. We hypothesize that there is a phylogenetic signal in the microbial biogeochemistry of the soil C cycle, explaining the variation in the degradation of soil organic matter in response to external forcing. The information generated will help to better understand the functional significance of the identity of microorganisms in complex, natural communities.

This work is supported by the Office of Biological and Environmental Research (OBER) of the U.S. Department of Energy (DOE).

54. Microbial community structure and activity during thawing of mineral cryosols of the Canadian High Arctic: meta genomics, transcriptomic and proteomics

Alice Layton[1], Tatiana Vishnivetskaya[1], Archana Chauhan[1], Daniel Williams[1], **Susan Pfiffner[1]**, B. Rekepalli[1], Karuna Chourey[2], **Robert L.Hettich[2]**, Tommy Phelps[1], Maggie C. Y. Lau[4], Brandon Stackhouse[4], Nadia Mykytczuk[3], J. Ronholm[3], G. Lemarche-Gagnon[3], **Lyle Whyte[3]**, **Tullis Onstott[4*]**(tullis@princeton.edu)

[1]University of Tennessee, Knoxville, Tennessee; [2]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]McGill University, Montreal, Quebec;[4]Princeton University, Princeton, New Jersey

Project Goals:

1. Perform ~2 year thawing experiments on well-characterized, intact cores of the Arctic active- layer and permafrost from a proposed reference ecosystem site where CO_2 and CH_4 fluxes, temperatures, humidity, soil moisture, nutrients, microbial diversity and activities and isotopic analyses are currently being measured.
2. Characterize the vertical flux of volatile organic acids, O_2, H_2, CO_2 and CH_4 the isotopic systematics of CO_2 and CH_4 and changes in the transcriptomics, proteomics and C cycle networks in these cores as the permafrost thaws under water saturated and water under saturated conditions.
3. Compare the fluxes measured for the cores with those measured in the field.
4. Based upon these thawing experiments construct a 1D biogeochemical reaction/transport model that predicts the CO_2 and CH_4 release into the atmosphere as permafrost thaws and compare these predictions with observations at the reference ecosystem site.

http://www.princeton.edu/southafrica/permafrost-project/

Permafrost thawing and the resulting microbial decomposition of previously frozen organic C, represent a significant potential positive feedback of CO_2 and CH_4 from terrestrial ecosystems to the atmosphere. However, to date, most studies have focused on relatively high carbon soil sites with less available information regarding microbial activity and the potential for CH_4 generation in low carbon/low nitrogen mineral soils such as those found on Axel Heiberg Island, Nunavut, Canada near the McGill Arctic Research Station. As part of a long-term 18-month thawing experiment (Stackhouse *et al*. 2014), soil samples were collected from 1-meter long intact cores consisting of active-layers and permafrost for metagenomic analyses using a 2 x 100 bp paired- end Illumina protocol. Seventy-seven metagenomes, averaging 5 Gbp/ library and representing a matrix of sequence data spanning 5 treatment conditions, 4 depths and 5 time points (unthawed to 18 months thawed) were annotated in MG-RAST. Taxonomic and functional characterization of the metagenomes indicate specific microbial community differences between the upper active layers and the underlying permafrost layer with opposite depth gradients between alpha- Proteobacteria which were more abundant towards the soil surface and Actinobacteria which were more abundant in the lower layers. Over a 12-month thaw period, the microbial community structure in the upper 5 cm active layer remained remarkably constant. In contrast, the microbial community structure shifted towards higher concentrations of Firmicutes and beta-Proteobacteria with time in the lower 65 cm and permafrost layers, i.e. more similar to that in the upper 5 cm soils. Potential CH_4 cycling pathways primarily consisted of methane oxidizers in the upper layers with a paucity of methanogenic archaea in the lower layers. Other differences in carbon- cycling pathways exist with depth, included aromatic ring oxidation and CO_2 fixation potential in the uppermost layer, and carbon reduction pathways in the permafrost layer.

Nitrogen-cycling pathways also differed by depth with N_2 fixation and denitrification pathways present in the upper layers and nitrification pathways in the permafrost layer. The metagenomic sequences generated during the study are being co-assembled to generate high quality databases in order to better identify peptide sequences in metaproteomes and metatranscriptomes. A co-assembly of sequences from ten 5 cm libraries generated 83,205 contigs >200bp. Within these contigs, nearly full-length sequences for the genes encoding particulate methane monoxygenase (*pmo*C, *pmo*A and *pmo*B) with high similarity to high-affinity Type II methanotrophs were identified (Lau *et al*. 2014). Mapping of ~30 million metatranscriptomic sequences generated from total RNA isolated from the upper 5 cm active layer soils, collected during Arctic summer, resulted in the identification of *pmo* transcripts from surface soils underlying the moss from the polygon interior and polygon trough samples but not in a non-vegetated polygon interior sample. The use of co- assembled contigs also formed the database by which the pmoB protein was found in two active- layer metaproteomic samples from the core thawing experiment. These combined results support parallel *in situ* field gas flux (CH_4, CO_2) and laboratory column analyses indicating these abundant Arctic mineral cryosols are acting as methane sinks rather than methane sources in the moss and wedge samples. Additional analyses of these combined metagenomic, metatranscriptomic and metaproteomic datasets provide a strategy for linking metabolic potential of the soils with metagenomic sequences and will improve our understanding of the microbial community, phytocommunity and ecosystem gradient.

References

M.C.Y. Lau, B.T. Stackhouse, A.C. Layton, A. Chauhan, T.A. Vishnivetskaya, K. Chourey, N.C.S. Mykytczuk, P.C. Bennett, G. Lamarche-Gagnon, N. Burton, J. Ronholm, W.H. Pollard, C.R. Omelon, D.M. Medvigy, R.L. Hettich, S.M. Pfiffner, L.G. Whyte and T.C. Onstott. An active atmospheric methane sink in high Arctic mineral cryosols. Science, *in review*, 2014.

B.T. Stackhouse, T.A. Vishnivetskaya, A.C. Layton, S.M. Pfiffner, N.C. Mykytczuk, L.G. Whyte, N. Saad, T.C. Onstott, Vertical gas fluxes, geochemical characteristics, and microbial community of polygonal active layer and permafrost during progressive spring thaw. Global Change Biology, *in post-review revision*, 2014.

This research was supported by the Genomic Science Program in the Office of Biological and Environmental Research in the U.S. DOE's Office of Science.

55. An Atmospheric CH_4 Sink in the High Arctic and its Implication for Global Warming

C.Y.M. Lau[1], B. Stackhouse[1], D. Medvigy[1], Y. Chen[1], A. Layton[2], T. Vishnivetskaya[2], **S.M. Pfiffner[2], L.G. Whyte[3]**, J. Renholm[3], J. Allen[3], N. Mykytczuk[3], P. Bennett[4], K. Chourey[5], **R.L. Hettich[5], T.C. Onstott[1*]** (tullis@princeton.edu)

[1] Princeton University, Princeton, New Jersey; [2] University of Tennessee, Knoxville, Tennessee; [3] McGill University, Montreal, Quebec; [4] University of Texas, Austin, Texas; [5] Oak Ridge National Laboratory, Oak Ridge, Tennessee

Project Goals:

1. Perform ~2 year thawing experiments on well-characterized, intact cores of the Arctic active-layer and permafrost from a proposed reference ecosystem site where CO_2 and CH_4 fluxes, temperatures, humidity, soil moisture, nutrients, microbial diversity and activities and isotopic analyses are currently being measured.
2. Characterize the vertical flux of volatile organic acids, O_2, H_2, CO_2 and CH_4 the isotopic systematics of CO_2 and CH_4 and changes in the transcriptomics, proteomics and C cycle networks in these cores as the permafrost thaws under water saturated and water under saturated conditions.
3. Compare the fluxes measured for the cores with those measured in the field.
4. Based upon these thawing experiments construct a 1D biogeochemical reaction/transport model that predicts the CO_2 and CH_4 release into the atmosphere as permafrost thaws and compare these predictions with observations at the reference ecosystem site.

http://www.princeton.edu/southafrica/permafrost-project/

Arctic permafrost underlies about 16% of the Earth's surface and contains ~500 Pg of C down to one meter. Organic-rich peatlands (averaging ~4 wt% SOC) comprise 19% of this area, whereas the remaining 81% is permafrost-affected mineral cryosols (0.5-1.5 wt% SOC). Temperatures in the Arctic are predicted to increase ~6°C over the next 100 years which increases the depth of the active layer, the seasonally thawed soil above the permafrost. Thawing permafrost peat deposits (e.g. Stordalen Mire, Sweden) are currently CH_4 sources. Field measurements, intact core studies and microcosm experiments performed by us over the past few years on mineral cryosols associated with ice wedge polygons from Axel Heiberg Island (AHI) in the Canadian high Arctic indicate that they are sinks for atmospheric CH_4.

After 1.5 years of thawing at 4°C, 1 m long intact cores of the active layer and underlying permafrost mineral cryosols collected from ice wedge polygons at AHI continue to exhibit uptake of atmospheric CH_4 even for water saturated cores. The measured core fluxes are consistent with flux measurements performed in the field over the past two years, which range from 0.005 to 0.89 mg CH_4-C/m^2-yr and which have revealed significant differences in the atmospheric CH_4 consumption fluxes between the polygon interiors and the polygon troughs.

Metagenomic analyses reveal a diverse population dominated by aerobic bacterial heterotrophs (Vishnivetskaya *et al*. 2014) of which ~1% is comprised of methanotrophs (Stackhouse *et al*. 2014). Metagenome and metaproteome analyses of these mineral cryosols have revealed the presence of USCα, USCγ and Cluster 1 *pmoA* genes, the three genotypic groups that are recognized as high affinity, atmospheric CH_4 oxidizers (Lau *et al*. 2014; Martineau *et al*. 2014). Our microcosm studies on atmospheric CH_4 uptake rates are consistent with published results from high latitude organic-rich soils

and temperate forest soils and indicate a temperature dependency for the cellular rate of CH_4 oxidization that is approximately twice that reported for methanogenesis. This temperature dependency when combined with annual temperature records from nearby Eureka weather station suggests that these high Arctic ice wedge polygons are significant annual sinks for atmospheric CH_4. Because the maximum atmospheric CH_4 uptake rate coincides with the summer time dips in the recorded atmospheric CH_4 and peaks in $\delta^{13}C$, we propose that seasonal variations in the high latitude atmospheric CH_4 are partially modulated by the activity of atmospheric CH_4 oxidizers. We also propose that this sink will increase with increasing Arctic temperatures and will lessen the interannual increases in atmospheric CH_4.

References

M.C.Y. Lau, B.T. Stackhouse, A.C. Layton, A. Chauhan, T.A. Vishnivetskaya, K. Chourey, N.C.S. Mykytczuk, P.C. Bennett, G. Lamarche-Gagnon, N. Burton, J. Ronholm, W.H. Pollard, C.R. Omelon, D.M. Medvigy, R.L. Hettich, S.M. Pfiffner, L.G. Whyte and T.C. Onstott. An active atmospheric methane sink in high Arctic mineral cryosols. Science, *in review*, 2014.

C. Martineau, Y. Pan, L. Bodrossy, E. Yergeau, L.G. Whyte and C.W. Greer, Atmospheric methane oxidizers are present and active in Canadian high Arctic soils. FEMS Microbiology Ecology, *in review*, 2014.

B.T. Stackhouse, T.A. Vishnivetskaya, A.C. Layton, S.M. Pfiffner, N.C. Mykytczuk, L.G. Whyte, N. Saad, T.C. Onstott, Vertical gas fluxes, geochemical characteristics, and microbial community of polygonal active layer and permafrost during progressive spring thaw. Global Change Biology, *in post-review revision*, 2014.

T.A. Vishnivetskaya, A.C. Layton, M.C.Y. Lau, A. Chauhan, K.R. Cheng, A.J. Meyers, J.R. Murphy, A.W. Rogers, G.S. Saarunya, D.E. Williams, S.M. Pfiffner, J.P. Biggerstaff, B.T. Stackhouse, T.J. Phelps, L. Whyte, G.S. Sayler and T.C. Onstott. Commercial DNA extraction kits impact observed microbial community composition in permafrost samples. *FEMS Microbiol Ecol*., *FEMS Microbiol Ecol*. 87:217-230. doi: 10.1111/1574-6941.12219, 2014.

This research was supported by the Genomic Science Program in the Office of Biological and Environmental Research in the U.S. DOE's Office of Science.

56. Microbial Community and Functional Responses to Rainfall Manipulations in a Prairie Soil

Ari Jumpponen[1*] (ari@ksu.edu), Lydia H. Zeglin[2], Peter J. Bottomley[2], Charles W. Rice[1], Maude M. David[3], Emmanuel Prestat[1,3], Janet K. Jansson[3], Susannah G. Tringe[5], Nathan C. VerBerkmoes[4], Robert L. Hettich[4] and **David D. Myrold[2]**

[1]Kansas State University, Manhattan, KS; [2]Oregon State University, Corvallis, OR; [3]Lawrence Berkeley National Laboratory, Berkeley, CA; [4]Oak Ridge National Laboratory, Oak Ridge, TN; [5]Joint Genome Institute, Walnut Creek, CA

http://cropandsoil.oregonstate.edu/content/meta-omics-analysis-microbia-carbon-cycling-responses-altered-rainfall-inputs-native-prairie-soils

Project Goals: Prairie soils are important stocks of sequestered carbon. However, responses in soil carbon processing to predicted environmental change are still unknown. This project aims to determine soil microbial community and functional responses to ambient and altered precipitation schemes with a specific emphasis on microbial processes that are fundamental to soil carbon dynamics. To do this, the project combines an array of omics tools ranging from targeted locus sequencing to soil transcriptomics and proteomics.

Large pools of carbon (C) are processed and stored in prairie soils: grasslands cover 6.1-7.4% of the earth's land area and store 7.3-11.4% of global soil C. Current global change predictions suggest that precipitation may change little in annual volume, but more in frequency, resulting in more variable regimes across the North American Great Plains with less frequent but larger rainfall events. Between these large pulse events, soil systems will experience extended and often extreme droughts. Despite the importance of prairie soils and therein-residing microbial communities for C sequestration, our understanding of microbial community and functional responses or the resultant C processing to the variability in soil hydrology is cursory at the best. Our current program has assessed soil microbial communities (PLFA- and qPCR-inferred biomass, bacterial 16S RNA and DNA; fungal 28S RNA and DNA), microbial function (respiration, C utilization efficiency, extracellular enzyme activity (EEA)) and microbial gene expression (soil transcriptomic and proteomic analysis, focused on components of C processing pathways) under ambient and altered precipitation schemes. We have taken advantage of the Rainfall Manipulation Plot (RaMP) infrastructure at the Konza Prairie Long-Term Ecological Research site in the Flint Hills region in NE Kansas. The RaMPs provide a long-term replicated (n=6 per treatment) field experiment that aims to mimic predicted shifts in precipitation intervals, while keeping the total precipitation volumes unchanged. In the course of our current program, we sampled soils before, during, and after rainfall events from experimental units that represent ambient rainfall (Ambient) as well as experimental units that experienced a 50% increase in the dry intervals between precipitation events and fewer but larger rainfall resulting in same precipitation volume over the growing season, simulating "droughty" conditions (Altered).

Results to date indicate that rainfall events caused rapid and similar microbial respiration responses in Ambient and Altered treatments. Further, microbial biomass increased rapidly after the rainfall, more so in the Altered plots. These results suggest that increasing precipitation intervals may increase microbial C use efficiency and lead to greater potential for C sequestration belowground. Concurrent with these responses, biomass C:N ratio and fungal:bacterial ratio increased as soil water content decreased; however, our next-generation sequence data suggested only minor community responses. These results suggest that

(i) coarse changes in relative abundance of microbial domains (bacteria vs. fungi) dominate within-domain taxonomic responses, and/or (ii) physiological shifts within the communities are more prevalent than community turnover as a result of altered precipitation frequency and soil water contents.

Both QPCR and gene expression data suggest bacterial sensitivity to rainfall manipulation, in that bacterial 16S rRNA gene abundance was lowest in the driest soils, and relative abundance of mRNAs for oxidative phosphorylation was highest immediately after rainfall (like microbial respiration). Our results also highlight microbial EEA responses. Rainfall induced bulk soil cellulolytic activity and expression of bacterial cellobiose transport genes in moist soils, but not in dry soils. This result suggests rapid processing of available organic matter once soil water contents increase, and that a threshold of microbial respiration and/or labile C availability must be exceeded to induce SOM decomposition.

Beyond identifying microbial mechanisms that contribute to enhanced soil C storage potential in drier soils, the gene expression data also suggest shifts in stress-related and C-processing metabolic pathways. Corresponding proteomes for each soil transcriptome are complete, and are being aligned to better understand microbial function at the protein level in field samples. Also, our efforts to improve resolution and annotation of Omic data likely provide further insight into soil microbial functional responses. Extended dry conditions caused the most notable shifts in both microbial structure and function, highlighting the importance of extreme events and preceding conditions in understanding soil microbial C cycling dynamics. The combination of multi-Omic and activity-level data acquisition confirmed that dynamic soil water content affects microbial C cycling processes at both the cellular and ecosystem levels. In all, our results suggest that, while sensitive to rainfall events and seasonal drying, microbial communities and activities are resilient over multi-annual time scales to shifts in precipitation frequency and express a variety of physiological strategies to cope with drought stress in prairie soils.

57. Transcriptome profiling of *Nitrosomonas europaea* grown singly and in co-culture with *Nitrobacter winogradskyi*

José Pérez[1], Alex Buchanan[1], Rebecca Ferrell[2], Jeff Chang[1], Frank Chaplen[1], Neeraja Vajrala[1], Peter J. Bottomley[1], **Daniel J. Arp**[1], and Luis A. Sayavedra-Soto[1]* (sayavedl@science.oregonstate.edu).

[1]Oregon State University, Corvallis OR and [2]Metropolitan State University of Denver, Denver CO.

URL: http://www.science.oregonstate.edu/bpp/Labs/arpd/

Project Goals: To create predictive models of ammonia--oxidizing bacteria (AOB) and nitrite-oxidizing bacteria (NOB) that incorporate metabolism, the regulatory interactions that influence metabolism, and the signaling network for interaction with the environment. These models will provide powerful tools for understanding responses of nitrifying organisms to different environmental conditions, and predicting how they will behave in response to changes in the environment.

Nitrification is the aerobic microbial process in which ammonia (NH_3) is oxidized to nitrate (NO_3^-). Nitrification can be carried out sequentially in two steps by the action of two groups of mostly chemolithoautotrophic bacteria. In the first step the AOB extract energy for growth from the oxidation of NH_3 to nitrite (NO_2^-). In the second step the NOB oxidize the NO_2^- produced by the AOB to NO_3^- to extract energy for growth. In aerobic ecosystems, these two groups of bacteria tend to work in concert and, in most situations NO_3^- accumulates, rather than NO_2^-.

In a first effort to establish whether there are interactions between AOB and NOB in the process of nitrification, *N. europaea* (AOB) and *N. winogradskyi* (NOB) were co---cultured with NH_4^+ as the sole growth substrate. The transcriptome profile of *N. europaea* grown singly and in co-culture with *N. winogradskyi* was examined and provided clues to the adaptations that might be contributing to differential growth of the co-culture compared to single culture.

Growth parameters and mRNA levels showed discernible differences between growing singly and in co-culture. Co-culture growth in a medium containing 60 mM NH_4^+ resulted in a cell density greater than that of the compounded single chemostat cultures when grown at equivalent 60 mM concentrations of NH_4^+ or NO_2^-, respectively. Oxygen substrate dependent consumption rates (NH_4^+ or NO_2^-) for each nitrifier allowed the estimation of the relative cell density contribution to the co-culture and suggested that the increase in cell density was due mainly to *N. europaea*.

The analysis of the transcriptome of *N. europaea* showed that the mRNA of 726 genes were at different levels between single culture and in co-culture. Compared to single culture, *N. europaea* in co-culture had 279 genes (38%) with higher mRNA levels and 447 genes (62%) at lower mRNA levels.

Examples of genes at lower mRNA levels (Table 1) included genes encoding for biosynthetic functions (flagella synthesis, amino acid synthesis, iron dependent metabolism, carbon fixation related), and genes involved in stress responses (nitrite reduction, glutathione synthesis, DNA repair DNA and oxidative stress). *N. europaea* apparently benefits from the interaction while in co-culture with *N. winogradskyi* compared to when grown singly. Transcriptome and physiological analyses of co-cultures are revealing interactions that go beyond nitrite acting as the growth substrate for NOB. The data is being used to construct predicting models.

Table 1: *N. europaea* mRNA fold changes of selected genes in co-culture.

Locus tag	Function	Gene name	Description	Fold: RNAseq
NE0202	Energy generation related	---	F0F1 ATP synthase	0.84
NE0207				1.28
NE1764	Electron transport chain	*nuo*	NADH dehydrogenase	1.64
NE1767				1.72
NE0102	Electron transport chain	*cytC552*	cytochrome 552	0.82
NE2377	Fe---S cluster	*bolA*	2Fe2S homeostasis	0.70
NE0582	Sulfur uptake	*sbp1*	sulfate/thiosulfate binding protein	0.78
NE0852	Sulfur reduction	*yvgQ*	sulfite reductase subunit beta	1.20
NE0448	NH_3 uptake	*amtB*	ammonia transporter	1.81
NE1919		*cbbQ*		-2.59
NE1920	Carbon metabolism	*cbbS*	RuBisCo	-1.69
NE1926	Carbon uptake	*cynT*	carbonic anhydrase	0.70
NE2149	Glycolate pathway	*cbbZ*	phosphoglycolate phosphatase	-0.40
NE0675	Glycolate pathway	*glcD*	glycolate oxidase subunit	-1.20
NE0589	Carbon fixation	*ppc*	phosphoenolpyruvate carboxylase	-1.20
NE1649	Fatty acid synthesis	*acpP*	acyl carrier protein	1.16
NE0925	Stress: N_2 inhibition	*nirK*	Nitrite reductase: cytochrome C	-9.93
NE1736	NO2--- pump	*nitT/tauT*	nitrate/taurine transport	-1.05
NE0308	NO2--- avoidance	*flgH*	flagellar basal body protein	-1.26
NE0870	Oxidative stress	*sodB*	superoxide dismutase	-0.80
NE0104	Glutathione biosynthesis	---	dihydroxy-aciddehydratase	-1.42
NE1669	Antioxidant	*coq7*	ubiquinone biosynthesis monooxygenase	-1.11
NE1638	Heavy metal pump	*czcA*	cobalt-zinc-cadmium efflux pump	-1.08
NE1721	Iron uptake	---	iron complex receptor	-1.24

58. Metaproteomics reveals key aspects of microbial community mediated carbon cycling in thawing Arctic Permafrost

Karuna Chourey[1], Alice Layton[2], Tatiana Vishnivetskaya[2], Archana Chauhan[2], Maggie C. Y. Lau[3], Brandon Stackhouse[2], Rebecca Sanders[3], **Susan Pfiffner**[2], Satish Myneni[3], **Lyle Whyte**[4], **Robert L. Hettich**[1]* (*hettichrl@ornl.gov*), **Tullis. C. Onstott** [3]

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]University of Tennessee, Knoxville, Tennessee; [3]Princeton University, Princeton, New Jersey; [4]McGill University, Montreal, Quebec

Project Goals:

1. Perform ~2 year thawing experiments on well-characterized, intact cores of the Arctic active- layer and permafrost from a proposed reference ecosystem site where CO_2 and CH_4 fluxes, temperatures, humidity, soil moisture, nutrients, microbial diversity and activities and isotopic analyses are currently being measured.
2. Characterize the vertical flux of volatile organic acids, O_2, H_2, CO_2 and CH_4 the isotopic systematics of CO_2 and CH_4 and changes in the transcriptomics, proteomics, and C cycle networks in these cores as the permafrost thaws under water saturated and water unsaturated conditions.
3. Compare the fluxes measured for the cores with those measured in the field.
4. Based upon these thawing experiments, construct a 1D biogeochemical reaction/transport model that predicts the CO_2 and CH_4 release into the atmosphere as permafrost thaws and compare these predictions with observations at the reference ecosystem site.

http://www.princeton.edu/southafrica/permafrost-project/

Microbial activity plays an important role in the fate of carbon compounds sequestered in permafrost ecosystems. Nonetheless, how the microorganisms influence carbon cycling is not clear due to the paucity of data on cellular activity of indigenous microbial communities in Arctic permafrost. Our work focuses on studying microbial activity in cryosol, (mineral cryosols above the permafrost table) obtained from Axel Heiberg Island (AHI), Canada. We sampled soil cores from different sites within AHI to explore the range of microbial activities in the soil. The microbes and their activities in various soil cores were examined in (1) unaltered state (control), (2) after thawing (4^{o}C and 10^{o}C and (3) after a combination of nutritional amended microcosms (glucose, acetate and lactate) incubated at temperatures above freezing. Microbial proteins were extracted using established protocols and identified with liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). The resulting MS datasets were matched against a compiled protein database assembled from genome sequences of closely related bacteria identified in AHI samples using a 16S rRNA 454 pyrosequencing approach. Metaproteomics of 5g of unaltered cryosols yielded limited protein information, suggesting dormant cryosol microflora and/or low biomass. In contrast, 5g of thawed (4°C or 10^{o}C, 1- 3 months) cryosol layers and/or nutritionally amended microcosms yielded increased protein identifications (~350 proteins), including DNA polymerase, energy metabolism proteins (phosophoribulokinase, malate dehydrogenase etc.), GroEl, DnaK, F0F1 ATP synthase, acetoacetyl CoA reductase, acetoacetyl CoA transferase. In particular, proteome data was obtained for species of *Bradyrhizobium* and known methanotrophs *Methylosinus* and *Methylocystis*, the latter of which are known to express high affinity methane monooxygenase capable of oxidizing atmospheric methane.

The cryosols were characterized with respect to bulk soil organic carbon, as analyzed by C K-edge x-ray absorption near edge structure, and extractable soil organic carbon (OC), as analyzed by ^{1}H nuclear magnetic spectroscopy and electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Results indicated that both the unaltered and thawed soils were highly aliphatic in composition. Although the soils are composed of 1-5% total OC, the amount of extractable soil OC was low (< 10% of the total organic carbon) and the composition did not significantly vary after a year of thawing. The limited variability and low amounts of extractable OC in AHI cryosols suggests limited bioavailability of this OC pool.

To ascertain if increasing amounts of readily bioavailable carbon can promote higher microbial growth in cryosol, another round of experiments was set up (1 gm soil in 10 ml culture media) and incubated at $10^{o}C$ for 3 months with different carbon sources [Glucose, Tryptic Soy Broth (TSB), Potato dextrose broth (PDB) and diluted R2A]. Increases in microbial biomass, as evidenced by increases in the amount of total extracted protein, was noted in all cultures except the relatively low nutrient 1/10 R2A supplemented cultures. MS analysis of the enrichment cultures yielded identification of ~ 1500 proteins spanning across multiple functional categories, indicating the presence of robust, actively dividing cells with varied metabolic pathways.

Metagenomic sequencing using a 2 x 100 bp paired-end Illumina protocol was performed on all enrichment cultures to measure shifts in the microbial community profile due to differences in carbon substrate types. The metagenomic and metaproteomics analyses indicated that *Bacilli* (phylum Firmicutes) was highly enriched in the PDB, glucose and TSB enrichment cultures, *Arthrobacter* (phylum Actinobacteria) was highly enriched in PDB, 1/10 R2A and glucose enrichment cultures, and *Clostridium* (phylum Firmicutes) was highly enriched in PDB, glucose and TSB enrichment cultures. Other genera enriched only in one type of media were *Pseudomonas* in PDB and *Bacteroidetes* class in 1/10 R2A. All media enrichments showed a relative decline in Alpha-Proteobacteria, including the order rhizobiales which is more dominant in the native upper active layer of the cryosol. The deeper cryosol layers revealed an abundance of spore formers belonging to order *Clostridiales*

In conclusion, AHI cryosols represent a unique permafrost system, which is low in utilizable organic carbon leading to a predominantly dormant microbial consortium in cryosol layers. However, upper active layers have a high abundance of Alpha-Proteobacteria (rhizobiales), some of which are capable of utilizing CO_2 and methane. With the gradual warming of cryosols coupled with an increase in greenhouse gases, the upper layer microbial consortia will likely cycle C1 carbon and thus provide utilizable carbon source to the deeper cryosol layers, reviving the dormant microbial community.

This research was supported by the Genomic Science Program in the Office of Biological and Environmental Research in the U.S. DOE's Office of Science.

59. Mechanism of Mercury Binding by Methanobactin from Methylocystis strain SB2

Bipin S. Baral[1], Alan A. DiSpirito[1*] (aland@iastate.edu), Nathan L. Bandow[1], **Jeremy D. Semrau[2]**

[1]Iowa State University, Ames, Iowa; [2]University of Michigan, Ann Arbor, Michigan;

http://sitemaker.umich.edu/methanobactin

Project Goals: Methanotrophs are ubiquitous in the environment, and despite their critical function in many different ecosystems, the biogeochemical factors that affect their activity and community structure are poorly understood. It is known that copper plays a key role in methanotrophic physiology, but the mechanism used by these microbes for copper acquisition was only recently discovered. This compound, methanobactin (mb), is the first example of a "copper-siderophore" or chalkophore. Mb binds many different metals, including mercury. Further, recent data show that different methanotrophs make different forms of mb that have varying metal affinities. The general objectives of this proposal are thus to consider how mb made by different methanotrophs alters the bioavailability of metals of concern to the DOE and how this affects: (1) the physiology, metabolism and gene expression in methanotrophs; (2) the broader microbial community structure and activity in laboratory soil columns, and; (3) the bioavailability of different metals in subsurface environments.

One of the persistent and substantial problems in remediation of hazardous waste sites is the mobilization and uncontrollable transport of radionuclides and heavy metals from these sites to surrounding areas. Some microbially-mediated processes can at least temporarily immobilize and reduce the toxicity of these materials through dissimilatory reduction that leads to precipitation and sorption under anaerobic conditions. As such, microbial-mediated processes can limit the dispersal of these materials and thus also limit the exposure of surrounding areas. Microorganisms, however, have effective and ubiquitous mechanisms to solubilize different metals and that non-specific binding of metals by these biogenic metal chelators may increase their solubility, mobility, and bioavailability. Here we are examining how the metal chelating agents analogous to siderophores in methane-oxidizing bacteria i.e., methanotrophs, binds copper and mercury individually and in mixed metal environments. Such studies will enable us to determine how methanotrophic activity may affect the copper and mercury mobility in subsurface waters, including at DOE sites.

Recent work in our laboratories has identified the genetic basis of mb and that many, but not all methanotrophs can synthesize mb (1-5). Interestingly, mb contains two heterocyclic rings, either imidazole (imi), oxazolone (oxa) or pyrazinedione (pyr) rings with an associated enethiol group, which together are responsible for metal binding (6-8). Given the structure of mb, it is quite possible that some if not all mbs can also bind toxic metals such as mercury and that mb made by one methanotroph may affect the bioavailability of metals to other methanotrophs. Our findings show that mb from *Methylosinus trichosporium* OB3b (mb-OB3b) does indeed bind mercury in addition to copper, and in doing so, reduced toxicity associated with Hg(II) to both α- and γ- Proteobacteria methanotrophs (9. At Hg to mb-OB3b ratios $\leq$ 1.0, Hg is coordinated via the two oxa rings and the associated enethiols groups. At Hg to mb-OB3b ratios $\geq$ 1.0, both oxa ring and associated enethiol group can each bind Hg separately. Interestingly, mercury binding by mb-OB3b was evident both in the presence and absence of copper, despite the fact that mb had a higher affinity for copper due to the rapid and irreversible binding of mercury by mb. Metal analyses indicated that Hg(II), after bound by mb- OB3b, may have been reduced to Hg(0) but was not volatilized. Rather, mercury remained associated with mb, and also was found

associated with methanotrophic biomass. It thus appears, although the mercury-mb complex was cell-associated, mercury was not removed from mb.

In this report, Hg binding by the structurally unique methanobactin from *Methylocystis* strain SB2 (mb-SB2) was examined and compared to mb-OB3b. Mb-SB2 is functionally similar (1,3), but structurally different (7) to mb-OB3b (6). In mb-SB2 one of the oxa rings is replaced by a imi ring and the redox active amino acids in mb-OB3b, Cys, Met, and Try, are replaced with Ala or are missing . Here mb-SB2 is shown to bind the common forms of Hg found in the environment, Hg^{2+} HgCN and CH_3Hg^+ as well as stimulate the solublization of metallic Hg. In general, the binding constrants and spectral; UV-visible absorption, fluorescent and circular dichroism, properties of mb-SB2 differed with different forms of Hg. Hg(II) and HgCN were coordinated by both oxa and imi ring and the associated enethiol groups whereas CH_3Hg^+ was coordinated via the oxa ring and both enethiol groups. The spectral, kinetic and thermodynamic changes following Hg binding by mb-SB2 also differed from the changes associated with mb-OB3b. Like mb- OB3b, copper did not displace Hg bound to mb-SB2 and was not volatilized. However, in constrast to mb-OB3b, mb-SB2 preferentially bound Hg over Cu in mixed metal environments. The perferential binding of Hg over Cu was related to the kinetics of Hg and Cu binding. The results suggest mb can bind Hg, even in environments with low Hg to Cu ratios.

References (1) Bandow, et al. 2012. *J. Inorgan. Biochem.* 110:72-82. (2) Semrau, et al. 2013. *Environ Microbiol*. 15:3077-3086. (3). Vorobev, et al. 2013. *Appl. Environ Microbiol*. 79:5918-5926. (4) Jagadevan & Semrau. 2013. *Appl. Microbiol. Biotechnol*. 97:5089-5096. (5) Vorobev, et al., 2014. *Appl. Environ. Microbiol* (submitted). (6) Kim et al. 2004. *Science* 305: 1612 – 1613, (7) Krentz et al. 2010. *Biochemistry* 49: 10117 – 10130, (8) El Ghazouani et al. 2012. *Proc. Nat. Acad. Sc. USA* 109:8400 – 8404, (9) Choi et al. 2006. *J. Inor. Biochem.* 100: 2150 – 2161.

This research was supported by the Office of Science (BER), U.S. Department of Energy.

60. Metal Uptake by Methanotrophs: Genetic Basis for the Biosynthesis of A Novel Chalkophore and Molecular Spectroscopic Analyses of Mercury Detoxification

Jeremy D. Semrau[1*] (jsemrau@umich.edu), Alexey Vorobev[1], Sheeja Jagedevan[1], Bipin Baral[2], Nathan Bandow[2], and Alan A. DiSpirito[2],

[1]University of Michigan, Ann Arbor, Michigan; [2]Iowa State University, Ames, Iowa

http://sitemaker.umich.edu/methanobactin

Project Goals: Methanotrophs are ubiquitous in the environment, and despite their critical function in many different ecosystems, the biogeochemical factors that affect their activity and community structure are poorly understood. It is known that copper plays a key role in methanotrophic physiology, but the mechanism used by these microbes for copper acquisition was only recently discovered. This compound, methanobactin (mb), is the first example of a "copper-siderophore" or chalkopore. Mb binds many different metals, including mercury.

Further, recent data show that different methanotrophs make different forms of mb that have varying metal affinities. The general objectives of this proposal are thus to consider how mb made by different methanotrophs alters the bioavailability of metals of concern to the DOE and how this affects: (1) the physiology, metabolism and gene expression in methanotrophs; (2) the broader microbial community structure and activity in laboratory soil columns, and; (3) the bioavailability of different metals in subsurface environments.

One of the persistent and substantial problems in remediation of hazardous waste sites is the mobilization and uncontrollable transport of radionuclides and heavy metals from these sites to surrounding areas. Some microbially-mediated processes can at least temporarily immobilize and reduce the toxicity of these materials through dissimilatory reduction that leads to precipitation and sorption under anaerobic conditions. As such, microbial-mediated processes can limit the dispersal of these materials and thus also limit the exposure of surrounding areas. Microorganisms, however, have effective and ubiquitous mechanisms to solubilize different metals and that non-specific binding of metals by these biogenic metal chelators may increase their solubility, mobility, and bioavailability. Here we are examining how the expression of metal chelating agents analogous to siderophores in methane-oxidizing bacteria i.e., methanotrophs, alters the bioavailability of various metals (copper and mercury) and how this: affects the (1) physiology, metabolism and gene expression in methanotrophs; (2) broader microbial community structure and meta-transcriptome, and; (3) bioavailability and risk associated with various metals. Such studies will enable us to determine how methanotrophic activity may affect the structure of subsurface microbial communities as well as the sustainability of subsurface waters, including at DOE sites.

Recent work in our laboratories has identified the genetic basis of mb and that many, but not all methanotrophs can synthesize mb. Interestingly, mb contains two heterocyclic rings, either imidazole, oxazolone or pyrazinedione rings with an associated enethiol group, which together are responsible for metal binding. Given the structure of the rings, it is quite possible that mb can also bind toxic metals such as mercury and that mb made by one methanotroph may affect the bioavailability of metals to other methanotrophs. Our findings show that mb from *Methylosinus trichosporium* OB3b does indeed bind mercury in addition to copper, and in doing so, reduced toxicity associated with Hg(II) to both α- and γ-Proteobacteria methanotrophs.

Interestingly, mercury binding by mb was evident both in the presence and absence of copper, despite the fact that mb had a much higher affinity for copper due to the rapid and irreversible binding of mercury by mb. Metal analyses indicated that Hg(II), after bound by mb, may have been reduced to Hg(0) but was not volatilized. Rather, mercury remained associated with mb, and also was found associated with methanotrophic biomass. It thus appears, although the mercury-mb complex was cell-associated, mercury was not removed from mb.

It was also found that the amount of biomass-associated mercury in the presence of methanobactin from *M. trichosporium* OB3b was greatest for *M. trichosporium* OB3b and least for the tested γ-Proteobacteria methanotroph *(Methylomicrobium album* BG8), suggesting that methanotrophs may have selective mb uptake systems that may be based on TonB-dependent transporters, but that such uptake systems exhibit a degree of infidelity. Further, it was found that the addition of mb from *M. trichosporium* OB3b stimulated the growth of other methanotrophs in the absence of mercury but in the presence of copper. As methanotrophs expressing the particulate methane monooxygenase (the predominant form of methane monooxygenase expressed) require copper for high activity, it may be that methanobactin from *M. trichosporium* OB3b increased the bioavailability of copper, thereby increasing activity of pMMO in other methanotrophs. If so, this suggests that mb made by one methanotroph may actually be taken up by others (as also suggested by the mercury uptake data). Collectively, these studies raise several interesting questions, including do all methanotrophs in a mixed community produce mb, or do some species act as "cheaters" and rely on mb made by other microbes to meet copper requirements for metabolism or for detoxification of metals such as mercury? How do methanotrophs that make mb ensure that they are able to effectively compete with such cheaters for copper? It appears that the mb uptake systems have some infidelity, and that *in situ*, such infidelity may be optimized as a general competition strategy.

Ongoing work is characterizing the transcriptome of both α- and γ-Proteobacteria methanotrophs under a range of copper concentrations to determine how copper affects overall gene expression in methanotrophs. We have also sequenced the genome and transcriptome of the novel facultative methanotroph, *Methylocystis* strain SB2, grown with different carbon sources. We have also recently collected soil samples from the Integrated Demonstration Site of the DOE Savannah River Site where methanotrophs are known to exist. We are in the process of characterizing the initial microbial community composition and will also construct a series of soil columns to characterize how mb affects: (1) copper and mercury mobility and bioavailability in the presence of soils from this site, and; (2) dissolution of soil-associated minerals. The resultant effects on the broader microbial community structure and function will be determined via metagenomics and metatranscriptomics.

Publications (1) Bandow, et al. 2012. *J. Inorgan. Biochem*. 110:72-82. (2) Semrau, et al. 2013. *Environ Microbiol*. 15:3077-3086. (3). Vorobev, et al. 2013. *Appl. Environ Microbiol*. 79:5918-5926. (4) Jagadevan & Semrau. 2013. *Appl. Microbiol. Biotechnol*. 97:5089-5096. (5) Vorobev, et al., 2014. *Appl. Environ. Microbiol* (submitted).

This research was supported by the Office of Science (BER), U.S. Department of Energy.

61. Pathways to Carbon Liberation: a Systems Approach to Understanding Carbon Transformations and Losses from Thawing Permafrost

Virginia Rich[1], Gene W. Tyson[2], Matthew Sullivan*[1] (mbsulli@arizona.edu), Steve Frolking[3], Changsheng Li[3], Ruth Varner[3], Greg Hurst[4], Nathan Verberkmoes[5], Patrick Crill[6], Jeffrey Chanton[7], and **Scott Saleska**[1]

[1]University of Arizona, Tucson; [2]University of Queensland, Brisbane, Australia; [3]University of New Hampshire, Durham; [4]Oak Ridge National Labs, Tennessee; [5]New England BioLabs, Ipswich, Massachusetts; [6]University of Stockholm, Sweden; [7]Florida State University, Tallahassee

Project Goals: Our objective is to discover how microbial communities mediate the fate of carbon in thawing permafrost 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 a chronosequence of permafrost thaw in subarctic Sweden.

Abstract: The fate of carbon (C) in thawing permafrost is an outstanding challenge of modern biogeochemistry and climate change. Permafrost C pools are large (~1700 PgC), and C dynamics of thawing permafrost complex: old C decomposes as it is liberated from thawing permafrost as CO_2 or CH_4, with a significant fraction cycling through lake sediments, even as new C accumulates due to thaw-initiated ecological succession.

Microbes mediate C dynamics in thawing permafrost, but a mechanistic understanding of how to scale microbial population dynamics, genomic potential, and expression to ecosystem-scale processes is missing. A key question is: What is the interplay of microbial communities and organic matter chemical structure in the decomposition/preservation of organic C across a thaw gradient? And intriguingly, what (if any) is the role of phage (viruses that infect prokaryotic cells) in mediating these processes? Viruses appear to play a large role in driving oceanic microbial functions, but these phenomena are virtually unstudied in terrestrial systems. This endeavor linking microbial and viral dynamics, organic geochemistry and trace gas production will improve models of C cycling in permafrost systems, and clarify the fate of C under future climates.

This builds upon existing work by our team on methane cycling in Stordalen Mire, Sweden, along a permafrost thaw chronosequence encompassing permafrost palsas and their initial collapse stage, thawing bog sites (dominated by Sphagnum spp.), fully- thawed and inundated fen sites (dominated by Eriophorum spp or Carex spp.), and lakes. Our newly begun work will use cutting edge technologies in both biogeochemistry and molecular microbial ecology to advance systems biology research on microbial carbon cycling through: (a) systems-level mapping of chemical states and ages of organic matter along thaw gradients (via FT-ICR-MS and ^{14}C analysis) to associated microbial communities, biochemical potentials, and activities (via meta-genomics, -transcriptomics, –proteomics, and viral genomics), and to CO_2 and CH_4 fluxes; (b) bioinformatics designed to simultaneously enhance the DOE Knowledgebase, and (c) integrated ecosystem C-cycle modeling testable by soil organic chemical and microbial data.

Funded by the DOE Genomic Science Program of the United States Department of Energy Office of Biological and Environmental Research, grants DE-SC0004632 and DE-SC0010580.

62. Rapid Binning and Metabolic Profiling of Subsurface Microbial Community Metagenomic Data via an Interactive Online Knowledgebase

Brian C. Thomas, Andrea Singh, Laura A. Hug* (laura.hug@berkeley.edu), and **Jillian Banfield**

University of California - Berkeley, Berkeley CA

Project goals:
Our goal is to develop an online resource, ggKbase, for the management and rapid analysis of microbial community data, including metagenomic and transcriptomic sequences and proteomic profiles. The platform offers a custom system for genome binning, metabolic pathway curation, and community composition analysis. Importantly, these functions are combined with visualization tools that help guide analysis and interpretation of complex datasets, as well as visually summarize the results for publication. The ggKbase platform is useful for researchers interested in the rapid binning and metabolic assessment of metagenomic data.

URL: http://ggkbase.berkeley.edu/

Abstract:
The scientific focus of our research is to develop a predictive understanding of the microbiology of the subsurface, including the roles of microorganisms in carbon cycling. The questions metagenomic datasets can address are multiple: what organisms are present in a community, what is their relative abundance, and how do these change over time and space or in relation to changing physiological conditions. Beyond taxonomic profiling, metagenomics allows prediction of the metabolic potential of the community, including which processes may be occurring and through which intermediates. Transcriptomics and proteomics can be used to confirm the active metabolic processes and identify the key members within a community for nutrient cycling. Datasets are growing in size in line with increases in sequencing capacity, and the communities being examined are likewise scaling in complexity and diversity. Thus, there are major challenges relating to efficiency and accuracy of data analysis.

The ggKbase platform is designed as an interactive, online environment for the simultaneous and partially automated analysis of hundreds of genomes and associated omic data. ggKbase offers a suite of tools for rapid assignment of assembled fragments to organisms (binning) and metabolic prediction. Intrinsic to these analysis options are interactive visualizations that allow fast and intuitive examination of the data, streamlining the analysis process and providing summaries that can be used in publications. We will present several examples of applications of ggKbase for binning (Fig. 1) and metabolic analysis (Fig. 2) of complex subsurface communities, containing hundreds to thousands of organisms, many of which are highly novel compared to existing genome databases.

To date, the ggKBase platform has been extensively used for metagenomic analyses on complex communities from aquifer sediment, acid mine drainage, and the human gut. These projects have allowed tool development informed by user requirements from manual investigations of the data. The ggKbase is now being transitioned to support community users on a wide scale.

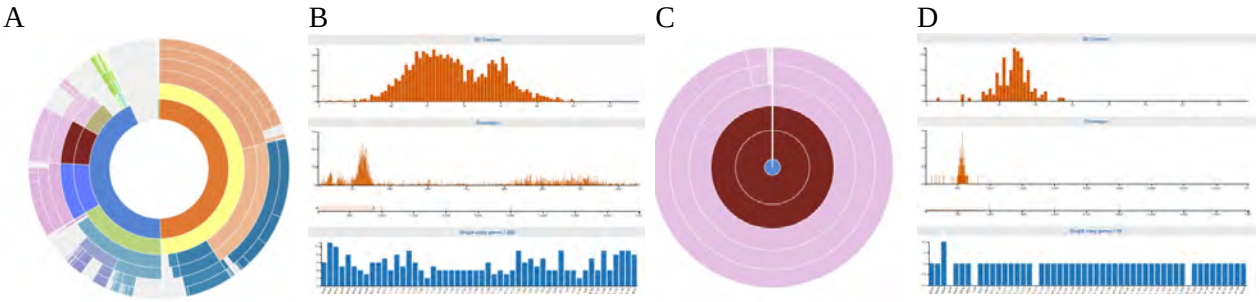


Figure 1: A) Interactive display of the phylogenetic profile of unbinned metagenomic data. Concentric rings relate to the taxonomic level to which a contig can be assigned based on the phylogenetic profile of encoded genes. B) Histograms profile the GC content, coverage, and single copy gene content of the whole dataset. Phylogenetic groups can be chosen, and subsets of these groups assigned to bins using interactive controls that select specific GC content and/or coverage ranges. For example, clicking a region in the taxonomic wheel in A (C) shows that a bin with consistent GC and coverage profile has been generated, and the single copy gene profile indicates that the putative bin could contain a relatively complete genome (D).

Figure 2: Example of the use of dataset-wide lists to profile both metabolic potential and genome completeness in binned data

63. Developing mass spectrometry approaches enabling multi-'omic' analyses of the dynamics, mechanisms, and pathways for carbon turnover in grassland soil under two climate regimes

Chongle Pan[2] (panc@ornl.gov), Tami Swenson[3], Susan Spaulding[1], Zhou Li[2], David Burstein[1], Stefan Jenkins[3], Brian Thomas[1], Mary Power[1], Robert Hettich[2], Blake Suttle[4], Trent Northen[3], **Jill Banfield**[1]

[1]University of California, Berkeley, [2]Oak Ridge National Laboratory, [3]Lawrence Berkeley National Laboratory, [4]Imperial College, London

Project goals: The goal of this project is to investigate how altered rainfall impacts carbon stored in grassland soil by characterizing effects on microbially mediated carbon decomposition and nitrogen cycle processes. The project will use genome targeted metagenomics, stable isotope resolved metaproteomics, and community metabolomic analyses. The study will leverage a long term ecological research site in California that features an ongoing experimental manipulation of water inputs to mimic shifting precipitation regimes that are relevant to climate change scenarios. New data visualization tools will be developed for the DOE Systems Biology Knowledgebase to facilitate integration of the information generated in the study and scaling of data to higher order process understanding.

Our program is focused on a decadal-scale climate manipulation experiment on a meadow in the Angelo Coastal Range Reserve in northern California. The reserve is a relatively pristine coastal environment with a Mediterranean climate, characterized by wet winters and long summer droughts. In environments such as this, changes in water abundance and the timing of rainfall may profoundly impact soil conditions, vegetation type, and ecological characteristics such as soil carbon storage and carbon turnover. For this reason, 13 years ago, UC Berkeley researchers initiated a series of experiments in which the intensity and seasonality of rainfall was manipulated (in 24 of the 36 experimental plots) to replicate two scenarios predicted by the current climate models. Prior research showed that extension of the spring rainy season significantly increased plant primary productivity (Suttle *et al*. 2007) and that the period immediately following the first autumn rainfall is critical for carbon breakdown.

Microorganisms are the primary agents responsible for breakdown and turnover of soil carbon compounds. After six years of rainfall manipulation, our group conducted preliminary 16S rRNA soil microbial community surveys that revealed seasonal responses that displayed the potential impact of extreme weather events (Cruz *et al.* 2009). The longer-term responses and soil system behavior are completely unknown. We will study the period around the first fall rainfall event, when soil-associated carbon fixed during spring growth is rapidly metabolized, focusing on climate manipulations that differ 1) in the amount of spring rainfall (above-ground carbon stocks) and 2) on the period of time following the first fall rain events (soil microbial communities) and 3) soil depth.

We have developed powerful approaches to investigate community metabolism using metaproteomics and metabolomics in model Acid Mine Drainage Biofilms that we are actively working to extend to study the Angelo soils. Soil presents tremendous challenges to protein and metabolite analysis and so our current focus has been on developing extraction protocols using the Angelo soil samples. In 2013, eight total sets of samples were collected before and after the first fall rainfall. For each set, samples and 10-cm-deep soil cores were collected from four depths (0-40 cm) within two 0.25-m^2 quadrants and flash-frozen in the field in a dry ice/ethanol bath. Samples were aliquotted for metagenomic sequencing, community proteomics measurements, and metabolomics analysis.

A set of metagenomic samples from before the first fall rain and three different days afterward were extracted with a modified PowerMax Soil DNA Isolation Kit protocol. The 10-minute vortexing step was replaced with a non-shaking water bath at 65 ^{o}C and only gentle inversions every 10 minutes. This produced approximately 1 microgram per gram of soil, of high-quality DNA (260/280 ratios of 1.6-1.8) of fragment sizes larger than 23 Kb. These samples are currently in the JGI sequencing queue.

To optimize protein extraction for Angelo soils, we compared two methods: the SDS-boiling method (Chourey *et al*., 2010) and the commercial MoBio NoviPure kit. SDS-boiling co- extracted contaminants that complicated protein concentration estimation and sample clean-up. It was critical to minimize the presence of soil contaminants in proteome samples, such as humic acids, because they can create many challenges for liquid chromatography and mass spectrometry analysis. On the other hand, the NoviPure kit produced colorless clean samples that are compatible with downstream preparation. The yield of protein extraction was estimated to be approximately 50 microgram of protein per gram of soil. The obtained samples were analyzed by shotgun proteomics using LTQ Orbitrap mass spectrometer. The quality of ion chromatograms was comparable to those from well-characterized systems, such as the Acid Mine Drainage microbial communities. Many peptides were detected with good chromatographic peaks and informative tandem mass spectra. This indicated effective protein extraction from Angelo soils, which will enable in-depth characterization of soil community proteomes.

For metabolomics analysis, total metabolites (extracellular and intracellular) were measured by fumigating soil samples with chloroform vapors for 24 hours. Unfumigated soil was used to analyze extracellular metabolites. Extractions were performed comparing buffers and the addition of organic solvents for soil metabolite analysis. Liquid chromatography/ mass spectrometry and gas chromatography/ mass spectrometry were used to evaluate the various extraction conditions. Fumigation is found to have the strongest effect on the range and abundance of metabolites detected, producing approximately two times more molecular features than unfumigated soil. Using this approach we detected a wide range of metabolites including amino acids, mono- and disaccharides, nucleosides, lipids and other metabolites consistent with our program goals. .

The developed technologies provided the foundation for integrated –omics analyses of a representative grassland soil microbial community. The results will characterize dynamic metabolic processes within this community during a key period of carbon turnover.

This work conducted was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-SC0010566 and by the Eel River Critical Zone Observatory (NSF 12-575) for Angelo Coastal Range Reserve.

64. Metagenomics-Enabled Predictive Understanding of Microbial Communities to Climate Warming: Results from Long Term Soil Incubations

Rosvel Bracho[1], **Edward A.G. Schuur[1], Susan Natali[2],** Elaine Pegoraro[1], **Liyou Wu**[3], Huaqun Yin[3], Kai Xue[3], Mengting Yuan[3], Jin Zhang[3], Ye Deng[3], **Zhili He**[3], Jim Cole[4], **James M. Tiedje**[4], **Konstantinos Konstantinidis**[5], Junyi Liang[3], Katherine Todd-Brown[3], **Yiqi Luo**[3] **and Jizhong Zhou**[3]

[1]Department of Biology, University of Florida, Gainesville, FL; [2]Woods Hole Research Center, Falmouth, MA; [3]Institute for Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, Norman, OK; [4]Center for Microbial Ecology, Michigan State University, East Lansing, MI;[5]Center for Bioinformatics and Computational Genomics and School of Biology, Georgia Institute of Technology, Atlanta, GA.

http://ieg.ou.edu

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 climate warming, altered precipitation and soil moisture regime a in a tundra and temperate grassland ecosystem; (ii) determine temperature sensitivity on recalcitrant C decomposition; (iii) determine microbiological basis that is underlying temperature sensitivity of recalcitrant C decomposition; and (iv) develop integrated bioinformatics and modeling approaches to scale information across different organizational levels. This study focused on using laboratory incubations of soils as an isolated system to understand the influence of microbial processes on the release of C, and their response to changes in temperature.

Long-term aerobic laboratory incubations were performed at two temperatures (15 & 25°C) to determine the temperature sensitivity of microbial respiration (Q_{10}) in soils from a field warmed tundra (Alaska, AK) and a temperate grassland site (Oklahoma, OK). Three different layers were incubated from the AK site: two surface layers (0 – 15 cm & 15-25 cm with high carbon content) and a layer perennially frozen (45-55 cm) while 0-25 cm soils were incubated from the grassland site. The OK site included soils from control and field-warmed plots combined with a root exclusion treatment over 8 years. Carbon fluxes were measured periodically over the course of the incubation and total C respired at each time (C_T) was fitted with a two-pool C model to estimate pool sizes and decay rates of fast (C_F) and slow (C_S) decomposing C fractions. We estimated Q_{10} using three different methods: (1) a short-term method ($Q_{10\text{-}S}$), where soils were exposed to 6 different temperatures ranging from 5 to 30°C over the period of <1 week, while measuring C fluxes over this range at 14, 100 and 280 days during the long-term incubation; (2) the equal carbon method ($Q_{10\text{-}EC}$), which estimates the ratio of time that it takes for a soil to respire a given amount of carbon at each incubation temperature through the incubation experiment. $Q_{10\text{-}EC}$ was estimated for fast and slow decomposing C pools; (3) a separate modeling approach, which is an assimilation method of estimating Q_{10} for bulk soils and decomposing C pools from incubation data. The method includes a one pool (1P) model, a two pool (2P) model, a three pool (3P) model, and a three pool model with C transfer between the pools (3PX). The microbial community phylogenetic and functional composition, structure, and dynamics were analyzed by 16S rDNA sequencing and functional gene array GeoChip 5.0.

Results from the AK soils showed that after one year of incubation, C_T in the top 15cm could be as high as 25% and 15% of the initial soil C content at 25°C and 15°C incubations respectively. The fast decomposing C pool (C_F) accounted for up to 5% of the initial C content in the top 15 cm soils. Both, C_T and C_F decreased with depth but no field warming effect was detected. Overall average turnover time for C_F was ~ 60 days at these laboratory conditions.

Turnover time for C_S varied from 10 years in top soils to ~60 years in soils near the permafrost at 15°C incubation and decreased by half at the higher incubation temperature. Total C respired (C_T) in soils from OK accounted for 5% at 15°C and 10% of the initial soil C content at 25°C after one year of incubation. Fast decomposing C reached up to 4% of the initial soil C content at 25°C with an average turnover time of up to 100 days. The overall short term $Q_{10\text{-}S}$ for AK soils was 2.55 ± 0.03. Neither treatment nor depth nor day of incubation, nor incubation temperature had any effect on Q_{10}-s, however, interactions of field treatment*depth and treatment*day of incubation were significant. Estimated $Q_{10\text{-}EC}$ was 1.2 ± 0.4 and 2.2 ± 0.06 for C_F and C_S, respectively, with no significant differences with field treatment, varying depth or incubation temperature. For OK soils, the overall $Q_{10\text{-}S}$ was 3.2 ± 0.3. The AIC analysis indicated that the two-pool model was the best fit given the incubation data over one year.

The results of the dissimilarity analysis of the 16S rRNA amplicon sequencing data from the OK soils showed warming effect on microbial communities. After two weeks at 25°C, samples from the deep collar had more abundant populations of *Acidobacteria, Actinobacteria, Crenarchaeota, Firmicutes, Gemmatimonadetes, Nitrospira, Planctomycetes, Verrucomicrobia*, but *Armatimonadetes, Bacteroidetes, Chloroflexi, Proteobacteria* were less abundant. When incubated at 25°C for three months, more differences were observed; however after nine months incubation, the community differences had disappeared. At the bacterial phylum level, no significant warming effect was observed for the soils outside of the deep collar across all sampling time points. These results indicate that the warming effects were more significant when roots were excluded and for those populations responding to labile carbon. GeoChip results showed significant differences in microbial functional diversity in the OK soils. For the incubations at 25˚C, most of the carbon degradation genes had higher abundances in the warmed soils outside deep collars after two weeks, but lower abundance after nine months. With the soils inside deep collars, the carbon degradation genes had lower abundances across the three time points. Warming effects were much less for the incubation at15°C.

The dissimilarity analysis of 16S rRNA sequences was also done for the CiPEHR soils. A significant warming effect was observed in the soils from 0-15 cm when incubated at 25°C for two weeks. In those samples, *Proteobacteria, Actinobacteria, Acidobacteria, Chloroflexi, Firmicutes,* and *Gemmatimonadetes* had higher abundances in the warmed soils, while *Planctomycetes, Chlamydiae, Armatimonadetes, Verrucomicrobia, Bacteroidetes* had lower abundances. Significant differences in microbial population abundances between warmed and control soils from 0-15 cm were also observed when incubated at 15°C after three months, and the soils from >35 cm after 9 months. From GeoChip analyses, differences were observed among soil depths, between treatments, incubation temperatures, and incubation time. For the incubations at 25°C, consistent differences in carbon degradation genes between warmed and un- warmed soils from 15-25 cm were observed along the incubation time points.

This work is supported by the US Department of Energy, Biological Systems Research on the Role of Microbial Communities in Carbon Cycling Program (DE-SC0004601).

65. Live-cell Micropatterning for Analysis of Inter-cellular Behavior in Multi-species Systems

Nathaniel Cady[1,*], William F. Hynes[1], William R. Harcombe[2], Daniel Segre[3], and **Christopher J. Marx**[4,5]

[1]SUNY College of Nanoscale Science & Engineering, Albany, NY, USA (ncady@albany.edu), [2]University of Minnesota, Minneapolis, MN, USA [3]Boston University, Boston, MA, USA [4]Harvard University, Boston, MA, USA [5]University of Idaho, Moscow, ID, USA

Microorganisms exist in dynamic environments where nutrients, energy sources and signals are often in flux. Modeling cellular behavior within such environments is complex, and requires a fundamental understanding of substrate dynamics (i.e., concentration, location, and diffusion rate) and cellular responses to these substrates (i.e., uptake, biomass conversion, growth, efflux). We have used the open-source software platform COMETS (Computation of Microbial Ecosystems in Time and Space) to model metabolite resource sharing in multi-species bacterial systems and to generate predictions of spatiotemporal growth responses. To validate predictions made by COMETS, we have developed a microscale cell printing technology to produce complex patterns of bacterial micro-colonies. Micro-colonies printed with this system can be arranged two-dimensionally on agarose-based media to assess distance-dependent growth phenomena between multiple species. We have demonstrated the utility of this approach using combined COMETS modeling and cellular micropatterning for a synthetic consortium of an *E. coli* and *S. enterica* under conditions that require metabolite sharing between these organisms.

Micropatterning results strongly correlate with COMETS predictions, demonstrating the potential of this approach as a validation tool for computational modeling of bacterial growth within the context of resource sharing.

66. Mechanisms leading to the evolution of novel glucose utilization by *Shewanella oneidensis* MR-1

Lon M. Chubiz[1*] (lchubiz@fas.harvard.edu) and **Christopher J. Marx**[1,2]

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138;
[2]Department of Biology, University of Idaho, Moscow, ID 83844 http://marxlab.rc.fas.harvard.edu

Project goals: Our broader project goals are to integrate novel experimental techniques and computational approaches with a tractable, model syntrophy to predictably understand the role of spatial structure on the function and evolution of microbial communities that mediate metal reduction. This syntrophic microbial community is composed of an engineered, lactate fermenting *Escherichia coli* and *Shewanella oneidensis* MR-1, a preeminent model for anaerobic metal reduction. A limitation to this effort has been an incomplete characterization of the metabolic capabilities of *S. oneidensis*. Under common laboratory conditions, *S. oneidensis* rapidly evolves to utilize glucose as a carbon and energy source. The goal of this study has been to elucidate the molecular mechanisms underlying this novel phenotype, to refine the existing genome-scale metabolic model for *S. oneidensis*, and to explore options to restrict use of glucose by *S. oneidensis* in future experiments.

Carbon metabolism is historically a defining feature in microbial species identification and characterization. In the past decade, with the rapid increase in microbial genomic information, use of carbon utilization as a species identifier has lost favor in light of molecular-based phylogenetics. However, this same information has fueled the construction of numerous genome-scale metabolic models, and provided an unprecedented look into the role microbial metabolic diversity plays in global carbon and nutrient cycling; changing how we perceive the impact of microbes on the environment. That being said, we often view carbon metabolism at as a fixed, intrinsic characteristic of a microbe. What is apparent from meta-genomic data and experimental evolution studies that the suite of carbon substrates a micro-organism consumes is not necessarily a static regimen, but rather a plastic, malleable phenotype. Here, we explore one such metabolic innovation; the molecular mechanisms that lead to the evolution of novel glucose utilization by *Shewanella oneidensis* MR-1 – a preeminent model for diversity in anaerobic respiration.

One of the defining characteristics of *S. oneidensis*, in contrast to many other Shewanellaceae, is its inability to catabolize complex carbon substrates such as glucose, or other carbohydrates. This has lead to much speculation regarding the ecological niche of *S. oneidensis.* Under aerobic growth conditions, we and a number of other researchers have observed spontaneous mutations resulting in emergence of glucose catabolism (Glu^+). To date, the underlying mechanisms driving this change in phenotype, and their consequences on cellular metabolism, have remained unknown. Via whole genome re-sequencing of Glu^+ mutants, we discovered the recurrent source of the Glu^+ phenotype was the result of a large deletion causing constitutive expression of genes involved in *N*-acetyl glucosamine (GlcNac) metabolism. We show that these genes are necessary for glucose transport and phosphorylation of glucose to glucose-6-phosphate – allowing for subsequent glycolysis. Through ^{13}C-labeling experiments, we found that the resulting glucose-6-phosphate was oxidized through the Entner-Doudoroff Pathway, confirming speculation of an incomplete Embden-Meyerhof-Parnas Pathway in *S. oneidensis*. To explore other possible mutations that may result in a Glu^+ phenotype, we performed saturating transposon mutagenesis. In addition to mutations in GlcNac metabolism, we identified several insertions proximal to putative carbohydrate transporters or carbohydrate kinases.

These data suggest a common mechanistic theme leading to novel glucose metabolism – coopting divergent carbon transporters and kinases to produce glucose-6-phosphate. Collectively, these results shed light on cryptic carbon metabolism in *S. oneidensis*, providing additional data to further refine genome-scale metabolic models in this important model organism. More broadly, we demonstrate a framework to explore latent metabolisms in other environmentally relevant microbes that may have profound impacts on global carbon and nutrient cycling.

67. In Situ Correlated Imaging of Chemically Communicating Microbial Communities

Eric Lanni[1]* (lanni1@illinios.edu), Rachel Masyuko[2], Callan Driscoll[3], Nameera Baig[2], Joshua Shrout[3], Jonathan Sweedler[1] and **Paul Bohn**[2]

[1]Department of Chemistry and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801
[2] Department of Chemistry and Biochemistry and Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556
[3] Department of Civil and Environmental Engineering and Earth Sciences and Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556

Project Goal: Our goal is to develop correlated mass spectrometry imaging (MSI) and confocal Raman microscopy (CRM) analytical techniques and apply them to model systems that closely mimic the rhizosphere. The biological system we are developing involves two microbes and a plant root. We seek to understand the biological processes occurring at the single cell and interactions and communication events among multiple cells at the molecular level. By spatially mapping the distribution of various molecular constituents within the model system using both MS and CRM imaging, we are able to acquire complementary information that gives great insights on: (1) the function of different molecular messengers found within microbial communities, (2) key molecular changes that occur as microbial species transition from single cells to communities, and (3) plant-microbe interactions.

http://www.bohnresearchgroup.com

Multiplex, label-free, and non-targeted chemical mapping of sample surfaces can be performed with mass spectrometric imaging techniques such as matrix-assisted laser desorption ionization and secondary ion mass spectrometry (MALDI and SIMS) as well as confocal Raman microscopy (CRM). CRM provides 3D-resolved chemical information based on the molecular vibrations of the components, while MSI yields highly specific chemical maps of the sample surface by direct detection of ionized molecules. Correlated imaging with MSI and CRM – applying these methods to the same sample and combining the resulting data with high spatial fidelity – has the potential to reveal information about molecular composition, distributions, and temporal changes that is not available from either method used alone.

Correlated imaging offers great potential for elucidating the behavior of bacteria. Many bacteria naturally aggregate and adhere to surfaces forming biofilms – multicellular communities held together by a self-produced extracellular matrix. Bacteria within biofilms live in a complex microbial community that exhibits primitive homeostasis, a circulatory system, and metabolic cooperativity, and imparts multiple benefits such as increased antibiotic resistance and enhanced virulence.[1] Corresponding to differences in phenotype, bacteria growing in a biofilm show distinct gene expression profiles relative to planktonic cells[2, 3] and also respond differently to environmental perturbations.

Spatially and temporally mapping the chemical composition of bacterial biofilms can provide valuable insight into the functions of molecular constituents, how they respond to their immediate microenvironment, the mechanisms of biofilm formation, and ultimately how they may be controlled or engineered to be useful.

Initial efforts in this project focused on the development of a correlated MS/CRM imaging platform. We have addressed the technical challenges associated with correlating information from two different spectral imaging techniques. Specifically, we have (1) developed sample preparation protocols that are amenable to both MSI and CRM experiments, (2) implemented a workflow that preserves the sample integrity for both analyses, and (3) developed a spatial registry method that is compatible with both CRM and MSI and precisely demarcates microscopic regions of interest as well as the sample orientation.

The new technologies for correlated MS and CRM imaging described above have been applied to study the biofilm formation process in the bacterium *Pseudomonas aeruginosa*. By characterizing planktonic cells and biofilms using CRM and MSI we are able to compare their molecular compositions and observe the major changes in molecular composition associated with the biofilm formation process. Of special interest are quinolone signalling molecules involved in biofilm production as well as multi- functional secreted glycolipid (rhamnolipid) compounds which act as surfactants, growth cues, and virulence factors.[4] CRM of *P. aeruginosa* biofilm revealed bands at 1030 cm^{-1} and 1068 cm^{-1} that are not present in the spectra of planktonic cells and indicate the presence of glycolipids, likely rhamnolipids. MSI analysis of the same sample confirmed the presence of rhamnolipids, identifying multiple rhamnolipid congeners. Furthermore, unique distributions of the rhamnolipid congeners were observed by MSI, which may indicate differing functions during biofilm growth.[5]

Our current and ongoing work focuses on extending the capabilities of correlated MS and CRM imaging to study microbial attachment on plant roots, as well as visualizing additional compounds of interest within bacterial biofilms.[6]

The authors acknowledge funding from the Department of Energy Office of Science (BER) through grant DE-SC0006642.

References

1. D. Lopez, H. Vlamakis and R. Kolter, *Cold Spring Harbor Perspectives in Biology*, 2010, **2**.
2. V. E. Wagner, D. Bushnell, L. Passador, A. I. Brooks and B. H. Iglewski, *Journal of Bacteriology*, 2003, **185**, 2080-2095.
3. J. W. Costerton, Z. Lewandowski, D. E. Caldwell, D. R. Korber and H. M. Lappinscott, *Annual Review of Microbiology*, 1995, **49**, 711-745.
4. Abdel-Mawgoud, F. Lépine and E. Déziel, *Applied Microbiology and Biotechnology*, 2010, **86**, 1323-1336.
5. J. Tremblay, A.-P. Richardson, F. Lépine and E. Déziel, *Environmental Microbiology*, 2007, **9**, 2622-2630.
6. R. Masyuko, E. Lanni, J.V. Sweedler, and P.W. Bohn *Analyst* 2013, **138**, 1924-1939.

68. Multidimensional Chemical Analysis of Microbial Communities by Integrating Fluorescence, Vibrational and Mass Spectrometric Microspectroscopic Imaging

Rachel Masyuko[1*](rmasyuko@nd.edu), Sarah Melton[2], Amber Bible[2], Eric Lanni[3], Callan Driscoll[4], Joshua Shrout[4], Jonathan Sweedler[3], Jennifer Morrell-Falvey[2], Mitchel Doktycz[2] and **Paul Bohn**[1]

[1] Department of Chemistry and Biochemistry and Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556 [2] BioSciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831 [3] Department of Chemistry and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801
[4] Department of Civil and Environmental Engineering and Earth Sciences and Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556

Project Goal: Our goal is to develop a multi-modal analysis program by integrating fluorescence, Raman, and matrix-assisted laser desorption ionization MS imaging (MALDI MSI) microspectroscopies and apply them to a system that closely mimics the rhizosphere. Specifically, we are developing a three-component system comprised of two microorganisms and a plant root. We seek to understand the biological interactions, processes, and communication events that occur at the intra- and inter-species level between the microorganisms as well as the interactions occurring between the plant root and the individual microorganisms by following the spatial and temporal characteristics of the various biomolecules within the system.

http://www.bohnresearchgroup.com

Bacteria are the most abundant organisms on the earth and play significant roles in processes such as nitrogen and carbon cycling, mass transfer of transition metals, and degradation of organic contaminants in the environment. Many bacteria naturally aggregate and adhere to surfaces forming biofilms, which are multicellular communities held together by a self-produced extracellular matrix. Bacteria within biofilms live in a complex microbial community that exhibits primitive homeostasis, a circulatory system, and metabolic cooperativity.[1] Corresponding to differences in phenotype, bacteria growing in a biofilm show different gene expression profiles than planktonic cells.[2] As a result, bacterial cells within biofilms respond differently to environmental perturbations than their planktonic counterparts. Biofilms can convey either detrimental or beneficial effects, depending on the setting. Characterizing and spatially mapping the chemical and molecular composition of bacterial biofilms would provide valuable insight into the behaviour of biofilm-forming bacteria including the function of molecular chemical constituents, how the bacteria respond to external cues contained in the chemical or physical environment, the mechanisms of biofilm formation and maturation, and how they may be managed.

Despite the great importance of answering these questions, there are a relative dearth of methods that can provide spatiotemporal chemical information that would inform them. Thus, in this work, fluorescence, Raman and matrix-assisted laser desorption ionization mass spectrometry imaging (MALDI MSI) microspectroscopies have been applied for the analysis of three different bacterial species; *Pseudomonas* GM41, *Pseudomonas* GM17 and *Pantoea* YR343. Both biofilms and planktonic bacteria have been characterized for the three species allowing us to (1) visualize and identify different molecular species in their composition, (2) differentiate biofilms and planktonic bacteria based on their molecular composition, and (3) differentiate the different species based on their molecular and chemical composition.[4] These powerful characterizations are made possible by the use of multivariate statistical analysis.

Further work has focused on *Pantoea* YR343, which was isolated from the rhizosphere of poplar and displays a robust root colonization phenotype. Colonies of *Pantoea* YR343 have a yellowish color due to the production of carotenoids. Consistent with this, the Ramam spectrum of *Pantoea* YR343 displays prominent peaks at 1150 cm^{-1} and 1525 cm^{-1}, suggesting that the color may be due to the presence of a carotenoid. To test this, we deleted the *crtB* gene, which encodes a phytoene synthase that is responsible for converting geranylgenanyl pyrophosphate to phytoene, an early step in the biosynthesis of β-carotene. A *Pantoea* mutant strain lacking *crtB* grows with similar kinetics as a wildtype strain, but displays a whitish color, consistent with the loss of pigment. The Raman spectrum from the *crtB* mutant lacks the prominent peaks at 1150 cm^{-1} and 1525 cm^{-1}, consistent with our hypothesis that these peaks arise from a carotenoid species, as predicted. Further analysis to confirm the exact identity of the carotenoid molecule is underway using LDI. Future studies are also aimed at using multimodal analyses to examine microbial colonization of plant roots.

The authors acknowledge funding from the Department of Energy Office of Science (BER) through grant DE-SC0006642 and the Genomic Science Program, Plant-Microbe Interfaces Project (ORNL).

References

1. Lopez D, Vlamakis H, Kolter R *Cold Spring Harbor Perspectives in Biology* **2010**, *2*.

2. Costerton JW, Lewandowski Z, Caldwell DE, Korber DR, Lappinscott HM *Annual Review of Microbiology* **1995**, *49*, 711-745

3. Cogdell RJ, Frank HA *Biochimica et Biophysica Acta*, **1987**, *895*,63-79

4. Masyuko R, Lanni E, Sweedler JV, Bohn PW *Analyst* **2013**, *138*, 1924-1939.

69. From Structure to Function: Metagenomics-Enabled Predictive Understanding of Soil Microbial Feedbacks to Climate Warming

Kai Xue[1], Mengting Yuan[1], **Liyou Wu**[1], **Zhili He**[1], Ryan Penton[2], **Jim Cole**[2], **James M. Tiedje**[2], Rosvel Bracho-Garrillo[3], **Edward A.G. Schuur**[3], Chengwei Luo[4], **Konstantinos Konstantinidis**[4], Junyi Liang[1], **Yiqi Luo**[1], and **Jizhong Zhou**[1,*](jzhou@ou.edu)

[1]Institute for Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, Norman, OK; [2]Center for Microbial Ecology, Michigan State University, East Lansing, MI; [3]Department of Biology, University of Florida, Gainesville, FL; [4]Center for Bioinformatics and Computational Genomics and School of Biology, Georgia Institute of Technology, Atlanta, GA

http://ieg.ou.edu

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. Towards this goal, we are pursuing the following objectives: (i) To determine the responses of microbial community structure, functions and activities to climate warming, altered precipitation, soil moisture regime and/or clipping in the tundra and temperate grassland ecosystems; (ii) To determine the temperature sensitivity and substrate priming on recalcitrant C decomposition; (iii) To determine microbiological basis underlying temperature sensitivity of recalcitrant C decomposition; 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 the Knowledge Base (KBase).

Ecosystem responses to climate warming in tundra. We have examined the tundra ecosystem responses to climate warming after one and half year in an experiment site of Carbon in Permafrost Experimental Heating Research (CiPEHR), which was established in September, 2008 at a moist acidic tundra site in Interior Alaska (AK). In this site, snow fences were used to increase soil temperature, coupled with early spring snow removal to control the snow-water equivalents in both warming and control plots. Our results showed that soil warming treatment increased both soil temperature and moisture significantly. In addition, the proportion of labile C pool 2 (mainly cellulose) was higher under warming than control. Furthermore, the gross primary productivity (GPP) was significantly increased by warming, mainly driven by the enhanced growth of graminoid. Warming also extended the length of growing season through earlier bud break (first appearance of bud) and delayed senescence. The ecosystem respiration (ER) was significantly higher under warming, but the net ecosystem exchange (NEE) did not differ between warming and control. Together, these results indicated that warming substantially altered plant and soil properties at the permafrost ecosystem.

We have used integrated metagenomic technologies to analyze the responses of microbial communities in experimental warming site of CiPEHR in tundra ecosystem. Although α- diversity of GeoChip-based microbial functional community did not differ significantly between warming and control, the Inverse Simpson index based on both 16S and 28S rRNA sequences decreased marginally significantly by warming. Also, three complimentary non-parametric multivariate statistical tests (ANOSIM, Adonis, and MRPP) based on various distance indexes (Euclidean, Horn and Bray) and

detrended correspondence analysis consistently revealed that functional community structure differed substantially between warming and control, although significant differences were not detected with sequence data. Specifically, stimulations of broad groups of pathways/functional genes/populations involved in degrading both labile and recalcitrant C in permafrost were observed, which could potentially result in accelerated C decomposition. Moreover, warming stimulated not only functional genes involved in aerobic but also anaerobic microbial processes, which could lead to greater positive climate feedbacks by releasing more CO_2, CH_4 and N_2O. Furthermore, warming significantly enhanced functional genes involved in nutrient cycling processes and hence may result in higher nutrient availability. Overall, our results imply that soil microbial community was sensitive to short-term warming in tundra ecosystem we tested and highlighted the importance of microbial community mediated ecosystem feedbacks to climate warming in permafrost.

Ecosystem responses to climate warming in temperate grassland. We have investigated the ecosystem responses to climate warming in an experimental warming facility in a temperate grassland (tall grass prairie) ecosystem in the US Great Plains in Central Oklahoma (OK) after one and half years. Results showed that warming significantly increased the soil temperature, but decreased the soil moisture significantly. In addition, the proportion of labile C pool 2 (mainly cellulose) in total soil organic C was increased by warming at 39%. Moreover, though the increase of GPP and the decrease of ER were not significant, the NEE was significantly decreased by warming, representing more C were sequestrated in the ecosystem. Together, these results indicated that warming changed some measured soil attributes, but not plant community.

We have also used metagenomic technologies, e.g. GeoChip and 16S rRNA amplicon sequencing, to analyze the responses of microbial communities at the experimental warming site in the temperate grassland ecosystem. Results showed that both functional and phylogenetic structures of soil microbial communities did not alter significantly by warming. Only a small portion of functional genes involved in carbon, nitrogen, phosphor and sulfur cycling showed significant differences between warming and control. Interestingly, the magnitudes of changes in these functional genes induced by short-term warming (one and half year) in the temperate grassland ecosystem were much lower than in the tundra ecosystem as mentioned above. Overall, these results reveal that soil microbial communities in the temperate grassland we tested were less sensitive to short-term warming.

Development of amplicon-sequencing approaches for uncovering functional gene diversity. Specific functional genes involved in nitrogen cycling and fungal community composition have been targeted for 454 amplicon pyrosequencing. The targeted genes included the fungal 28S LSU region, Archaeal *amoA* for ammonia oxidation, *nirK* and *nirS* for nitrite reduction, *nosZ* for nitrous oxide reduction and *nifH* for nitrogen fixation. Our findings showed that warming in temperate grassland significantly affects the overall community structure for *nifH*, *nirK* and *nosZ*. In addition, overall richness increased significantly for all genes except Archaeal *amoA*. Only *nirK* evenness, richness and overall structure were found to change in tundra.

This work is supported by the US Department of Energy, Biological Systems Research on the Role of Microbial Communities in Carbon Cycling Program (DE-SC0004601).

70. Linking Iron and Nitrogen Cycling in *Anaeromyxobacter dehalogenans*

Jenny R. Onley[1*] (jmerryfi@utk.edu), Silke Nissen[1,2*](snissen@utk.edu), Karuna Chourey[3], Xiaoxin Liu[1], Robert Hettich[3], **Kirsti. M. Ritalahti**[1], **Joanne Chee-Sanford**[4], **Konstantinos Konstantinidis**[5], **Robert A. Sanford**[6], and **Frank E. Löffler**[1,2,7]

[1]University of Tennessee, Knoxville; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge; [3]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge; [4]US Department of Agriculture-Agricultural Research Service, Urbana, Illinois, [5]Georgia Institute of Technology, Atlanta, Georgia, [6]University of Illinois, Champaign; [7]Center for Environmental Biotechnology, Univ. of Tennessee, Knoxville

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 expression of genes contributing to N-transformation is required to link desirable (i.e., N-retention) and undesirable (i.e., N_2O emission) activities with measurable microbial parameters. Linking molecular- and organismal-level information with environmental factors that control N- and C-turnover are desirable to interpret field-scale observations, and predict the impact of land management practices on greenhouse gas (N_2O, CO_2) emissions. Such integrated approaches generate novel information at multiple scales of resolution and contribute to system-level understanding of key nutrient cycles in soils. In the present work, we examined the contributions of the common soil bacterium, *Anaeromyxobacter dehalogenans* to N-cycling under different conditions, and examined the *c*-type cytochromes expressed by this organism.

Anaeromyxobacter dehalogenans strains gain energy from the reduction of a variety of electron acceptors including oxidized metals and nitrogen species. *A. dehalogenans* lacks the ability to denitrify (i.e., *nirK* and *nirS* are absent), but reduces nitrate via the dissimilatory nitrate reduction to ammonium (DNRA) pathway, with nitrite as an intermediate. Interestingly, *A. dehalogenans* strains possess a complete and functional "atypical" *nos*Z gene cluster conferring the ability to reduce N_2O to dinitrogen (1). The addition of 1.0 mM nitrite to ferric iron-grown cultures resulted in stoichiometric conversion of nitrite to N_2O via an abiotic mechanism (i.e., chemodenitrification). In contrast, nitrite added to fumarate-grown cultures did not result in N_2O formation.

Abiotic control vessels containing medium without cells and amended with nitrite and ferrous iron (as ferrous chloride) resulted in N_2O production. These observations suggest that the abiotic reaction of ferrous iron with nitrite led to N_2O formation (i.e., chemodenitrification). The N_2O produced in live cultures was reduced by *A. dehalogenans* to dinitrogen. Cell enumeration using quantitative real-time PCR (qPCR) demonstrated growth with abiotically produced N_2O in cultures that had reduced ferric to ferrous iron. Despite the absence of key denitrification genes (i.e., *nirK* and *nirS*), *A. dehalogenans* contributed to the conversion of nitrite (a DNRA intermediate) to dinitrogen via a coupled abiotic-biotic process involving ferrous iron, the end product of ferric iron reduction catalyzed by the same organism. Ferric iron minerals are common to soils and sediments suggesting that coupled abiotic-biotic processes (i.e., chemodenitrification followed by enzymatic N_2O reduction) contribute to N-cycling and affect N_2O flux.

Additional studies attempted to link *c*-type cytochrome expression profiles with *A. dehalogenans* ecophysiology. *c*-type cytochromes are key components for electron transfer to terminal electron

acceptors in *A. dehalogenans* and other metal-reducing populations. Deletion mutant and biochemical studies identified a number of *c*-type cytochromes involved in electron transfer to oxidized metals in *Shewanella* spp. and *Geobacter* spp.; however, only a fraction of the entire *c*-type cytochrome pool has been functionally characterized. The genome of *Anaeromyxobacter dehalogenans* strain 2CP- C encodes 69 *c*-type cytochromes, including two *nrfA* gene copies (Adeh_0910, Adeh_2902) and two *nrfH* gene copies (Adeh_0911, Adeh_2903) involved in DNRA. To identify and subsequently characterize unique *c*-type cytochromes involved in soluble and amorphous ferric iron reduction and N-metabolism, *A. dehalogenans* strain 2CP-C was grown in batch cultures with ferric citrate, ferric oxyhydroxide, and nitrate. Fumarate-grown strain 2CP-C cells were included as a control. Whole cell lysates were subjected to trypsin proteolysis and analyzed using a biphasic LC-MS/MS (Liquid chromatography-tandem mass spectrometry) setup. Distinct *c*-type cytochrome expression patterns were observed in cells grown with the different electron acceptors. Interestingly, several *c*-type cytochromes, including NrfA (Adeh_2902) were expressed in cells grown with ferric iron or nitrate but could not be detected in control cultures grown with fumarate. The second *nrf*A gene copy (Adeh_0910) was only expressed in cells grown with nitrate.

These results indicate that the *A. dehalogenans* strain 2CP-C uses the same *c*-type cytochromes for electron transfer to nitrate and to ferric iron, thus linking iron and N- cycling. Further the data illustrate that the role of a microorganism or a microbial community in N-cycling cannot be predicted based on gene content (i.e., the genome sequence or metagenomic datasets) alone, and confirm that non-denitrifiers possessing atypical *nosZ* genes contribute to N_2O consumption (1). In experimental systems, *A. dehalogenans* effectively contributes to denitrification, even without enzymes converting nitrate to N_2O, by coupling biotic with abiotic reactions. The contributions of non- denitrifying populations such as *Anaeromyxobacter* spp. to complete denitrification (i.e., N_2 formation) and associated C-turnover in soil ecosystems is being explored.

Reference:

1. 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. **Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils.** *Proc. Natl. Acad. Sci. U. S. A.* **109**: 19709-14.

This research was supported by the US Department of Energy, Office of Biological and Environmental Research, Genomic Science Program, Award DE-SC0006662.

71. Detecting Nitrous Oxide Reductase (nosZ) Genes in Soil Metagenomes: Method Development and Implications for the Nitrogen Cycle

Luis Orellana-R (lhorellana@gatech.edu),[1*] Luis Rodriguez-R,[2] Steve Higgins,[3] **Joanne Chee-Sanford,[4] Robert Sanford,[5] Kirsti Ritalahti,[3,6,7] Frank Löffler,[3,6,7]** and **Konstantinos Konstantinidis[1,2]**

[1]School of Civil and Environmental Engineering, [2] and School of Biology, Georgia Institute of Technology, Atlanta, Georgia; [3]Department of Microbiology, University of Tennessee, Knoxville, Tennessee; [4]US Department of Agriculture—Agricultural Research Service, Urbana, Illinois; [5]Department of Geology, University of Illinois, Urbana, Illinois; [6]Bioscience Division, Oak Ridge National Laboratory, Oak Ridge; and Department of Civil and [7]Environmental Engineering, University of Tennessee, Knoxville, Tennessee.

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 expression of genes contributing to N-transformation is required to link desirable (i.e., N-retention) and undesirable (i.e., N_2O emission) activities with measurable microbial parameters. Linking molecular- and organismal-level information with environmental factors that control N- and C-turnover can predict the impact of land management practices on greenhouse gas (N_2O, CO_2) emissions. Such integrated approaches generate novel information at multiple scales of resolution and contribute to system-level understanding of key nutrient cycles in soils. In the present work, we developed and applied bioinformatic approaches to study the abundance and diversity of the nitrous oxide reductase gene, currently known as the only gene directly involved in the reduction of N_2O to an innocuous form, N_2.

Abstract: The anthropogenic fixation of N_2, by means of the Haber-Bosch process, has led to the overuse of synthetic, nitrogen-based fertilizers in agriculture. As a consequence of the increased nitrogen content in soils, the atmospheric N_2O concentration increased nearly 20% relative to preindustrial era levels. Microbial processes including ammonium oxidation, dissimilatory nitrate reduction to ammonium, and primarily denitrification contribute to N_2O emissions. The key enzyme for mitigating N_2O emissions is the nitrous oxide reductase (NosZ), which catalyzes N_2O reduction to N_2 and is generally attributed to denitrifying microorganisms. We have recently described a novel group of "atypical" functional NosZ encoded on the genomes of denitrifiers and incomplete denitrifiers, most of which were missed in previous PCR-based surveys (Sanford et al., PNAS 2012). We analyzed the abundance and diversity of both types of *nosZ* genes in whole-genome shotgun metagenomes obtained from sandy and silty-loam soils in Illinois that typify the Midwest US corn belt, frequently used in bioenergy production. We tested different algorithms and defined appropriate cut-offs for detecting typical and atypical *nosZ* fragments based on *in silico* generated (mock) metagenomes. Based on the determined cut-offs, more than 71 distinct reference representatives (obtained from clustered sequences at 95% amino acid identity), encoding typical and atypical NosZ, were detected in both soil types. Remarkably, more than 70% of the total *nosZ*-encoding reads in both soils were classified as atypical. About 12% of the total *nosZ* reads were taxonomically assigned to the *Anaeromyxobacter* genus, indicating the potential relevance of these organisms for N_2O reduction. Further analyses indicated that atypical *nosZ* genes outnumbered typical counterparts in most publicly available soil metagenomes from various important soil ecosystems in North America, underscoring the ecological importance of atypical *nosZ* in soils. Our work provides a bioinformatic strategy to detect target genes in data-rich short-read metagenomes and has implications for better understanding N-cycling in soils.

References:

1. 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. **Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils.** *Proc. Natl. Acad. Sci. U. S. A.* **109**: 19709–14.
2. C. Luo, L. M. Rodriguez-R, and K. T. Konstantinidis. **A User's Guide to Quantitative and Comparative Analysis of Metagenomic Datasets**. Methods Enzymol. 531:525- 47. 2013.

***Funding Statement**: This research was supported by the US Department of Energy, Office of Biological and Environmental Research, Genomic Science Program, Award DE-SC0006662.*

72. Cultivation and PCR-based Approaches Elucidate the Functional Diversity of Soil Fungal Populations Contributing to Nitrogen Cycling in Soils

Steven A. Higgins[1]* (shiggin8@utk.edu), Allana Welsh[2], **Robert A. Sanford**[2], **Joanne Chee- Sanford**[3], Luis Orellana-R[4], **Konstantinos Konstantinidis**[4], **Kirsti. M. Ritalahti**[1], and **Frank E. Löffler**[1,5]

[1]University of Tennessee, Knoxville, TN; [2]University of Illinois, Urbana, IL[3]; US Department of Agriculture-Agricultural Research Service, Urbana, IL; [4]Georgia Institute of Technology, Atlanta, Georgia, [5]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

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 expression of genes contributing to N-transformation is required to link desirable (i.e., N-retention) and undesirable (i.e., N_2O emission) activities with measurable microbial parameters. Linking molecular- and organismal-level information with environmental factors that control N- and C-turnover are desirable to interpret field-scale observations, and predict the impact of land management practices on greenhouse gas (N_2O, CO_2) emissions. Such integrated approaches generate novel information at multiple scales of resolution and contribute to system-level understanding of key nutrient cycles in soils. In the present work, we developed and applied molecular approaches to study the diversity of fungi contributing to nitrous oxide production within two field sites in Illinois. The current study addresses the need for a better understanding of the contributions of fungal populations to N-turnover and associated C-flux in soils. Specifically, molecular tools were designed and applied to assess the fungal diversity and the abundance of fungal p450nor genes implicated in fungal denitrification and N_2O production.

The kingdom Fungi is estimated to comprise anywhere from 700,000 to 1,500,000 species. Of the known diversity, a majority is housed within the phyla Ascomycota and Basidiomycota. These taxa are distinguished by their activities in soils, such as nutrient transport and retention, engaging in beneficial symbiotic associations with plants, and enhancing overall soil health. Although recent efforts suggest that fungi play relevant roles for nutrient turnover in soils, their involvement in soil geochemical transformations is poorly understood. Anthropogenic activities (i.e., large-scale agriculture and industrialization) over the past two centuries have affected soil microbial activity, and in turn have led to increased greenhouse gas emissions from soil. Filamentous Ascomycetes and Basidiomycetes are key contributors to the degradation of plant- derived organic matter (e.g., lignin, cellulose), and affect soil carbon flux (1, 2). Moreover, members of these taxa have been implicated in denitrification, the conversion of nitrate/nitrite to gaseous products (N_2O, N_2) (3). Though relevant contributions of fungi to C- and N-turnover in soils have been recognized, molecular tools to selectively target fungal taxa and their functional genes involved in geochemical cycling are lacking. To address these shortcomings, 214 fungal isolates capable of using nitrate or nitrite as the sole nitrogen source in liquid medium were obtained from two physicochemically distinct soil sites in Illinois. The majority (70%) of the isolates were denitrifiers and produced N_2O from nitrite. Phenotypic characterization distinguished 15 morphotypes represented by isolates from both field sites. For functional characterization of denitrifying isolates, degenerate PCR primers amplifying an approximately 650-bp fragment of the fungal *p450nor* gene, which is responsible for N_2O production, were designed. The primer set amplified the *p450nor* gene from many denitrifying fungal isolates, and to date 13 *p450nor* genes were cloned and sequenced. To assess the diversity and dynamics of soil fungi from the two different soil types, automated ribosomal intergenic spacer analysis (ARISA) was conducted on samples collected over defined spatial and temporal scales (e.g., with depth and across seasons). The average number of ARISA fragment sizes representing unique operational

taxonomic units (OTUs) was higher in the well-drained sandy soil (n=95) than in the clay-containing silt loam (n=76). Overall fungal communities were significantly distinct across soil depths at any time of year, but assemblages shifted seasonally within depth. The distinction between fungal communities across spatiotemporal scales was more prominent in sand than silt loam. The ribosomal intergenic spacer regions from the 15 distinct fungal morphotypes were cloned and sequenced. This sequence information will guide the design of qPCR assays to target fungal populations and their dynamics in terms of abundance and activity. In concert with soil metadata analysis (e.g., soil moisture, pH, temperature), these efforts will quantify the fungal contribution to N- and C-turnover in agricultural soil ecosystems.

References:

1. **Rousk J, Brookes PC, Bååth E**. 2009. Contrasting Soil pH Effects on Fungal and Bacterial Growth Suggest Functional Redundancy in Carbon Mineralization. Appl. Environ. Microbiol. **75** :1589–1596.
2. **Fontaine S, Henault C, Aamor A, Bdioui N, Bloor JMG, Maire V, Mary B, Revaillot S, Maron PA**. 2011. Fungi mediate long term sequestration of carbon and nitrogen in soil through their priming effect. Soil Biol. Biochem. **43**:86–96.
3. **Shoun H, Fushinobu S, Jiang L, Kim S-W, Wakagi T**. 2012. Fungal denitrification and nitric oxide reductase cytochrome P450nor. Philos. Trans. R. Soc. Lond. B. Biol. Sci. **367**:1186–94.

This research was supported by the US Department of Energy, Office of Biological and Environmental Research, Genomic Science Program, Award DE-SC0006662.

73. From Genomes to Metabolomes: Interspecies Interaction in the Archaeal System *Ignicoccus-Nanoarchaeum* and in other Nanoarchaeota

Louie Wurch[1], Richard J. Giannone[2], Michelle Tigges[3], Brian Tripet[3], Tim Hamerly[3], Valérie Copié[3], Brian Bothner[3], Robert Hettich[2], **Mircea Podar***[1,2] (mpodar@utk.edu)

[1] University of Tennessee, Knoxville TN; [2] Oak Ridge National Laboratory, Oak Ridge, TN; [3] Montana State University, Bozeman MT.

Project Goals: Archaea harbor specific genomic, biochemical and membrane level adaptations that enable them to thrive under chronic energy stress. Nanoarchaeota are hyperthermophilic archaea dependent on direct interaction with other archaea and appear to have pushed that limit even further by becoming nutritional and energetic parasites. Two such systems have been characterized so far, one in our laboratory (Podar et al, 2013). This project integrates comparative genomics with molecular microbiology, proteomics, transcriptomics and metabolomics to understand the mechanisms and evolutionary history of such archaeal relationships. Principles and approaches resulting from this work will also be applied to studying more complex microbial consortia and syntrophic associations.

The marine crenarchaeon *Ignicoccus hospitalis* supports the propagation on its surface of the nanoarchaeote *Nanoarchaeum equitans*. We used concerted proteomic and transcriptomic analyses to probe into the functional genomic response of *I. hospitalis* as *N. equitans* multiplies on its surface. The expression of over 97% of the genes was detected at mRNA level and over 80% of the predicted proteins were identified and their relative abundance measured by proteomics. These indicate that little if any genomic information is silenced during the growth of *I. hospitalis* in the laboratory. The primary response to *N. equitans* was at the membrane level, with increases in relative abundance of most protein complexes involved in energy generation as well as that of several transporters and proteins involved in cellular membrane stabilization. Similar up-regulation was observed for genes and proteins involved in key metabolic steps controlling nitrogen and carbon metabolism, although the overall biosynthetic pathways were marginally impacted. Proliferation of *N. equitans* resulted, however, in selective down regulation of transcription factors, replication and cell cycle control genes as *I. hospitalis* shifted its physiology from its own cellular growth to supporting that of its ectosymbiont/parasite (Giannone et al, 2014). Metabolomic analysis (MS and NMR) has identified specific small molecules and metabolic profiles associated with *N. equitans* proliferation. The majority of metabolites were observed to be present at lower concentrations in the co-culture, suggesting that there is a higher energy demand when *N. equitans* is present. Additionally, trehalose, an osmolite that protects cells from osmotic stress, was readily observed in the co-culture. This compound may mitigate changes in cell volume and membrane-level stress likely to occur during association. Correlations with the proteomic data aim at developing an integrated molecular model for the interspecies interaction.

The interaction between the two organisms is specific. Co-cultivation with other related species of *Ignicoccus* revealed that *N. equitans* can attach to, but is not able to actively divide, on their surface. This suggests that specific mechanisms of interaction, likely involving membrane proteins have evolved to a high degree of specificity. Comparative genomics, cellular and membrane biochemistry, and are used to study interspecies interaction and metabolic transfers.

Comparisons with a Nanoarchaeota system we isolated from a terrestrial thermal system from Yellowstone are used to understand the broader mechanisms of interactions between symbiotic/parasitic archaea and their hosts.

References

1. Podar M, Makarova KS, Graham DE, Wolf YI, Koonin EV, Reysenbach AL. (2013) Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park. ***Biol Direct***. Apr 22;8:9.

2. Giannone RJ, Wurch LL, Heimerl T, Martin S, Yang Z, Huber H, Rachel R, Hettich RL, Podar M. (2014) Life on the edge: Functional genomic response of *Ignicoccus hospitalis* to the proliferation of *Nanoarchaeum equitans*. Submitted.

This research was supported by a grant from the U.S. Department of Energy, Office of Biological and Environmental Research (DE-SC0006654).

74. Greenhouse gas emissions from fertilized plots of bioenergy crops in Eastern Washington

Manmeet Pannu[1], Kelley Meinhardt[1], Anthony Bertagnolli[1], **Sally Brown[2]**, **Steve Fransen[3]**, **David Stahl[1]** and **Stuart Strand[1]*** (sstrand@uw.edu)

Civil and Environmental Engineering[1] & College of Forestry[2], University of Washington, Seattle, WA 98195, Department of Crop and Soil Science[3], Washington State University at Prosser, WA 99350

Project Goals: The impact of an expanded bioenergy industry on soil and atmospheric chemistry remains unknown. Nitrogen fertilization is required for enhanced plant growth but nitrifying organisms in soil oxidizes ammonium to nitrate and produces greenhouse gases as a byproduct. Ammonia oxidizing Archaea (AOA) are often observed in high abundances in soils, and have been implicated as the dominant assemblages responsible for the first step of nitrification (ammonia oxidation to nitrite). However, information regarding their role in nitrification in soils impacted by biofuels cultivars remains unknown. The goals of this project were to assess whether differing soil types and nitrogen delivery strategies 1) altered the rates of differing greenhouse gas emissions 2) impacted the community abundance, structure, and diversity of nitrifying assemblages 3) altered broader components of microbial diversity and metabolic activity associated with the nitrogen cycle.

Cellulosic ethanol (biofuel) is proposed as an alternative to fossil fuels. Nitrogen fertilizer is the highest energy input for biofuel production and is often applied in excess to enhance crop yields. However, excess nitrogen application leads to nitrogen leaching and to increased emissions of atmospherically active nitrous oxide (N_2O) and nitric oxide (NO). Nitrous oxide is a greenhouse gas with a global warming potential 300 times greater than carbon dioxide. Nitric oxide has an indirect impact on earth's radiative balance by catalyzing tropospheric ozone formation. Thus, in order to minimize adverse environmental impacts of crop production, efficient nitrogen delivery is essential. Our primary research objective was to determine the influence of plant species, soil type, N source, and microbial community on greenhouse gas emissions. Our studies focused on two experimental plots of irrigated switchgrass, a potential biofuels feedstock, in Eastern Washington at two sites, near Prosser (PR) and Paterson (PatN), having different soil characteristics. Experimental treatments included varying rates of chemical and biosolids fertilizer application, and the passive delivery of nitrogen via intercropping with alfalfa (N-fixer). During the 2013 field season (April-October), gas samples were collected immediately after irrigation at roughly monthly intervals and the flux of nitrogen oxides calculated using Fick's law.

Nitrous oxide flux data for the 2013 season revealed distinctive site differences. A greater flux was measured at the PatN site (sandy, pH ~6) than at the PR site (silt loam, pH ~8), suggesting an effect of soil texture and/or pH on nitrous oxide emissions. Previous published studies also observed higher N_2O fluxes at sites with acidic soils. Average N_2O fluxes were elevated immediately after fertilization and irrigation: 12.5 g N_2O-N /ha/d at PR and 15 g N_2O-N /ha/d at PatN. Further, biosolids application (PatN site) contributed to much higher emissions (25 g N_2O- N /ha/d) than observed for inorganic fertilizer or in control plots. Intercropping with alfalfa resulted in a maximum flux of 4 g N_2O-N /ha/d. To mimic field treatments in a controlled laboratory environment, undisturbed soil cores (without switchgrass) from the PatN field site were analyzed using a dynamic chamber system. N_2O fluxes increased immediately after watering soils amended with biosolids and inorganic fertilizer at field rates, but returned to ambient concentrations over time as the soils dried. Higher emissions from biosolids (250 N_2O-N g/ha/d) and inorganic fertilizer (75 N_2O-N g/ha/d) than observed in the field site was likely due to the absence of

switchgrass plants in these initial laboratory studies (now being replicated using switchgrass planted soils).

Our results clearly demonstrated that soil type, and the form and rate of nitrogen application, together greatly influence N_2O emission rates. Greater N_2O fluxes were observed with biosolids and in acidic soil (PatN) relative to slightly alkaline soil (PR). Complementary molecular studies have revealed that ammonia oxidizing archaea (AOA) are the dominant ammonia-oxidizing population in these intensively managed soils, greatly outnumbering ammonia-oxidizing bacteria. Since these molecular studies also revealed a distinctive shift in AOA population type - *Nitrososphaera* (54d9 clade) dominant at PatN and *Nitrosotalea* (subcluster 1.1) dominant at PR – physiological differences among AOA genotypes may also influence emission rates of atmospherically active gases.

75. Diversity of ammonia-oxidizing archaea in soils under managed and native conditions

Stuart Strand[2], Steve Fransen[3], Sally Brown[1], **David Stahl**[2], Dylan McCalmont[1]* (dmccalmo@uw.edu), Anthony Bertagnolli[2], Kelley Meinhardt[2], Manmeet Pannu[2]

[1]School of Environmental and Forest Sciences, [2]Civil and Environmental Engineering, University of Washington, Seattle, [3]Department of Crop and Soil Science, Washington State University, Prosser

The impact of an expanded bioenergy industry on soil and atmospheric chemistry remain unknown. Reduced nitrogen (R-Nr) application is required for enhanced plant growth. Nitrifying microorganisms in soil convert R-Nr to more mobile and biologically less favorable oxidized forms, reducing crop yields and increasing production of detrimental fugitive gases. Ammonia oxidizing Archaea (AOA) are often observed in high abundances in soils, and are implicated as the dominant assemblage responsible for the first step of nitrification, ammonia oxidation to nitrite. However, little information regarding their role in nitrification in soils of biofuels cultivars is available. To that end, the goal of this project was to assess whether differing soil types and nitrogen delivery strategies: 1) impacted the abundance, distribution, and diversity of nitrifying assemblages, 2) altered the release of different fugitive gases 3) altered broader components of microbial diversity and metabolic activities associated with the nitrogen cycle.

Ammonia oxidizing archaea (AOA) contribute to a significant portion of ammonia oxidation in soil. These organisms compete with plants for available N, having significant impacts on plant proliferation, as well as production of fugitive gases. AOA community distribution patterns are influenced by multiple factors, of which, biogeography has emerged as an important variable. Developing an understanding of community differences in AOA amid differing land management types may provide tools to understand differences in N use efficiency and other, broader impacts of AOA on soil and atmospheric biogeochemistry.

The goal of this study was to assess whether agriculturally managed soils displayed shifts in AOA community diversity in contrast to non---managed soils located in close proximity. Soil was collected from two sites in eastern Washington where AOA community diversity patterns have previously been examined in soils influenced by long---term management practices. At both sites soil was collected from the surface horizon (0---15 cm) of the adjacent native shrub---steppe (dominated by bunchgrass) and from switchgrass cultivated fields. AOA communities were evaluated by terminal restriction fragment length polymorphism (TRFLP) targeting subunit A of the *Archaeal* ammonia monooxygenase and analyzed using multivariate statistical approaches. At both the slightly alkalkine and slightly acidic agricultural stations, significant differences in AOA community diversity were observed based on the contribution of differing terminal restriction fragments (TRFs) to managed and native soils based on analysis of similarity (ANOSIM, R value greater than 0.6 $p<0.05$). In contrast, native soils displayed higher similarity to one---another, despite significant spatial separation, than either agriculturally influenced site. In native soils located adjacent to a slightly acidic switchgrass cultivated site, TRFs affiliated with members of the genus *Nitrososphaera* were highly detected that were of significantly lower abundance in cultivated sites. In contrast, TRFs attributed to *Nitrosotalea* were dominant in the switchgrass cultivated site, but were substantially lower in the cultivated site. In addition, a higher number of TRFs were observed in the non---managed areas, indicative of a more diverse AOA community. At the slightly alkaline site, similar differences between native and cultivated AOA communities were also observed. However, the most abundant TRFs in the native soils were non---detectable in the cultivated areas, suggesting a complete replacement of native ecotypes. Preliminary TRF identification suggests different phylogentically distinct

members of the genus *Nitrosophaera* are responsible for the observed shifts between native and cultivated soils. Taken together, our results suggest that agricultural land---management significantly alters AOA community diversity patterns. These results will be used to assess whether these soils are also attributed with differing rates of nitrogen usage and production of fugitive gases, parameters that would be useful for modeling the impacts of switchgrass cultivation on nitrogen cycling soil ecosystems.

76. Land management alters N-cycling metabolic potential in Pacific Northwest soils

Anthony Bertagnolli[1], Kelley Meinhardt[1], Manmeet Pannu[1], Sally Brown[2], Stuart Strand[1], Steve Fransen[3], **David A. Stahl**[1] (dastahl@uw.edu)

1Civil and Environmental Engineering, 2School of Environmental and Forest Sciences, University of Washington, Seattle, 3Department of Crop and Soil Science, Washington State University, Prosser

Project Goals: The impact of an expanded bioenergy industry on soil and atmospheric chemistry remain unknown. Reduced nitrogen (R-Nr) application is required for enhanced plant growth. Nitrifying microorganisms in soil convert R-Nr to more mobile and biologically less favorable oxidized forms, reducing crop yields and increasing production of detrimental fugitive gases. Ammonia oxidizing Archaea (AOA) are often observed in high abundances in soils, and are implicated as the dominant assemblage responsible for the first step of nitrification, ammonia oxidation to nitrite. However, little information regarding their role in nitrification in soils of biofuels cultivars is available. To that end, the goal of this project was to assess whether differing soil types and nitrogen delivery strategies: 1) impacted the abundance, distribution, and diversity of nitrifying assemblages, 2) altered the release of different fugitive gases, and 3) altered broader components of microbial diversity and metabolic activities associated with the nitrogen cycle.

Abstract: Cellulosic ethanol production from switchgrass has gained interest as an economically and environmentally sustainable alterative to fossil fuels. Successful plant propagation requires the application of inorganic nutrients, of which nitrogen-based fertilizers are the most expensive, and often converted to energetically and environmentally less favorable forms by nitrifying microorganisms. To assess whether alternative land-management practices associated with biofuels crop production differentially impact nitrogen transformations, metagenomes were generated from several soils in Eastern and Western Washington associated with different soil chemistries and agricultural histories. We focused on genes for nitrification (ammonia monooxygenase), general nitrogen cycling processes (including denitrification, dissimilatory nitrate reduction to ammonia, assimilatory nitrite/nitrate reduction, and urea utilization), and genes implicated in archaeal ammonia oxidation (e.g., the copper containing nitrite reductase). Abundances of ammonia monooxygenase genes were highest in long-term agricultural field plots in Eastern Washington, and mainly affiliated with *Thaumarchaeota*. Comparisons of short metagenomic fragments with full-length amoA gene sequences in clone libraries developed from the same sites revealed good consistency in diversity and rank abundance patterns, with *Nitrososphaera* and *Nitrosotalea* comprising the majority of the AOA ecotypes. Western Washington forest soils displayed lower levels of thaumarchaeal amoA, yet bacterial variants displayed similar abundance patterns in both eastern and western soils. Gene variants involved in canonical denitrification (nirS, nosZ) were elevated levels in some, but not all, agricultural soils. Notably, nosZ variants affiliated with the recently described 'atypical', non-denitrifier pathway were abundant in most soils, in some instances in numbers ten-fold greater than the nosZ of typical denitrifiers. Thus, emissions of N_2O from soils may be attenuated under condition favoring the expression of this novel reductase. The demonstration that AOA are the dominant ammonia-oxidizing population in these intensively managed soil systems now provides a framework for associating key transformation of nitrogen, including the production of atmospherically reactive oxides, with patterns of AOA species distribution and activity.

Research was supported by DOE Genomic Science and Technology for Energy and the Environment grant DE-SC0006869

77. Systems Level Insights into Alternate Methane Cycling Modes Phase II: Deciphering Mechanistic Details of Interspecies Cooperation Through Manipulation of Model Microbial Communities

Ludmila Chistoserdova*, Igor Oshkin, David A.C. Beck, and **Mary E. Lidstrom** University of Washington, Seattle WA

Project Goals: (1) To develop model synthetic communities active in methane utilization through employing axenic cultures of methane-oxidizing bacteria, non- methanotrophic methylotrophs and non-methylotrophic heterotrophs, and to manipulate these communities by modifying methane and oxygen partial pressures followed by observations on growth and performance of the communities and behavior of each of the partners; (2) To identify candidate factors involved in interspecies cooperation through utilizing comparative transcriptomics and metabolomics; (3) To validate hypotheses via strain and community manipulation employing site-specific mutagenesis and/or introduction of synthetic functional modules to metabolically modify strains of interest and by substituting modified strain variants for wild type strains in manipulated microcosms. To Integrate data and use phenotypes and growth characteristics for developing and constraining single strain and community metabolic models.

In the first phase of this project, we investigated populations of microbes in a freshwater lake (Lake Washington) sediment active in methane oxidation at different oxygen tensions, using systems biology approaches such as microcosm manipulations, (meta)genomics, (meta)transcriptomics, (meta)metabolomics as well as pure culture manipulations. The major discoveries from this project include identification of the *Methylococcaceae* family microbes and specifically the *Methylobacter* species as key players in methane oxidation not only in aerobic but also in microaerobic conditions, and a rapid transfer of carbon originating from methane to non-methanotroph species, suggesting that metabolism of methane must involve a food chain rather than a single type of microbe (1). We have also uncovered one mechanism by which *bona fide* methanotrophs may support non-methanotroph communities, a novel type of methane metabolism involving fermentation (2). At the same time we have observed that communities active in methane metabolism are not random combinations of species, instead these are communities involving specific bacterial types, the most prominent being the *Methylotenera* species (3).

In this new phase of the project that commenced on September 1, 2013, we are addressing the nature of these proposed relationships among different physiological groups of microbes and the mechanisms underlying their specificity. In order to obtain mechanistic details into interspecies cooperation as part of the methane cycle, we are now using both top-down and bottom-up community manipulation approaches. In the first, we manipulate natural communities using methane as the only carbon source and monitor complex community deconvolution toward dramatically simplified, several-species communities. In the second, we use pure cultures of bacteria, all originating from Lake Washington sediment, all with sequenced genomes, to build and manipulate simple synthetic communities. At this meeting we will present data from the top-down approach demonstrating that, under methane pressure, the complex natural communities simplify rapidly, with *Methylobacter* species becoming one of the dominant species under two different oxygen regimens (referred to as ‘high’ and ‘low’). The communities, sampled at multiple time points using 16S rRNA gene pyrosequencing, revealed dynamic behavior with respect to species accompanying *Methylobacter*, which, while representing a significant portion of the total population at each given time, progressed through a series of population sweeps. The observed community dynamics differed between the ‘high’ and ‘low’ oxygen condition microcosms in terms of the dominant species composition.

Notably, the *Methylotenera* species that were the first responders to the methane stimulus in both conditions were replaced by the *Methylophilus* species in the 'high' oxygen conditions, followed by successions of various non-methylotrophic heterotroph species.

In contrast, the *Methylotenera* species persisted in the 'low' oxygen conditions. Moreover, in these conditions, we were able to observe dynamics between two different ecotypes of *Methylotenera*. Insights into the physiology of the major species forming these simple methane-utilizing communities were gained through metagenomic sequencing revealing differences in the genomic blueprints of the dominant species, including differences in major pathways for carbon and nitrogen metabolism. These data from the top-down approach suggest the most prominent models to employ in the bottom-up synthetic community experiments.

References

1. Beck, D.A.C., M.G. Kalyuzhnaya, S. Malfatti, S.G. Tringe, T. Glavina del Rio, N. Ivanova, M.E. Lidstrom, and L. Chistoserdova. A metagenomic insight into freshwater methane-utilizing communities and evidence for cooperation between the *Methylococcaceae* and the *Methylophilaceae*. PeerJ 2013, 1:e23.
2. Kalyuzhnaya, M.G., S. Yang, O.N. Rozova, N.E. Smalley, J. Clubb, A. Lamb, G.A. Gowda, D. Raftery, Y. Fu, F. Bringel, S. Vuilleumier, D.A. Beck, Y.A. Trotsenko, V.N. Khmelenina, and M.E. Lidstrom. Highly efficient methane biocatalysis revealed in a methanotrophic bacterium. Nature Communications 2013, 4:2785.
3. Chistoserdova, L., M.G. Kalyuzhnaya, and M.E. Lidstrom. Cycling single-carbon compounds: from omics to novel concepts. Microbe 2013, 8:395-400.

This research is funded by the Grant DE-SC0010556.

78. Soil Microbial Community Composition as an Indicator of Agroecosystem N_2O Emissions

David S. Duncan[1,2*] (dsduncan@wisc.edu), Kelsea A. Jewell[1], David A. Williams[2], Garret Suen[1], and **Randall D. Jackson[1,2]**

[1]University of Wisconsin, Madison; [2]Great Lakes Bioenergy Research Center, Madison, WI

Project Goals: This project is part of a larger effort that aims to use high throughput molecular methods for microbial community characterization to understand the effects of different bioenergy cropping systems on the composition and function of soil microbial communities. The primary focus is on nitrogen cycling and linking microbial community composition to soil nitrous oxide emissions. This is a preliminary study exploring changes to the microbial community as measured by a variety of molecular methods during the establishment phase of a prairie restoration contrasted with continuous corn.

High throughput molecular methods for taxonomic and functional characterization have revolutionized microbial ecology. These changes have motivated several fields, including ecology and medicine, to consider microbial communities as sources of information about complex environments and systems. In the study of soils and soil-based processes, microbial communities are a particularly appealing source of information because of their capacity to integrate a broad range of environmental factors, while filtering these for biological relevance. At the same time, many key processes in soils, particularly those involving nutrient and carbon cycling, are directly mediated by microbes, so measurements of the community may indicate the biological potential of a system to respond to environmental stimuli. Among these processes, the nitrogen cycle, and particularly the emission of nitrous oxide (N_2O), has received considerable attention both because of its relative biochemical simplicity and conservation and because of its importance to global climate change.

Our project takes place within the context of the Great Lakes Bioenergy Research Center's (GLBRC) cropping systems experiment, located at Arlington, WI. Through this experiment, we have established eight managed agroecosystems whose biomass could serve as a feedstock for biofuel or bioenergy production. As part of GLBRC's mandate to evaluate and improve the sustainability of novel biofuels, the cropping systems experiment has included the collection of a variety of ecologically relevant environmental measures, including trace gas fluxes from soils and inorganic nitrogen pool sizes. In our project, we have collected soils from a range of the ecosystems under consideration, including continuous corn, miscanthus, switchgrass, and a mixture of prairie species. Our long term objective is to combine microbial community characterizations from these soils with the rich environmental dataset being generated to provide a robust estimation of the interactions among microbial community properties and environmental drivers in determining N_2O emissions.

Here, we present a preliminary component of that study, where we focused on the two most strongly contrasting cropping systems: continuous corn and restored native prairie. We sampled these systems in years 3-5 of establishment to capture their ongoing divergence from a common background state. We employed a variety of characterization techniques including phospholipid fatty acid (PLFA) analysis, targeted 16S ribosomal subunit and bacterial nitrous oxide reductase (*nosZ*) sequencing, and Illumina whole metagenome shotgun sequencing. Each of these techniques characterizes a different aspect of the microbial community. We analyze the extent to which each technique captures differences between the two cropping systems, as well as among years. In addition, we test the degree to which different sources of microbial community information can explain variation in aggregated annual N_2O fluxes within systems and within years.

79. Integrated Biological Conversion of Gaseous Substrate into Lipids

Peng Hu, Sagar Chakraborty, Hongjuan Liu, Amit Kumar, **Gregory Stephanopoulos**

Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA.

Email: gregstep@mit.edu
http://bamel.scripts.mit.edu/gns/

The development of alternative fuel sources and the implementation of appropriate waste gas mitigation technologies have become high global priorities. Although possibilities for alternative sources are being explored, there has been relatively little discussion on gaseous substrates - such as syngas - as potential feedstocks for bioconversion to liquid fuel production. Sources of syngas include effluents of cement or steel manufacturing or the product of coal, methane or biomass gasification. Building on our previous studies, we introduce a two process system where in one stage syngas is converted to acetate and in the other acetate is converted to lipids. In this study, we choose *Moorella thermoacetica* and *Yarrowia lipolytica* as our model microorganisms. Metabolic engineering and process modeling suggest that the acetic acid produced by *M. thermoacetica* can be fed into the bioreactor containing *Y. lipolytica* for lipid production in an integrated system. In the first stage CO_2 is reduced in the presence of H_2 to acetate (30 g L^{-1}), and afterwards this acetate containing stream is fed into a aerobic bioreactor where an engineered *lipolytica* produces a lipid titer of 46 g L^{-1} with a productivity of 0.27 g L^{-1} h^{-1} and lipid content of 59 %. Thus by combining syngas fermentation with lipid production, we show that waste gas can be effectively converted into lipids, and that this integrated bioprocess has potential to be an economically viable technology for the production of alternative fuels. In addition, this finding provides a versatile approach that has the potential for utilizing other feedstocks such as municipal solid waste and food industry waste for the production of alternative fuels and chemicals.

Acknowledgement: This research was supported by the Department of Energy's Genomic Science Program.

80. Risk and Escape Policies, Perspectives, and Practices: Issues and Implications for Biosystems Design R&D on Microbes, Algae, and Plants

Amy K. Wolfe[1*] (wolfeak@ornl.gov), David J. Bjornstad,[2] Barry L. Shumpert,[2] and Maria Fernanda Campa[3]

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee, [2]Retired, formerly Oak Ridge National Laboratory, Oak Ridge, Tennessee, [3]Bredesen Center for Interdisciplinary Research, University of Tennessee, Knoxville

Project Goals. This project that aims to identify circumstances that create, amplify, or diminish environmental and health risks associated with biosystems design research and development (R&D), thereby revealing new opportunities for avoiding or managing those risks. It takes a systems approach that places biosystems design risk and containment within the R&D context.

This poster describes a new research project that explores risk and containment issues associated with biosystems design (synthetic biology) R&D from social and institutional perspectives. It views the R&D context as part of a complex system, and focuses on key elements that have the potential to contribute to or reduce environmental and health risks associated with biosystems design R&D. Inquiries will explore such key elements of R&D context as social and institutional interactions, research setting, research approach, and organism studied. To identify circumstances and new opportunities for avoiding or managing biosystems design R&D-related risks, questions principally will address:

- sources of variability in R&D context that influence the potential for environmental and health risks;
- risk and escape implications of current risk and containment practices; and
- gaps in current containment practices that inadvertently may increase risks and signal the need for new practices.

Questions will be investigated from three different perspectives: (a) formal (e.g., journal articles) and informal (e.g., news articles, reports, and the like) accounts and documentation; (b) scientists engaged in biosystems design R&D; and (c) other parties key to the achievement of biosafety in R&D settings, such as biosafety committee members, laboratory safety managers, and waste handlers. Research will focus on primarily on biosystems design R&D oriented toward next- generation biofuels and a subset of the U.S.-based research projects and institutions engaged in this work. Secondarily, limited attention will be paid to dual use, European research or risk/containment efforts, and a wider array of scientists—including a portion of the do-it-yourself (DIY) community.

This research is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

81. A Systems-Level Analysis of Drought and Density Response in the Model C4 Grass *Setaria viridis*

Authors: **Thomas P. Brutnell** (tbrutnell@danforthcenter.org)[1]*, Ivan Baxter[1,2], Asaph Cousins[3], José R. Dinneny[4], Andrew D.B. Leakey[5], Todd Mockler[1], Hector Quemada[1], Seung (Sue) Rhee[4], Dan Voytas[6], [1]Donald Danforth Plant Science Center, St. Louis, MO; [2]USDA-ARS; [3]Washington State University, Pullman, WA; [4]Carnegie Institution for Science, Stanford, CA, [5]University of Illinois, Urbana-Champaign, IL; [6]University of Minnesota, St. Paul, MN

http://foxmillet.org/

Project Goals: To develop *S. viridis* as a model system to molecularly dissect traits that limit current bioenergy feedstock grasses. This includes developing an informatics toolkit to enable the interrogation of the genome, mapping genes that underlie water use efficiency and density tolerance, developing new gene knockout and replacement technologies and investigating gene flow for biocontainment. A web portal will host transcriptomic and QTL datasets generated over the course of this project as well as metabolic and gene annotation data for the *S. viridis* genome. The tools and datasets will not only serve this project but also provide a resource for the *Setaria* research community to mine and display community datasets.

Abstract text: Bioenergy grasses promise to provide a sustainable source of renewable fuels for the US bioenergy economy. These dedicated second generation bioenergy crops can be grown on marginal lands and with fewer inputs than traditional row crops such as corn which requires energy intensive annual planting and the addition of chemical fertilizers. To engineer bioenergy grasses with the desirable traits needed for large scale production, it will be necessary to develop model systems that are closely related to bioenergy feedstocks, but which are more amenable to genetic analysis. One of the most promising model species is *Setaria viridis*. *S. viridis*, like all major feedstock targets, is a C_4 panicoid grass. However, it is much smaller in stature, flowers within 6 weeks of planting, and can be easily transformed with genes of interest. The objectives of this project are to utilize genomic, computational and engineering tools to begin the genetic dissection of drought and density response in *S. viridis*. The ability of bioenergy feedstocks to utilize water efficiently and to produce abundant yields at high density will be major drivers in the development of improved varieties. Through the use of *S. viridis*, we hope to discover the molecular mechanisms that underlie this response and in doing so identify candidate genes and pathways for improving the closely related feedstock grasses. Importantly, we will also begin field studies to assess the ethical, legal and societal implications (ELSI) in anticipation of broad adoption of these technologies.

The specific aims of the proposal are to: 1) Identify QTL for the effect of drought and density on biomass and seed yield components of *S. viridis*. 2) Conduct in–depth physiological profiles in roots and leaves of a subset of selected lines 3) Integrate datasets and develop metabolic and gene networks for *S. viridis* to prioritize candidate genes for transgenic characterizations 4) Develop transformation technologies for *S. viridis* 5) Functionally examine the role of candidate genes deduced by network models; and 6) Develop protocols and best practices for monitoring gene flow in transgenic *S. viridis*. To achieve these aims we will produce one of the most extensive molecular characterizations of plant growth in the field to date, generating several million data points that will be collected from physiological and molecular genetic studies. We will develop novel informatics models and network tools that will guide future molecular characterization in *S. viridis* and guide breeding efforts in major feedstock targets. The ELSI activity will lay the foundation for future studies and establish *S. viridis* as the model system for bioenergy grass research. Work over the past year of this project has included: 1) Conducted replicated field trials of *S. viridis* x. *S. italica* RIL populations to examine the response of plant growth to water stress and density 2) Developed

a pipeline and performed QTL analysis of several traits examined in field studies 3) Established SetariaCyc, a genome-wide metabolic network of *Setaria italica* (available from the Plant Metabolic Network website) 4) Generated an *S. viridis* genome assembly and established a framework to visualize SNP, QTL and mutant datasets 5) Established a *Setaria viridis* focused web-portal that provides access to the *S. viridis* genome assembly and annotation including a genome browser (JBrowse), and BLAST, FTP functionalities 6) Constructed and transformed several constructs to perform site directed mutagenesis in *S. viridis* 7) Developed resources and protocols for conducting genetic analysis in *S. viridis* in field and laboratory settings 8) Conducted developmental, physiological and QTL studies of water stress responses in the root system.

82. Integrating energy transduction from light absorption to biofuel precursors in the alga *Phaeodactylum tricornutum*.

Graham Peers[1*] (graham.peers@colostate.edu), Nathan Sindt[1], Michael Caballero[1], Seijin Park[1], Corey Broeckling[1], **Bernhard Ø. Palsson[2]**, **Chris L. Dupont[3]**, **Andrew E. Allen[3]**

[1]Colorado State University, Fort Collins; [2]University of California, San Diego; [3]J. Craig Venter Institute, San Diego, California

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

We are using a systems biology approach to understand and increase the efficiency of photon and carbon utilization in the model marine diatom *Phaeodactylum tricornutum*. Our approach spans photo---physiology observations to metabolic modeling and pathway manipulation to identify important pinch---points of metabolism and increase the overall yield of fuel precursors. The work reported in this abstract is one avenue of this project.

We observed changes associated with cell metabolism during a shift from excess light energy to light fluxes that limit growth. This represents a switch in cellular energy states from one where energy is being shunted to storage products such as carbohydrate (chrysolaminarin) or lipid (triacylglycerol) to a condition where these storage products are being consumed to maintain cell division. This switch also captures the process of photoacclimation, whereby light-limited cells begin to increase cellular pigments concentrations – this process is a well recognized target for improving overall photosynthetic efficiency in dense cultures of algae like those found in an industrial scale algal photobioreactor.

We have identified a series of differentially regulated, novel targets for increasing yields of lipids. These include 4 putative components of chrysolaminarin biosynthesis and a putative acyl-CoA-dehydrogenase that may participate in the oxidation of fatty acid. We are using a novel knock-down vector to investigate the role of these proteins in carbon partitioning. We have also identified several proteins that play likely play role in photoacclimation including three light harvesting antenna proteins (LHCF5, -16 and -14) that are highly upregulated in response to low light.

We are working to increase data analyses throughput for the rapid, automated analyses of shotgun metabolomics data. We are developing software code that will query spectral databases to ID experimentally derived shotgun metabolite data sets. This will greatly increase the speed that metabolites can be identified. We are also developing tools to detect and quantify highly polar metabolites associated with the primary steps of photosynthesis (e.g. NADPH, fructose-6-phosphate).

The physiology of photosynthesis is dynamic over the course of a natural day/night cycle. *Phaeodactylum* induces a pathway to maintain an oxidized plastoquinone pool during the highest light fluxes of the afternoon and this is independent of non-photochemical quenching. We are currently investigating the role

of alternate electron transport pathways in this photoprotective response and are quantifying the energy flow through this pathway using membrane-inlet mass spectrometry techniques.

This research is supported by the Office of Biological and Environmental Research in the DOE Office of Science grant # DE-SC0008595. MC is supported by a NSF graduate student fellowship.

83. Identifying biological pathway targets for lipid production in *Yarrowia lipolytica*

Scott E. Baker [1]* (scott.baker@pnnl.gov), Siwei Wei[2], Young-Mo Kim[2], Kyle Pomraning[1], Bruce Arey[1], Alice Dohnalkova[1], Sue Karagiosis[3], Galya Orr[1], Tom Metz[2], **Gregory Stephanopoulos[4]**

[1]Environmental Molecular Sciences Laboratory, [2]Biological Sciences Division and [3]Energy Processes and Materials Division, Pacific Northwest National Laboratory, Richland, WA; [4]Massachusetts Institute of Technology, Cambridge, MA

Project Goals: Develop tools for cell-wide measurement of metabolites and lipids, which, along with genetic, light and electron microscopic and transcriptional analysis, will allow the construction of genome-scale metabolic models, as well as models of transcriptional regulation, that will guide the further metabolic engineering of *Yarrowia lipolytica*.

Yarrowia lipolytica is an oleaginous ascomycete yeast that is well known for its ability to accumulate large amounts of lipids. Despite a growing scientific literature focused on lipid production by *Yarrowia*, there remain significant knowledge gaps regarding the key biological processes involved. We are taking a multidisciplinary approach to identifying and characterizing the key pathways involved in *Yarrowia* lipid accumulation. A combination of mass spectrometry based metabolomics, light and electron microscopy and forward and reverse genetics is being applied to deepen our understanding of how this organism regulates and controls lipid production. We collected samples every 12 hours during a 5 day time course of *Yarrowia* growth under lipid producing conditions.

From these samples, we performed metabolomic analysis, super resolution fluorescence microscopy, helium ion microscopy and transmission electron microscopy. Our results indicate that cell wall biosynthesis may pull carbon flux away from lipid production making this an attractive target for future strain improvement efforts. In addition, we have developed a Nile Red-based genetic screen for Yarrowia mutant strains with altered lipid accumulation. We are currently characterizing these strains.

This Biosystems Design research is funded by the U.S. Department of Energy, Office of Biological and Environmental Research under the Genomic Science Program.

84. Measurement and manipulation of phenylpropanoid metabolic flux in Arabidopsis

Peng Wang[1]* (wang1155@purdue.edu), Rohit Jaini[2], Antje Klempien[1], John Morgan[2], Natalia Dudareva[1], **Clint Chapple**[1]

[1] Department of Biochemistry, Purdue University, West Lafayette, IN; [2] School of Chemical Engineering, 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 heterogeneous phenolic polymer that constitutes about 30% of the carbon fixed by photosynthesis in terrestrial plants. Cross-linked with polysaccharides in the plant secondary cell wall, lignin provides strength and hydrophobicity to plant cell walls, but makes it problematic to utilize lignocellulosic biomass for forage, paper making, and biofuel production. Lignin is derived from the phenylpropanoid pathway, the architecture of which is well understood based upon the biochemical and genetic investigations into the enzymes and the genes encoding them conducted to date. In contrast, we lack a systematic and quantitative view of the factors that determine carbon flux into and within this branched metabolic pathway in plants. Several enzymatic steps in the lignin biosynthetic pathway have been hypothesized to be critical integrators of phenylalanine and lignin biosynthetic flux. To explore the control of carbon allocation in the lignin biosynthetic pathway, we are developing a kinetic model of the pathway in Arabidopsis and performing metabolic control analysis to test the regulatory role of several key steps. We have established two experimental systems for flux analysis using wild-type Arabidopsis: developing seedlings and excised mature stems. Our data from the seedling system show that the *de novo* synthesis of phenylpropanoids and lignin commences early during Arabidopsis development. We have also found that excised stems continue to grow and lignify for in excess of 48 hours when incubated in appropriate medium, and show a distribution of PAL and 4CL activities consistent with the pattern of lignin deposition observed. When $^{13}C_6$-ring labeled phenylalanine was supplied to excised stems, corresponding isotopologues of a number of intermediates have been detected and quantified by LC/MS/MS, and incorporation of $^{13}C_6$-ring labeled monolignols into lignin was demonstrated by DFRC/GC/MS. We are developing LC/MS/MS methods for detection of the remaining intermediates in the phenylpropanoid pathway. We will analyze the isotope abundance and metabolite concentrations in the pathway in these two established systems, and use these data for kinetic modeling and metabolic control analysis. Following the generation of kinetic model in wild-type Arabidopsis, we will take advantage of monolignol biosynthetic mutants in hand to evaluate and refine the model.

This research is supported by the Office of Biological and Environmental Research in the US Department of Energy.

85. Enhancing H_2-Production and Nutrient Exchange in a Symbiotic Bacterial Coculture

Alexandra L. McCully[1], Gina C. Gordon[1], Breah LaSarre[1], Timothy A. Kremer[1], David Liu[1], Amanda L. Posto and **James B. McKinlay[1]*** (jmckinla@indiana.edu)

[1]Indiana University, Bloomington

Project goals: Specialized features of individual microbial species can be combined in a complementary manner in cocultures or consortia to produce useful fuels and chemicals. The challenge of maintaining stable microbial relationships has impeded progress in characterizing and implementing such consortia. The goals of our project are to develop a stable hydrogen gas-producing microbial coculture and to use genetic, biochemical, evolutionary, and systems biology approaches to characterize and manipulate microbial interactions and enhance hydrogen production.

Hydrogen gas is an important commodity chemical and is being considered as a future biofuel. Many fermentative bacteria, like *Escherichia coli*, produce H_2 from carbohydrates under anaerobic conditions but at low yields due to the obligate excretion of organic acids and alcohols. Photosynthetic purple nonsulfur bacteria, like *Rhodopseudomonas palustris*, consume fermentation products under anaerobic conditions and use some of the electrons to produce H_2 via nitrogenase. It has long been realized that combining these two classes of microbes can result in higher H_2 yields from carbohydrates (1). However, little progress has been made with such cocultures in the last thirty years likely due to the non-trivial challenge of maintaining stable relationships between the two species. Using defined mutations and environmental conditions we recently developed a stable coculture of *E. coli* and *R. palustris*. In this coculture *E. coli* ferments carbohydrates and excretes carbon for *R. palustris* and *R. palustris* fixes N_2 gas and excretes nitrogen for *E. coli*. One species cannot survive without the other.

We are currently examining the environmental, metabolic, and evolutionary factors that influence coculture H_2 production and nutrient exchange. An environmental perturbation that profoundly affects the H_2 yield is whether the cocultures are shaken or not. When cocultures are not shaken, N_2 gas diffusion is limited and results in a subpopulation of *R. palustris* that is starved for nitrogen. When starved, *R. palustris* still metabolizes carbon and redirect electrons away from biosynthesis to H_2 production (2). The H_2 yield from non-shaken cocultures was 17- times higher than in shaken cocultures. Despite the starving subpopulation, non-shaken cocultures were still viable with reproducible trends through serial transfers.

A metabolic factor that affects coculture H_2 production and nutrient exchange is the CO_2-fixing Calvin cycle. The Calvin is known to compete for electrons against H_2-production via nitrogenase in *R. palustris* monocultures and genetically disrupting Calvin cycle flux leads to higher monoculture H_2 yields (3). We recently determined that deleting the genes encoding the Calvin cycle enzyme, Rubisco, results in a lower *R. palustris* monoculture growth rate than when the gene encoding the Calvin cycle enzyme, phosphoribulokinase, is deleted (4). The low Rubisco mutant growth rate is likely the result of toxic metabolite accumulation that is avoided in phosphoribulokinase mutants (5). Furthermore, we found that phosphoribulokinase mutants have a higher specific rate of H_2 production than Rubisco mutants in monoculture. In coculture, we similarly found that genetically disrupting *R. palustris* Rubisco activity led to a lower coculture growth rate. However, disrupting *R. palustris* phosphoribulokinase activity led to a higher coculture growth rate, likely due to an increased redirection of electrons to nitrogenase and higher rate of nitrogen excretion. Cocultures with either Calvin cycle mutant produced more H_2 than cocultures containing the *R. palustris* parent strain.

We will continue to use our ability to genetically manipulate each species as we determine the effect of *E. coli* fermentation product profiles on coculture traits and ultimately employ transcriptomic and fluxomic approaches to characterize interspecies interactions in this simple defined community.

References:

1. Odom JM, Wall JD. 1983. Photoproduction of H_2 from cellulose by an anaerobic bacterial coculture. Appl. Environ. Microbiol. 45:1300–1305.

2. McKinlay JB, Oda Y, Ruhl M, Sauer U, Harwood CS. 2014. Non-growing *Rhodopseudomonas palustris* increases the hydrogen gas yield from acetate by shifting from the glyoxylate shunt to the tricarboxylic acid cycle. J. Biol. Chem. 289:1960–1970.

3. McKinlay JB, Harwood CS. 2011. Calvin cycle flux, pathway constraints, and substrate oxidation state together determine the H_2 biofuel yield in photoheterotrophic bacteria. MBio. 2:10.1128/mBio.00323–10.

4. Gordon GC, McKinlay JB. 2014. Calvin cycle mutants of photoheterotrophic purple non-sulfur bacteria fail to grow due to an electron imbalance rather than toxic metabolite accumulation. J. Bacteriol. In press.

5. Wang, D., Zhang, Y., Pohlmann, E. L., Li, J., & Roberts, G. P. (2011). The poor growth of Rhodospirillum *rubrum* mutants lacking RubisCO is due to the accumulation of ribulose-1,5-bisphosphate. J. Bacteriol. 193: 3293–3303.

This work was supported by the Office of Science (BER), U.S. Department of Energy grant DE-SC0008131.

86. Regulation of Cellular Nitrogen Metabolism in the Model Marine Diatom *Phaeodactylum tricornutum*

Andrew. E. Allen (PI)[1*], Christopher L. Dupont (Co-PI)[1], McCarthy, James[1], Adriano Nunes-Nesi[2], Sabrina, Kleeson[2], Alisdair Fernie[2]

[1.] Microbial and Environmental Genomics, J. Craig Venter Institute, San Diego, CA, USA

[2] Integrative Oceanography Division, Scripps Institution of Oceanography, University of California, San Diego, CA 92093, United States

[3]Max-Planck-Insitut für Molekulare Pfanzenphysiologie, 14476 Potsdam-Golm, Germany

Project Goals:

While the complete genome sequence of a centric and pennate diatom, forward and reverse genetic techniques and *in silico* modeling have enabled our laboratory and others to begin characterizing unknown genes, pathways and interactions; nevertheless, key information sets necessary to a systems biology approach to diatom biology remain undeveloped. Our proposed goals focus on two critical gaps in the diatom knowledge base: *i)* although *in silico* models of carbon and nitrogen metabolism depend on information about subcellular locations of metabolic pathway constituents, very few have yet to be experimentally verified; *ii)* overall pathways and mechanisms controlling cellular carbon and nitrogen sensing, assimilation, and flux, in diatoms remain largely undescribed and have not been formally linked to lipid metabolism. Using a combination of transcriptomics, proteomics, phosphoproteome, metabolomic and stable isotope metabolite flux profiling in steady state cultures, along with directed enzyme localization and biochemistry experiments, we will evaluate lipid metabolism within the overall context of cellular nitrogen and carbon metabolism.

Abstract:

The unique evolutionary footprint of diatoms may have fostered the evolution of peculiar and unique biochemical pathways contributing to the ecological success of diatoms in the modern ocean. Most notably, a complete metazoan-like urea cycle appears to have been acquired from the host of the secondary endosymbiotic event that gave rise to the Chl c algae. In metazoans, the urea cycle is involved in the catabolism of amino acids and the generation of urea for export. The presence of the urea degrading enzyme urease, acquired from the endosymbiont, strongly suggests an alternative function in diatoms. In marine diatoms, which are frequently subjected to nitrogen limitation, we hypothesize that the urea cycle functions in an anabolic capacity to repackage and recycle inorganic C and N from both endogenous and exogenous sources (Allen et al., 2011). Like green algae and vascular plants, diatom genomes also appear to encode plastid targeted Glutamine Synthetase-Glutamine oxoglutarate aminotransferase (GS-GOGAT) components; unlike green lineage eukaryotes, however, diatoms also express distinct mitochondrially targeted GS-GOGAT genes. This mitochondrial GS-GOGAT cycle, in tandem with a mitochondrial urease, might allow for a rapid redistribution of urea cycle-derived nitrogen metabolites to amino acids following the cessation of nutrient limitation. We propose that a two-part uptake system, involving a plant-like outer membrane transporter and a metazoan-like mitochondrial transporter, delivers urea from the extracellular mileau to the mitochondria. Genomic analyses and metabolite flux studies show that the ammonium produced by urease is assimilated using a complete GS- GOGAT cycle found in the mitochondria, with ancillary fixation through CPS III and the urea cycle. In contrast, nitrate-derived ammonium is clearly assimilated through a plastid-localized GS-GOGAT cycle, with a transfer to the urea cycle metabolite pool via arginosuccinate synthase. Comparative genomic analyses suggests this bifurcated nitrogen assimilation system may be present in other phytoplankton of the chromaveolate lineage. RNAi and TALEN mediated knockdown of mitochondrial urease and mitochondrial and

chloroplast localized GS levels are providing additional insights into overall cellular regulation of nitrogen metabolism.

Nitrate reductase (NR) is also enzyme central to overall cellular nitrogen assimilation and metabolism. NR was predominantly believed to be involved in reduction of nitrate as part of nitrogen assimilation. However, mounting evidence suggests a multifunctional role in marine diatoms. First, NR is highly upregulated under cold temperature-high light conditions; this been hypothesized to suggest that NR provides an alternative electron sink for photosynthetically derived electons and reductants that are in excess due to an imbalance between carbon assimilation and growth (Lomas and Gilbert 1999; Parker and Armbrust 2005).. Additionally, in NR-YFP transgenic overexpressors, nitric oxide production is greatly increased. This signaling molecule has been implicated in apotosis and cell-cell signaling in diatoms; although the source of NO in plant cells remains controversial, the peroxisome and NR have each independently, but never together, been implicated in NO production. It is tempting to speculate that NR could be fueling NO production in diatom peroxisomes. In any case, it appears clear that NR is at the center of nitrogen assimilation, signaling, and energy balance. In order to investigate this in more detail we have performed a series of immunolocalization experiments intended to examine NR localization *in vivo* in response to cellular nitrogen status and nitrogen source. NR localization appears to oscillate between the cytosol, peroxisome, and association with the vacuole. RNAseq, proteomic, phosphoproteomic, and metabolomic experiments aimed at preliminary characterization of the diatom response to cellular nitrogen status and nitrogen source have also been performed. In conjunction with the various data types collected to date, an initial genome-scale model of nitrogen metabolism has been constructed.

Allen, A.E., Dupont,C.L., Oborník, M., Horák, A., Nunes-Nesi, A., McCrow, J.P., Zheng, H., Johnson, D.A., Hu, H., Fernie, A.F., Bowler, C. 2011. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473(7346):203-7

Lomas, M. W., and P. M. Gilbert. 1999. Temperature regulation of nitrate uptake: A novel hypothesis about nitrate uptake and reduction in cool-water diatoms. Limnology and Oceanography 44:556-572.

Parker, M. S., and E. V. Armbrust. 2005. Synergistic effects of light, temperature, and nitrogen source on transcription of genes for carbon and nitrogen metabolism in the centric diatom Thalassiosira pseudonana (Bacillariophyceae). Journal of Phycology 41:1142-1153.

Lomas, M. W., and P. M. Gilbert. 1999. Temperature regulation of nitrate uptake: A novel hypothesis about nitrate uptake and reduction in cool-water diatoms. Limnology and Oceanography 44:556-572.

DOE-DE-SC0006719, Integrated Systems Biology of Metabolic and Transcriptional Networks in a Model Photosynthetic Microbe

DOE-DE-SC0008593, Optimization of Energy Flow Through Synthetic Metabolic Modules

87. Inactivation of *Phaeodactylum tricornutum* urease gene using TALEN-based targeted mutagenesis

Philip D. Weyman [1], Karen Beeri [2], Stephane C. Lefebvre [2], Josefa Rivera [2], Adam L. Heuberger[3], Graham Peers[4], Christopher L. Dupont(Co-PI) [2], **Andrew E. Allen (PI)**[2,5*]

[1]J. Craig Venter Institute, Department of Synthetic Biology and Bioenergy, 4120 Torrey Pines Rd., La Jolla, California, 92037 USA, [2]J. Craig Venter Institute, Department of Microbial and Environmental Genomics, 4120 Torrey Pines Rd., La Jolla, California, 92037 USA, [3]Colorado State University, Proteomics and Metabolomics Facility, Fort Collins, Colorado, 80523 USA, [4]Colorado State University, Department of Biology, Fort Collins, Colorado, 80523 USA, [5]Integrative Oceanography Division, Scripps Institution of Oceanography, University of California, San Diego, CA 92093, United States

Project Goals: ***Targeted Genetic Manipulation and Systems Biology*:** Our initial genome scale modeling and targeted genetic manipulation activities will focus on cellular modules fundamental for cellular regulation of primary metabolism and overall photosynthetic efficiency. These include pyrenoid formation and function in regulation of the diatom carbon concentrating mechanisms (CCM), a quantitative understanding of linear and alternative electron flow through photosystems (Nogales et al. 2012), a functional characterization of metabolites that are exchanged between the chloroplast and mitochondria and their carrier proteins and manipulation of metabolite partitioning to storage lipids. We will employ an iterative approach of computational modeling followed by genetic manipulation and physiological characterization to characterize these systems.

Abstract:
We describe a method for genome editing in the model diatom *Phaeodactylum tricornutum* using transcription activator-like effector nucleases (TALENs). For a proof of concept, the gene encoding for the enzyme urease was targeted for interruption. A mix of both heterozygous and homozygous mutants was obtained using TALEN-assisted homologous recombination. Southern blot and PCR analyses of genomic DNA confirmed efficient and specific insertion of the knockout cassettes within the urease gene. Western blot analyses confirmed lack of urease in homozygous lines, which were unable to grow on urea as sole nitrogen source. The homozygous urease knockout lines are able to grow on other nitrogen sources and metabolomics analysis revealed a buildup of metabolites indicative of disruption of the urea cycle such as urea, arginine, and ornithine. Numerous high carbon metabolites were enriched in the homozygous mutant, suggesting a disruption of cellular C and N repartitioning. The presented method improves the molecular toolkit for diatoms and clarifies the role of urease in the urea cycle.

DOE-DE-SC0008593, Optimization of Energy Flow Through Synthetic Metabolic Modules

88. Assembly of entire Phaeodactylum tricornutum chromosomes in yeast

Philip D. Weyman[1], Bogumil J. Karas [1], Bhuvan Molparia[1], Jelena Jablanovic[1], Wolfgang J. Hermann[1], Ying-Chi Lin[1], Christian Tagwerker[1], Isaac T. Yonemoto[1], Vladimir N. Noskov[3], Ray-Yuan Chuang[3], John I. Glass[3], Clyde A. Hutchison, III[1], Hamilton O. Smith[1], J. Craig Venter[1,2,3], Christopher L. Dupont (Co-PI) [2], **Andrew E. Allen (PI)** [2,4*]

[1]J. Craig Venter Institute, Department of Synthetic Biology and Bioenergy, 10355 Science Center Dr., San Diego, CA 92121, USA, [2]J. Craig Venter Institute, Department of Microbial and Environmental Genomics, 10355 Science Center Dr., San Diego, CA 92121, USA, [3]J. Craig Venter Institute, Department of Synthetic Biology and Bioenergy, 9704 Medical Center Dr., Rockville, MD 20850, USA, [4]Integrative Oceanography Division, Scripps Institution of Oceanography, University of California, San Diego, CA 92093, United States

Project Goals: ***Synthetic Module Generation and Transplantation***: Using yeast as a platform (Gibson et al. 2008), we will create synthetic, extra-chromosomal elements to install new biological functions into diatoms. These functions may include traits found in related diatom species, or novel traits encoded by a heterologous pathway. As part of this phase, existing diatom promoter sequences will be characterized and evolved to create a suite of endogenous control elements needed in metabolic redesign.

Abstract:
Synthetic genomic approaches offer unique opportunities to use powerful yeast and Escherichia coli genetic systems to assemble and modify chromosome-sized molecules before returning the modified DNA to the target host. The model diatom Phaeodactylum tricornutum has an average G+C content of 48% and a 27.4 Mb genome sequence that has been assembled into chromosome-sized scaffolds making it an ideal test case for assembly and maintenance of eukaryotic chromosomes in yeast. We present a modified chromosome assembly technique in which eukaryotic chromosomes as large as ~500 kb can be assembled from cloned ~100 kb fragments. We used this technique to clone fragments spanning P. tricornutum chromosomes 25 and 26 and to assemble these fragments into single, chromosome-sized molecules. We found that addition of yeast replication origins improved the cloning, assembly, and maintenance of the large chromosomes with moderately high G+C content in yeast. Furthermore, purification of the fragments to be assembled by electroelution greatly increased assembly efficiency. These techniques offer new opportunities to design large biosynthesis pathways for expression in algae and open the door to the development of large DNA fragment replacement in algae for efficient genome editing.

DOE-DE-SC0008593, Optimization of Energy Flow Through Synthetic Metabolic Modules

89. Towards a Comprehensive Knowledge Base for the Marine Diatom *Phaeodactylum tricornutum*

Jennifer Levering[1*](jlevering@ucsd.edu), Alessandra A. Gallina[1], Philip Miller[1], Adam Feist[1], Karsten Zengler[1], Graham Peers[2] (Co-PI), Bernhard Ø Palsson[1] (Co-PI), Christopher L. Dupont[2] (Co-PI), **Andrew E Allen (PI)**[3,4]

[1]Bioengineering Department, University of California San Diego, La Jolla, CA, USA and [2]Colorado State University, Fort Collins, CO, [3]J. Craig Venter Institute, La Jolla, CA, USA, [4]Integrative Oceanography Division, Scripps Institution of Oceanography, University of California, San Diego, CA 92093, United States

Project Goals: Genome-scale metabolic models are fundamental for the analysis of cellular processes at a system level and represent an ideal organizational framework for analyses of functional genomics experimental data and computational studies. In recent years, there has been an increasing interest in high-quality metabolic reconstructions of phototrophic organisms and robust computational tools to integrate 'omic' data from these organisms within genome-scale models. The approach of the project is to combine cutting-edge genome manipulation and physiological characterization with metabolic modeling. The ultimate goal is the exploration of next generation biofuels through a more comprehensive understanding of light-driven lipid metabolism in a model marine diatom.

Bottom-up reconstructions are biochemically, genetically and genomically structured knowledge-bases that contain information such as reaction stoichiometry, reaction reversibility, and the association between genes, proteins and reactions. The first step in the genome-scale metabolic network reconstruction process involves the generation of a draft reconstruction based on the organisms genome annotation and manually curated reference models. To obtain an automated reconstruction the RAVEN Toolbox was used. The draft reconstruction accounts for 589 genes associated with 835 reactions and 1027 metabolites distributed across 5 compartments, namely cytosol, chloroplast, endoplasmic reticulum, mitochondria and peroxisome. In the second step the draft reconstruction is manually refined. The functions in the automated draft reconstruction will be evaluated against organism-specific literature and data. Finally, using the COBRA Toolbox the manually curated reconstruction is converted into a mathematical model. This model will be evaluated and tested against well-known metabolic capabilities of *Phaeodactylum tricornutum* (Pt) such as growth rate, by-products and secretion.

The genome-scale model (GEM) requires an organism-specific biomass objective function which accounts for both the composition of the cell and the energetic requirements necessary to generate biomass. An initial effort has been made in order to gather Pt-specific biochemical data from the available literature. Some of the main differences with the used references models include lipid content (e.g. better representation of C14 and C16 acyl chains), photosynthetic pigments and carbohydrate storage (i.e. chrysolaminarin). The main conclusion has been the need to generate additional data to account for variability of cellular biochemical composition in different growth conditions. Currently, experimental determination of the biochemical composition is being carried out. Given the importance of a comprehensive organization of the available data and information to be used during the model curation, an effort is being made to generate a Pt- specific bibliome database. This type of database represents a useful concept and tool likely to become increasingly used. The up-to date available literature has been collected and currently 1212 publications are being

manually screened and categorized. A web-based interface has been implemented in order to provide a user-friendly tool for the scientific community.

A Postgres relational database was developed to provide storage and retrieval capabilities for the GEM. The relational database stores the components of the model each of which can be queried for comparative analyses. Management of the model versions and development history are also made possible through the relational database. The database supplies the model as an exported SBML or through an SQL interface for direct retrieval and data management. A web interface will provide flexible public access and a graphical report generation facility.

Besides reconstructing the metabolic network, a regulatory network is being built based on RNA sequencing data under multiple conditions. The first step in the network inference is the reduction of dimensionality by grouping genes that are co-regulated into clusters using cMonkey. The resulting clustering is used to obtain a global regulatory network for Pt suggesting regulatory interactions. This network is used to explore Pt's global expression under novel perturbations.

Supported by DOE-DE-SC0008593, Optimization of Energy Flow Through Synthetic Metabolic Modules

90. Developing Synthetic Biology Tools to Improve Nutrient Acquisition of Energy Crops

Patrick Shih[1] and **Dominique Loqué**[1]*(dloque@lbl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley, CA

http://www.jbei.org/research/divisions/feedstock/cell-wall-engineering/

Project Goals: The main goals of this research are the generation of "universal" expression tools for plant root engineering and to utilize them to improve crop nutrient acquisition.

Plant growth is highly dependent on it root system since it anchors plants to the ground; it is responsible to acquire all essential mineral nutrients and water; and it has an important role in plant-rhizosphere interactions. The first green revolution has been mainly driven by the use of fertilizers since deficiency of any essential mineral elements (e.g. N, P, K) negatively impacts photosynthesis and plant growth. Unfortunately the excessive use of N-fertilizers has come at high environmental and economic costs; and its production utilizes a substantial proportion of worldwide energy consumption. The current development of dedicated crops for energy production that can be grown on marginal lands will reduce competition with food products and the pressure on high-quality arable lands utilization. Even if energy crops are selected for their ability to grow on marginal lands, it is important to note that these lands are not suitable for food crop production because the poor quality of the soil (e.g., poor nutrient availability, low water content, vulnerability to erosion, and, heavy-metal pollution) abates their yield potential. Therefore, even if bioenergy crops are better adapted than food crops to grow on marginal lands, they will most likely not reach their full yield potential without fertilizer application unless they are further optimized for growth with low nutrient availability.

Our current focus is to design "universal" expression tools that are functional across widely diverse plant species, and to utilize them to engineer metabolic pathways, which will be designed to optimize nutrient acquisition by energy crops. We expect that the results of this research will generate a diversity of building blocks for plant engineering and will directly contribute to advance the DOE's mission for the sustainable production of bioenergy.

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

91. Adaptive Radiation after Gene Transfer Leads to Population Specialization and Enhanced Glycan Cycling

Jan-Hendrik Hehemann[1]* (hehemann@mit.edu), Manoshi Sen Datta[1], Nathan Cermak[1], Andreas Hentschel, Christopher Corzett[1], Xiaoqian Yu[1], **Martin Polz**[1], and **Eric Alm**[1]

[1] Massachusetts Institute of Technology, Cambridge, MA

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.

Polysaccharides produced by plants and algae represent the major source of food for life on Earth. The rate-limiting factor for their turnover is the activity of microbial carbohydrate active enzymes, or CAZymes, which split polysaccharides into smaller, consumable sugars. How microbes interact to drive glycan cycling in systems such as the human gut or the ocean is poorly understood. Genomic comparisons of 76 *Vibrio* genomes revealed a common degradation pathway for alginate, an important structural polysaccharide in brown algae. This pathway was acquired once via horizontal gene transfer and subsequently modified by gene loss and duplication in different lineages, leading to a cascade of interacting populations in the degradation of alginate. Some populations specialize in the breakdown of the polymer, while others have lost the dedicated enzymes for polysaccharide degradation (alginate lyases), but have retained the import proteins and enzymes (oligo-alginate lyases) for oligosaccharide degradation. Hence, some populations pioneer polymer breakdown, while others harvest the alginate oligosaccharides that diffuse away from the primary degraders, revealing an efficient mechanism of resource exploitation at the community level. Overall, this work illustrates that molecular understanding of CAZymes can be used to reconcile microbial food webs, thereby advancing our knowledge of how polysaccharides are so rapidly metabolized by natural microbial communities.

This work is supported by the Office of Biological and Environmental Research in the Department of Energy Office of Science (DE-SC0008743). Jan-Hendrik Hehemann is supported by Human Frontiers Fellowship.

92. The Ecology of Algal Polysaccharide Degradation: Characterizing Novel Fucoidan-Degrading Bacteria

Christopher Corzett[1]* (corzett@mit.edu), Jan-Hendrik Hehemann,[1] **Martin Polz,**[1] and **Eric Alm**[1]

[1] Massachusetts Institute of Technology, Cambridge, MA

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 (~90% dry mass). Identifying the microbes and metabolic pathways responsible for degrading these sugars is not only crucial to understanding marine carbon flow, but also offers vast potential for biofuel production utilizing seaweed feedstocks. Fucoidans are an important class of structurally heterogeneous sulfated polysaccharides found in brown seaweeds, yet few organisms have been shown to metabolize this abundant carbohydrate. Using environmental samples from coastal waters we have isolated and characterized numerous representatives from diverse genera (*Vibrio*, *Lentimonas*, *Stappia*, *Neptunomonas*, *Alteromonas*, *Tenacibaculum*) capable of using fucoidan as a sole carbon source, and demonstrated enzymatic degradation of fucoidan polysaccharides using cellular extracts.

Preliminary findings also reveal differences in the dynamics and extent of fucoidan degradation among closely related isolates. These differences in enzymatic capabilities could reflect metabolic specialization and indicate resource partitioning occurs during the degradation of specific algal polysaccharides. Ongoing comparative genomic and transcriptional analyses aim to identify and characterize the enzymatic machinery and metabolic modules required for 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).

93. Cloning and Initial Characterization of the Laminarinases from Vibrio breoganii 1C10.

Ahmet Badur[1], Matthew Plutz[1], Geethika Yalamanchili[1], Jan-Hendrik Hehemann[2], Martin Polz[2], and **Christopher V. Rao**[1]

[1]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana- Champaign;
[2]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology

Project Goals: This project will harvest 'biomass to biofuel' pathways from algae- associated 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 have proven potential as feedstocks for biofuel production, and have several advantages over other marine algae. Laminarin is one of the major polysaccharides of brown algae. It is one of the least complex carbohydrates in brown algae and consists of β-1,3 and β-1,6 linked glucose residues. Glycoside hydrolases (β-1,3-glucanases) that cleave the β-1,3 linkage belong to the seven GH families: GH3, GH5, GH16, GH17, GH55, GH64 and GH81. However, the glycoside hydrolase that degrades the β-1,6 linkage in laminarin remains unknown.

We are investigating the mechanism of laminarin metabolism in marine *Vibrios*. As a first step, we have cloned, purified, and performed an initial characterization of four laminarinases from *Vibrio breoganii* 1C10. These results represent a first step towards identifying the pathways for laminarin metabolism in marine *Vibrios.*

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.

94. Cloning and Characterization of the Alginate Lyases from *Vibrio splendidus* 12B01 and 13B01

Mathew Plutz, Ahmet Badur, Geethika Yalamanchili, and **Christopher V. Rao**

Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana- Champaign

Project Goals: This project will harvest 'biomass to biofuel' pathways from algae- associated 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 have proven potential as feedstocks for biofuel production, and have several advantages over other marine algae. Alginate is one of the major polysaccharides of brown algae. It is a heterogeneous polymer of two uronic acids, β-D-1,4-mannuronate (M) and its C5 epimer α-L-1,4-guluronate (G). These residues appear as homopolymeric MM and GG blocks or in a heteropolymeric distribution GM. Alginate is degraded by alginate lyases which are abundant in marine bacteria, but the enzymes for initial attack and subsequent catabolism of the building blocks are poorly characterized or unknown.

We have investigated two strains of *Vibrio spendidus* (12B01 and 13B01) for their ability to degrade alginate. We performed a preliminary analysis of the secretome from 12B01 and 13B01 using shotgun proteomics. LC-MS/MS identified 5 putative lyases in the 12B01 secretome and 2 in the 13B01 secretome. We have also cloned, purified, and enzymatically characterized eight alginate lyases from 12B01 and six from 13B01. These enzymes were identified based on sequence homology to known alginate lyases and include the lyases identified by LC-MS/MS.

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.

95. Functional overexpressions and characterizations of lipogenesis-related genes in oleaginous yeast Yarrowia lipolytica

Kangjian Qiao[1,*] (kjqiao@mit.edu), Andrew Silverman[1,*](asilver@mit.edu), Peng Xu[1], and **Gregory Stephanopoulos[1]**

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA

Project Goals: We focused on achieving a fundamental understanding of the metabolic pathways of the oleaginous yeast *Yarrowia lipolytica* and developing tools to characterize and engineer it. 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.

As an oleaginous yeast, *Yarrowia lipolytica* can naturally accumulate more than 20% of its biomass as triacylglycerols. Recently, due to the availablilty of its sequenced genome and a limited set of genetic tools, it was extensively investigated as a model organism for *de novo* lipid biosynthesis and accumulation. In spite of previous efforts, many gene functions of *Y. lipolytica* are still unknown. Therefore, in our study, we applied a systematic approach to identify and investigate a diverse group of lipogenesis-related genes of *Y. lipolytica*. Many aspects to lipid synthesis are represented in our set of gene targets, including glycerolipid synthesis pathways, the fatty acid synthesis pathway, NADPH generation, transcriptional and protein-level regulators, lipid transporters, and central carbon metabolism. So far, 35 of the selected genes were cloned into a previously established expression platform, featuring the strong constitutive TEF (Transcription Initiation Factor-1α) promoter, and transformed into *Y. lipolytica* po1g; elevated transcription levels for each construct was demonstrated by RT-qPCR. Strains were evaluated in fermentations using either glucose or acetate as sole carbon source and the total lipid titer and content was quantified via GC-FID.

Preliminary results are encouraging, as many of our single-overexpression constructs demonstrated an increase in lipid titer and dry cell weight fraction over wild-type in our fermentations. In particular, overexpression of DGA2 (YALI0E07986g) was demonstrated to significantly boost the lipid content (~55%) as its counterpart DGA1 (YALI0E32769g). On the other hand, overexpression of two other diacylglyceride acyltransferases DGA (YALI0F06578p) and PDAT (YALI0E16797g) elevate the lipid content by 1.4 fold. Together, this evidence supports the hypothesis that integrating intracellular free fatty acid into neutral lipid (located at lipid droplet) is rate limiting step and overexpression of appropriate enzymatic steps can efficiently activate de novo lipogenesis and storage. Moreover, moderate increase (1.2 – 2.1 fold) of lipid titer and content were observed by overexpressing most genes in the Kennedy pathway using both glucose and acetate as the carbon sources, indicating that an increase in the concentration of all Kennedy pathway intermediates could facilitate the biosynthesis of triacylglycerides. Aside from just Kennedy pathway intermediates, overexpressing glycerol-3-phosphate dehydrogenase to produce more of the glycerol backbone significantly increases lipid titer and content in glucose fermentations. Additionally, overexpressing the delta-12 desaturase didn't seem to facilitate lipogenesis but to a considerable extent changed the lipid distributions: the content of C18:2 is increased by 9- fold.

Our results strongly suggest that the oxidative pentose phosphate pathway (oxPPP) is the major NADPH generation pathway for lipid synthesis in *Y. lipolytica*. Of the several gene products that convert $NADP^+$ to NADPH, only those involved in oxPPP conferred increases in lipid titer and content to

strains that overexpressed them, and every oxPPP gene (ZWF1, SOL3, and GND) overexpression led to increased lipogenesis. These increases in performance were only present with glucose as the carbon source. One surprising result was that the most effective of these was SOL3, the intermediate step (and the one that does not by itself produce NADPH), with increases in end-point fermentation lipid content over wild-type ranging from 13% in high carbon-nitrogen ratio media to 177% in lower C:N media. The dependence of comparative results of these strains on the carbon-nitrogen ratio in the media possibly signifies that in *Y. lipolytica*, oxPPP flux is increased in response to nitrogen exhaustion during fermentation.

Lastly, although only a few of our regulators have been tested so far, we saw significant improvements in glucose fermentation performance when overexpressing one particular transcription factor (YALI0C02387g, annotated as "YAS1"). YAS1 is a helix-loop-helix TF demonstrated to be involved in up-regulating genes needed to utilize alkanes as a carbon source. However, *Y. lipolytica* YAS1 also has high homology to the INO4 transcription factor in *Saccharomyces cerevisiae*, which is involved in expressing genes needed for glycerolipid synthesis, so the exact mechanism of YAS1 influencing lipid synthesis is unclear at this time. Moving forward, we aim to test more potentially lipogenic genes as well as conduct further experiments on our most effective targets to elucidate their mechanisms of action as they relate to lipid synthesis.

96. Characterization of Algal Polysaccharide Degrading Enzymatic Machinery and its Refactoring via DNA Assembler for Biofuels Production

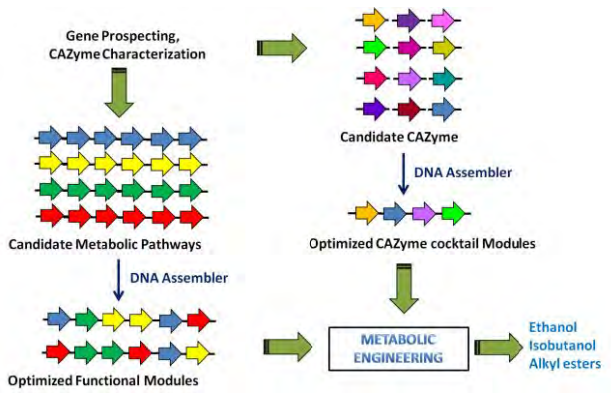
Sujit Sadashiv Jagtap[1,2]* (jagtap@illinois.edu), Eva Garcia Ruiz,[2] Jan-Hendrik Hehemann,[3] Martin F. Polz,[3] Jung-Kul Lee,[1] **Huimin Zhao**[2]

[1]Department of Chemical Engineering, Konkuk University, Seoul; [2]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Illinois; [3]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts

http://scs.illinois.edu/~zhaogrp/

Project Goals: We will employ new methods like DNA Assembler and COMPACTER developed by the Zhao group for rapidly assembling and optimizing complex multi-gene pathways for degrading complex algal polysaccharides using candidate genes from our collection of marine bacteria. A key advantage of this approach is that our designs will not be limited to the metabolic capabilities of a single organism. Rather, our methodology enables us to combine diverse metabolic functionalities from multiple bacteria in a single host. In particular, we will be able to assemble different combinations of genes in parallel from a large and diverse library of marine bacteria. In addition, we will use this approach to rapidly optimize these pathways by altering their promoters, ribosome binding sites, codon bias, and organization within operons for expression in the industrial bacterium *E. coli*. This will enable us to couple our designs with designs previously developed in *E. coli* for production of ethanol, isobutanol, and alkyl esters.

Algal polysaccharides constitute an important carbon and energy resource for marine organisms and have been considered as cost-competitive biomass for the production of biodiesel, bioethanol, and biohydrogen. It is restricted by the availability of tractable microorganisms that can metabolize alginate polysaccharides. More efforts are required to harness the enzymatic machinery that bacteria use to convert marine algal polysaccharides into bioenergy substrate. Here, we present the characterization of three oligoalginate lyases (Oals) and four alginate lyases (Alys) from *Vibrio splendidus* that catalyze depolymerization of alginate into monomer substrates which later converted to biofuels. OalA (PL-15) and OalB, OalC (PL-17) from *V. splendidus*, act on alginate polysaccharides and oligosaccharides release monosaccharides from the substrate terminus. OalA was purified as an active soluble form using MBP-tag affinity chromatography. Protein refolding of OalB and OalC using the flash dilution method resulted in refolding of the protein into its native structure and regaining full biological activity. Exotype OalA had activities toward both poly β-D-mannuronate (polyM) and poly α-L-guluronate (polyG), indicating that it is a bifunctional alginate lyase. The variation in the kinetic parameters for the OalA reaction as a function of substrate length suggests that the enzyme is well adapted to process the oligomers that are imported into the cell. The turnover number was inversely proportional to degree of polymerization of substrates. OalA showed highest activity towards the trimeric substrate, while OalB and OalC preferred dimers among various oligomers. Saccharification of



alginate by OalA produced high concentration of monosaccharides, a substrate for biofuels. Alys belonging to the PL-5 and PL-7 families acted on alginate polymers endolytically and in some cases exolytically to produce oligosaccharides and monomers. AlyA (PL-5) and AlyB (PL-7) specifically depolymerized heteropolymer poly MG in alginate molecules into oligomers. AlyC and AlyD belong to family PL-14 are specific for polyM, and polyG, respectively. Oals and Alys can be used as biocatalysts for saccharification of alginate since they can efficiently degrade alginate and alginate oligomers into alginate monosaccharides. The detection, quantification, and structure determination of the alginate products were carried out using TLC, HPLC, ESIMS or NMR. The characterized algal polysaccharide degradation enzymatic machinery and the redesigned alginate degradation cluster by addition of novel alginate lyases and removal of redundant enzymes are refactored using the DNA assembler method to create recombinant bacterial and yeast strains capable of producing biofuels and chemicals.

This research was supported by U.S. Department of Energy (grant # DE-SC0008743).

97. Resolving central metabolism of wild type and engineered *Yarrowia lipolytica* by ^{13}C- metablic flux analysis with multiple isotopic tracers

Woo Suk Ahn* (wsahn@mit.edu), Thomas M. Wasylenko* (twasylen@mit.edu) and **Gregory Stephanopoulos**

Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA

http://bamel.scripts.mit.edu/gns/

Project Goals: The project aims to develop ^{13}C-metabolic flux analysis (^{13}C-MFA) methodology for flux determination in *Yarrowia lipolytica* with high resolution and accuracy. The generated flux information will be applied to identify gene targets and develop metabolically engineered cells for biodiesel production.

The oleaginous yeast *Yarrowia lipolytica* can accumulate large quantities of lipids in the form of triacylglycerol (TAG) in its lipid body, making it an attractive host for biodiesel production. To increase TAG production, *Y. lipolytica* needs to be engineered through genetic manipulations, e.g. gene knockout and over-expression. Thus, the determination of *in vivo* metabolic fluxes can provide critical information for the identification of target genes and bottle- neck pathways and the evaluation of engineered strains.

Here, we constructed a metabolic model of *Y. lipolytica* and performed ^{13}C-MFA, which can estimate *in vivo* metabolic fluxes, including net fluxes as well as the rates of reverse reactions in central metabolism. Generally, ^{13}C-MFA utilizes substrates with ^{13}C-atoms, i.e. isotopic tracers which are incorporated into intracellular metabolites by biochemical reactions and the labeled metabolites are analyzed with mass spectrometry. However, one type of isotopic tracer provides only limited flux observability in terms of continuous atom transitions in central metabolism. To overcome this shortcoming, we combined analysis of multiple labeling data sets obtained by parallel experiments with different ^{13}C-labeled tracers. As a result, we achieved high flux observability in *Y. lipolytica* central metabolism with high resolution and accuracy compared to individual tracer experiments.

Using the combined analysis with ^{13}C-MFA, we evaluated wild type and engineered *Y. lipolytica* over-expressing acetyl-CoA carboxylase (ACC) and diacylglycerol acyltransferase (DGA). We successfully estimated most of the fluxes through the pentose phosphate pathway (PPP), glycolysis, citric acid cycle and lipid metabolism, including reaction reversibility, with the combined analysis. Furthermore, the flux maps indicated the metabolic rewiring of engineered cells for production of lipids. In the case of NADPH generating pathways, we observed a dramatic increase in the oxidative PPP flux in the engineered strain relative to the wild type strain. In contrast, we observed little change in the malic enzyme flux in the ACC DGA strain. Taken together, these results suggest the oxidative PPP is the primary source of NADPH for lipid over-production. In sum, the high resolution and accurate flux determination by combined analysis allowed us to evaluate engineered cells and to identify key enzyme targets for further engineering

This research is supported by the U.S. Department of Energy (DE-SC0008744).

98. Next generation of the Yeast Genome-Scale Model: Reformulation and Improvement for Future Data Integration

Antonio Marras[1,*], Amir Feizi[1], Jens Nielsen[1], **Greg Stephanopoulos**[2]

[1]Chalmers University of Technology, Gothenburg, Sweden, [2]MIT, Cambridge, MA
*marrasa@chalmers.se

Project goals: The main goal of this project is to extend the current yeast genome scale models towards the next generation in order to capture current knowledge and use them as a scaffold for future data integration and modeling.

Genome-Scale Modeling and Flux balance analysis (FBA) are powerful tools to investigate the metabolic behavior of microbial cells. The recent development in high-throughput technologies produced a huge amount of data to be integrated and interpreted in their biological context. Integration of such high-throughput data with Genome-Scale models has valuable benefits for the understanding of the cell behavior. Nevertheless, this integration process with genome-scale models have been challenging in the area of research also due to a lack of a detailed description of the functions which have to be integrated.

Accordingly, we aim to reformulate and improve the current genome-scale yeast reconstruction approach to include more biological knowledge such as enzyme complex information which the previous approaches were lacking. For this aim, we use the knowledge available from structural proteomics and molecular level biochemistry in order to extend and accurate the gene-protein-reaction (GPR) relationships. Furthermore, the translation machineries and secretory machinery models will be reconstructed and integrated to the reformulated metabolic model. This provides the possibility to connect and investigate the metabolic demands of important cell processes in the cell.

The reformulation of the genome-scale models to include more molecular level mechanistic details will broad our ability to integrate various *-omics* data in order to get more systemic level understanding while it also increase the predictability property of the models.

References

Lerman, J. A., Hyduke, D. R., Latif, H., Portnoy, V. A., Lewis, N. E., Orth, J. D., Schrimpe-Rutledge, A. C., Smith, R. D., Adkins, J. N., Zengler, K., et al. (2012) In silico method for modelling metabolism and gene product expression at genome scale. Nat Commun, 3, 929.

99. Regulation of lipid metabolism in *Yarrowia lipolytica*

Eduard J Kerkhoven[1,*], Jens Nielsen[1], **Greg Stephanopoulos**[2]

[1]Chalmers University of Technology, Gothenburg, Sweden, [2]MIT, Cambridge, MA
*eduardk@chalmers.se

Project goals: This project aims to elucidate the regulation of lipid metabolism in *Y. lipolytica* under different growth conditions. This will allow us to identify new targets to improve the TAG yield.

Oleaginous yeasts such as *Yarrowia lipolytica* are capable of accumulating lipids up to 70% of their biomass, predominantly in the form of triacylglycerols (TAGs), and this has fuelled interest in exploiting these fungi for the production of biodiesel. *Y. lipolytica* can use different carbon sources—including a genetic mutant that can grow on xylose—but TAG yields vary considerably under different growth conditions. To further optimise these yields, we are studying the metabolic fluxes and their regulation in *Y. lipolytica* in different growth conditions on a genomic scale.

An updated genome-scale model of metabolism (GEM) was generated for *Y. lipolytica*, based on literature data and homology with related species. A set of possible flux distributions can be obtained by random sampling of the GEM solution space, constrained by a small set of measured metabolic fluxes. The obtained statistics allow estimation of the significance of changes in metabolic fluxes in different conditions, and this can be compared to observed changes in transcript levels.

To validate the model, a genetic mutant overexpressing the last enzyme of TAG synthesis was compared to wild-type *Y. lipolytica* during growth at high and low C/N-ratios.

Measurements were taken of metabolic fluxes, lipid composition and transcript levels. Integration of the experimental data with the *in silico* model gives insight on the regulation of *lipolytica* metabolism and how the TAG yield can be further improved.

100. Predicting Gene Targets for Optimizing Oil Production in *Yarrowia lipolytica* by Cell-Wide Ensemble Modeling

Yun Lee[1]* (yun.lee@ucla.edu), Jimmy G. Lafontaine Rivera[1], James C. Liao[1, 2], and **Gregory Stephanopoulos[3]**

[1]Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, 5531 Boelter Hall, Los Angeles, CA; [2]UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, 611 Young Drive East, Los Angeles, CA; and [3]Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA

Diesel is a valuable fuel due to its high-energy density and better (than gasoline) thermal efficiency as transportation fuel. Additionally, it can be produced from renewable resources and, as such, it can contribute significantly to mitigation of greenhouse gas generation and enhancement of domestic feedstocks for production of transportation fuels. To date, all biodiesel production facilities rely on vegetable oils and animal fats as feedstocks. Such feedstocks, however, are very limited (maximum capacity of ~1B gallons/year). As a result, there is a need for the cost-efficient production of lipids from carbohydrate feedstocks, which are very plentiful throughout the world and rather well distributed in various forms (monosaccharides, starches, various types of cellulosic and hemicellulosic biomass).

Recently, it has been shown that an engineered strain of *Yarrowia lipolytica* is capable of converting, at very high yields, carbohydrates to oils and lipids for biodiesel production. The current conversion yields achieved by this engineered strain suggest that oil for biodiesel can be produced at a cost of approximately $3/gallon, depending on the feedstock used. Moreover, this engineered oleaginous yeast can metabolize not only simple sugars, such as glucose and sucrose, but also crude glycerol (byproduct of current biodiesel operations), and biomass and algae hydrolysates, among other feedstocks. The above figures of merit exceed significantly what has been reported to date on microbial and algal lipid production and could provide the basis of a cost-effective process for lipid production from renewable feedstocks.

However, further improvement of such strains to achieve near-optimal yields often defies biochemical intuition and stoichiometric reasoning. Alteration of enzyme levels, for example, does not change stoichiometry and, as such, cannot be captured in constraints- based models. As such, it needs to be represented by dynamic models involving enzyme kinetics. Unfortunately, the lack of kinetic parameters has hindered the utility of such approaches. Furthermore, such models typically require metabolite time-course measurement for parameter estimation and model validation. *Ensemble Modeling* (EM) was recently introduced to address these problems by mimicking the strategy of high-throughput screening of biological molecules. The approach starts from constructing a library (or ensemble) of models with different parameter sets such that all models are anchored to the same end point (flux distribution) of fermentation using the control strain. Thus, the ensemble of models spans all the allowable kinetic space while having the same steady-state flux distribution. The constraint to the end point (or steady state) significantly reduces the parameter space, and enables effective sampling of parameters. In the next step, the ensemble of models will be screened by fermentation flux measurements after genetic perturbation of enzymes (overexpression or knockout). To screen the models, each model in the ensemble is perturbed the same way as in strain construction experiments. Only the models that match the new experimental data are retained, and the rest are discarded. This model-screening step is repeated in the next round of strain engineering. After a few rounds of

screening, the remaining models in the ensemble become increasingly accurate, and can be used to generate possible targets for future strain engineering.

To apply EM to improve oil production in *Y. lipolytica*, we have reconstructed a cell- wide description of *Yarrowia* metabolism that encompasses all reactions necessary for lipid biosynthesis. Based on this reconstruction, we adopted the EM approach and generated an ensemble of kinetic models that have the same steady-state fluxes but differ in kinetic parameters. In this case, the flux data came from the 13C-labeling experiments and included measurements for various *Yarrowia* strains with distinct genetic backgrounds. Currently, all the models within the ensemble are anchored to the steady- state fluxes measured from a genetically modified strain where acetyl-CoA carboxylase (ACC1) and diacylglycerol acyltransferase (DGA1) are overexpressed. This ACC1+DGA1 strain has been shown to accumulate up to 62% of its DCW as lipids at an overall volumetric productivity of 0.143 g/L/h, making it a desirable starting point for further optimization.

To identify potential gene targets, we perturbed each model by increasing (up to 2-fold) and decreasing (down to 50%) the activity of each enzyme. Since not every model is capable of sustaining a steady state upon perturbations, we retained the models that survived the maximum perturbations and used their collective results to identify perturbations that were shown to improve the lipid accumulation. Interestingly, our results suggest that overexpression of hexokinase, oxoglutarate dehydrogenase, or cytosolic adenylate kinase, as well as knockdown of glutamate synthase or glutamine synthetase, will lead to a greater production of triacylglycerol compared to the control strain. These computational predictions will provide guidance in the next round of metabolic engineering, and the experimental results, whether they are consistent with the computational predictions or not, will be used for further screening of the ensemble.

101. Stoichiometric and kinetic modeling of phenylpropanoid metabolism in Arabidopsis

Longyun Guo[1], Rohit Jaini[2], Peng Wang[1], **Clint Chapple**[1], Natalia Dudareva1, John A. Morgan[2*] (jamorgan@purdue.edu)

[1]Department of Biochemistry, Purdue University, West Lafayette, IN; [2]School of Chemical Engineering, 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, a major component of the plant cell wall, is an inhibitory factor for lignocellulosic biofuel production because of its recalcitrance to degradation. Although progress has been made in altering plant lignin levels and composition, there still remains a lack of systematic rational design for manipulating lignin biosynthesis towards achieving optimum biofuel yield. We utilized stoichiometric modeling to analyze the energetic and carbon costs and yields of the major cell wall polymers lignin, cellulose and hemicellulose. Based upon the carbon and energetic cofactor requirements, the stoichiometric model provides quantitative constraints for computing optimal alterations to biomass composition through metabolic engineering.

In order to get a deeper mechanistic understanding of the control of lignin biosynthesis and its allosteric regulation, a preliminary kinetic model was generated starting with phenylalanine, and ending with synthesis of three lignin monomers, *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol. The parameters for the model were initially derived from *in vitro* enzymatic values and adjusted based on analysis of Arabidopsis plants with knockdown or knockout of several enzymatic steps, namely phenylalanine ammonia lyase (PAL), 4-coumarate:CoA ligase (4CL), hydroxycinnamoyl CoA:shikimate hydroxycinnamoyl transferase (HCT), *p*-coumaroyl shikimate 3'-hydroxylase (C3'H), caffeoyl shikimate esterase (CSE) and ferulate 5-hydroxylase (F5H). We compared the relative change of the ratio of lignin monomers as well as some intermediates between mutant and wild-type plants as predicted by the model against experimental results from the literature. Generally, the model is consistent with knockout experiments of C3'H, CSE and F5H, and also fits well with 4CL. However, the model was not predictive in the other knockdown cases. We are currently further refining the structure of the model and adjusting parameters through ^{13}C ring labeled Phe feeding experiments. In addition to modeling native phenylpropanoid metabolism, we are simulating the production of 2- phenylethanol, a promising biofuel molecule by introducing the heterologous pathway that competes for the common substrate phenylalanine. The maximum attainable flux ratio towards 2-phenylethanol was computed from both the stoichiometric and kinetic models.

This research is supported by the Office of Biological and Environmental Research in the US Department of Energy.

102. Inducible Extreme Expression of Cellulases in Poplar

Charleson Poovaiah* (crpoovai@syr.edu), Yao Xiao and **Heather Coleman**

Syracuse University, Syracuse, New York http://biology.syr.edu/

Project Goals: The overall goals of the project is to verify in poplar the In Plant Activation (INPACT) technology[1], which enables inducible expression of genes and accumulation of proteins at very high levels *in planta* and to evaluate the ability of cellulases produced through this technology to hydrolyze cellulose to simple sugars for fermentation.

Cost of enzymes for biofuel production is one of the major limitations for the widespread use of biofuels from lignocellulosic biomass. Production of enzymes *in planta* would decrease the amount of additional enzymes necessary for hydrolysis of cellulose. In Plant Activation technology allows for very high inducible expression of recombinant proteins *in planta*.

INPACT uses the rolling circle replication of Gemini virus to produce high levels of gene amplification and protein production. In this project, 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 β-glucosidases, involved in the hydrolysis of cellulose will be expressed with constitutive and tissue specific promoters.

Cellulases from thermophilic organisms will be codon optimized and cloned into INPACT vectors for transformation and expression at very high levels in poplar. The expression of INPACT system will be controlled by the alcohol-inducible transgene expression system. The yield and hydrolytic activity of the cellulases produced will be tested.

References

Dugdale, Benjamin, et al. "In Plant Activation: An inducible, hyperexpression platform for recombinant protein production in plants." *The Plant Cell Online* 25.7 (2013): 2429-2443.

Funding statement: This project is funded by the U.S. Department of Energy Office of Biological and Environmental Research.

103. High--throughput genomic studies using CRISPRi

Esteban Toro[1]*, Xiaowo Wang, Guillaume Cambray[1], Honglei Li, Stanley Qi[3], and **Adam P. Arkin[1]**

etoro@berkeley.edu

Department of Bioengineering, UC Berkeley

We recently showed that dCas9 can be used as a transcriptional-control tool in a variety of species, ranging from bacteria to mammals. We have optimized this technology and adapted it to high throughput studies. Using next-generation sequencing and a library of oligos that targets >10000 annotated functional elements in the E. coli genome (coding sequences, promoters, transcription factor binding sites, small RNAs), we were able to investigate the effect of all 10000 features in a single tube, in a single experiment. Our results agree well with published results on coding sequence effects. Because our system is based on inducible phenotypes, however, we are able to investigate the effect of important genes in any condition. We demonstrate this by finding that while NrdA and NrdB are essential aerobically (and thus generally classified as essential by existing databases), they are dispensable anaerobically. Our data also provides a rich source of information about regulatory effects in E. coli which allows us, for example, to ascribe deleterious effects of transcription factor knock-downs to particular binding sites. Our technology can be easily applied to any organism in which dCas9 works, and thus should prove of general interest as a high-throughput discovery tool.

104. Repurposing the Yeast Peroxisome to Compartmentalize Engineered Metabolic Pathways

William C. DeLoache,[1] Zachary N. Russ,[1] **John E. Dueber**[1,2*] (jdueber@berkeley.edu) [1]University of California, Berkeley, [2]Lawrence Berkeley National Labs, Berkeley, California

Project Goals: Engineered metabolic pathways often suffer from undesired interactions with the production host's native cellular processes. Evolution has solved the problem of metabolic crosstalk by segregating distinct cellular functions into membrane-bound organelles. The goal of this project is to repurpose one of these organelles – specifically the peroxisome of *Saccharomyces cerevisiae* – for compartmentalizing heterologous metabolic pathways. Towards this goal, we are working to 1) improve the targeting efficiency of non- native enzyme cargo to the peroxisome, 2) determine the natural chemical composition of the peroxisomal lumen, 3) establish methods for altering this chemical composition to suit the needs of new enzymatic pathways, and 4) demonstrate successful compartmentalization of a model pathway. Ultimately, this work will contribute to the development of a synthetic organelle that can limit metabolic crosstalk and improve production efficiency for a variety of engineered pathways.

We selected the peroxisome as the basis for a synthetic organelle for a few key reasons. First, and most importantly, loss of the peroxisome does not adversely affect *S. cerevisiae* growth provided that long chain fatty acids are not used as the sole carbon source. Second, while peroxisomes occupy only a small fraction of the total cell volume under normal growth conditions, they are capable of expanding dramatically when induced. In some methanol utilizing yeasts such as *Pichia pastoris* and *Hansenula polymorpha*, methanol induction generates massive peroxisomes that often take up more than 70 percent of the total cell volume—an appealing prospect for high-flux engineered pathways. Third, unlike other organelles, the peroxisome's cargo protein is imported directly from the cytosol in the folded state, reducing the probability of protein misfolding for non-native cargo. Finally, many organisms have already repurposed the peroxisome to perform specialized functions—evidence of the organelle's inherent plasticity.

There are two natural targeting tags used for importing protein into the peroxisome: the C- terminal PTS1 tag used by the vast majority of native cargo and the N-terminal PTS2 tag used by just four peroxisomal enzymes. We compared the efficiency of the PTS1 and PTS2 tags by analyzing the peroxisomal import rate of a heterologously expressed enzyme. For our assay, we utilized the bacterial enzyme vioE, which produces the easily detectable green pigment prodeoxyviolacein (PDV). When vioE is expressed in the cytosol of *S. cerevisiae*, it generates large amounts of PDV, resulting in green colonies. However, when sequestered in the peroxisome, vioE is unable to access its substrate, and no PDV is produced. Using this assay, we found that PTS1 import of vioE was many times more efficient than PTS2. In fact, the peroxisomal PTS1 tag proved to be more efficient at sequestering vioE than canonical targeting tags for the vacuole, mitochondrion, or extracellular secretion. Even with a PTS1 tag, however, we still observed detectable levels of cytosolic vioE when the enzyme was expressed at the very high levels utilized by most metabolic engineers. To further improve the import efficiency of PTS1-tagged cargo, we constructed a randomized library of six amino acids preceding the native PTS1 tag and screened for colonies that showed reduced levels of vioE in the cytosol. The output of this screen showed a pronounced trend for improved import when basic residues were preceding the PTS1 tag. Based on these results, we now have a sequence-optimized targeting tag that is capable of importing protein to the peroxisome extremely efficiently. This tag is also modular—we have shown that it maintains its efficiency when fused to a variety of cargo proteins.

In addition to efficient targeting of protein cargo, our strategy demands that we have control over the metabolite pool within the peroxisome. We are addressing this problem in two ways:

1. We are attempting to determine what metabolites are naturally present in the peroxisome and to understand how they get there. Previous studies on the chemical composition of the peroxisome have employed either fluorescent biosensors or purified peroxisomes. We have opted to instead utilize an enzyme-based, *in vivo* approach that is more sensitive than fluorescent biosensors and more physiologically relevant than purified systems. Our assay uses our optimized PTS1 tag to target enzymes of interest to the peroxisome. By comparing enzymatic activity with and without peroxisomal localization, we can determine whether the substrate for each enzyme is present in the peroxisome. Thus far, we have found that a surprisingly high number of small metabolites are present in the lumen of the peroxisome, especially given the existence of only two known metabolite transporters on the membrane of the peroxisome. We believe these metabolites get through the peroxisomal membrane by freely diffusing through a non-specific pore protein—something that has been hypothesized in the literature but has yet to be identified or confirmed *in vivo*. Our current efforts are focused on identifying this pore protein so that we can knock it out in our engineered system.

2. In addition to clearing out metabolites that are natively found in the peroxisomal lumen, we are also developing methods for introducing new metabolites into the lumen via membrane transporters. By fusing the transmembrane segment of a native peroxisomal protein to a variety of plasma membrane transporters, we have demonstrated that we can redirect these transporters to the peroxisomal membrane. The apparent permeability of the peroxisome membrane to small metabolites makes it challenging to assay for the activity of these transporters. Thus, we are currently in search of transporters for larger metabolites that do not get into the peroxisomal lumen natively.

Finally, we are applying our strategy of peroxisomal compartmentalization to a model pathway. We are utilizing a branched pathway in which two enzymes act on the same substrate to generate different products. By producing the shared substrate in the peroxisome, we hope to show that we can control which enzyme acts on it simply be controlling the localization of the downstream enzymes (cytosol vs. peroxisomal). This system mimics a problem that is often encountered in metabolic engineering where a pathway intermediate is lost to a side pathway that happens to exist natively within the cell. We hope that our strategy of peroxisomal compartmentalization will help metabolic engineers overcome this problem as well as many others that arise when engineered pathways are mixed with a production host's native cellular machinery.

105. Expanding the breeder's toolbox for perennial grasses: Engineered doubled-haploid and gene containment systems

Sangwoong Yoon[2], Ray Collier[2], Maria Reguera[1], Maureen Daley[1], Matthew Wright[1], Abdel-Tawab YM[1], Sean Gordon[2], **John Vogel[2]**, **Eduardo Blumwald[1]**, **Roger Thilmony**[2], **Christian Tobias**[2*] (christian.tobias@ars.usda.gov)

Project Goals: The project aims at using a systems-based approach to develop new breeding tools for perennial grasses and apply these tools towards the improvement of switchgrass (*Panicum virgatum* L.). Our objectives are: (1) Accelerate conventional breeding using the fast generation of doubled haploid lines (developing a CENH3-based method in switchgrass); (2) Use the model perennial grass *Brachypodium sylvaticum* to identify combinations of transgenes that confer tolerance to multiple abiotic stresses; (3) Develop a gene containment system to minimize gene flow from transgenic switchgrass; (4) Create transgenic switchgrass plants containing the best combinations of transgenes identified in objective 2 and the gene containment system from objective 3; (5) Evaluate the best transgenic switchgrass plants from objective 4 in field trials.

Switchgrass is a large, perennial, outcrossing species native to most of North America. It carries wind-pollinated flowers and fields can produce substantial pollen during flowering. Small seeds which exhibit dormancy can be unintentionally transported by farm operations. These features which are not unique to switchgrass may improve individual fitness and preserve genetic diversity but also present difficult dilemmas for modern crop improvement by creating natural and artificial conduits for gene flow into the environment. This issue, along with concerns of the public and stakeholders with genetic diversity and maintaining differentiation among multiple potential end-uses are driving our work to develop genetic isolation technologies which will mitigate gene flow.

Evolving, complex, overlapping regulatory frameworks cover the release of engineered, transgenic switchgrass. To meaningfully evaluate new lines, field studies will need to be performed under strict oversight and gene containment measures will likely be required in some form. Our strategy employs a two component system that when combined in an F1 hybrid, will activate reproductive containment. The general design utilizes two types of transformation constructs that function together using recombinase-mediated excision.

The first component consists of a plant expression vector that may harbor four or more transgenes: (1) a 'gene of interest'; (2) a selectable marker; (3) the *Cre* recombinase; and (4) *Barnase* which encodes a ribonuclease that is toxic when expressed in plant cells. Transgenic plants generated with this construct will express the gene of interest, but the transgenic pollen produced from these plants will be inviable due to pollen-specific Barnase expression. We have demonstrated that the rice *Pollen Specific3* (*PS3*) promoter is active in switchgrass pollen and we are using this promoter to control Barnase expression. These hemizygous transgenic plants can be used as the female parent for generating hybrid switchgrass seed.

The second component of the proposed strategy utilizes plant expression vectors that are capable of harboring a second gene of interest, selectable marker, and the cytotoxic *Diphtheria Toxin A* (*DTA*) gene. Expression of DTA is initially blocked by the presence of a *lox*-flanked triple terminator 'stuffer' sequence (35Sterm, nosTerm and ocsTerm). This allows transgenic plants generated with these vectors to be fully fertile (due to the stuffer-mediated blocking of DTA expression). When these fertile plants are crossed with transgenic plants containing the first component (used as the female parent), hybrid plants containing both components will be generated. The hybrids constitutively

expressing the Cre recombinase will undergo site-specific recombination that excises the stuffer region flanked by the directly oriented *lox* sites. Following stuffer excision, the *DTA* open reading frame will be fused to the upstream pollen-specific promoter separated only by a single 34bp *lox* sequence, thus transgene expression and reproductive ablation will be activated. Initial testing of the hybrid transgene containment system will be performed using hemizygous transgenic lines for crossing. Hybrid genotypes containing both constructs will be identified with seedling selection and molecular screening.

Our gene containment strategy would function far more efficiently with availability of homozygous-inbred lines because these can be crossed to produce 100% hybrid seed containing both gene containment constructs. Fully homozygous lines are also advantageous for breeding, particularly for exploiting heterosis and major QTL or advancement of new transgenic lines. Rapid creation of inbreds via doubling of haploid lines is feasible in some species, but haploids are difficult to obtain in most species including switchgrass. We are currently exploring the potential of centromere-mediated genome elimination to efficiently produce haploids. This approach has been successfully applied to Arabidopsis, but not to other plant species. In Arabidopsis it was observed that substitution of wild-type CenH3 for one of several mutant versions can lead to creation of semi-fertile haploid-inducer lines. These are believed to act through an early embryonic defect resulting in mis-segregation and elimination of chromosomes derived from the haploid inducer line. Mature haploid embryos or young plants can subsequently be artificially doubled by chemical treatment to produce completely homozygous diploid lines. Switchgrass genotypes that we work with are tetraploid, and we have found that switchgrass contains two copies of *CenH3* which we have designated *PviCenH3-1* and *-2*. We are now attempting to create functional knockouts of both genes using TALENs. These lines will be assessed for haploid induction rates. As this is new and unproven technology for switchgrass, we are also attempting a parallel approach using *Brachypodium distachyon*. Evidence in support of haploid induction can be achieved more quickly in this species. To date we have targeted 10 separate loci using TALEN pairs designed to induce double strand DNA breaks. Each pair is coordinately transcribed via a single maize ubiquitin promoter and is separated by a T2A translational skipping sequence. Three pairs target *PviCenH3-1*, three target *PviCenH3-2*, and four target *BdiCenH3*.

106. Expanding the breeder's toolbox for perennial grasses: The use of the model perennial grass *Brachypodium sylvaticum* to identify combinations of transgenes conferring tolerance to multiple abiotic stresses

Maria Reguera[1], Maureen Daley[1], Matthew Wright[1], Yasser Abdel-Tawab[1], Sean Gordon[2], Sangwoong Yoon[2], Ray Collier[2], **Christian Tobias**[2], **Roger Thilmony**[2], **John Vogel**[2], **Eduardo Blumwald**[1] **(eblumwald@ucdavis.edu)**

[1]Dept. of Plant Sciences, University of California, Davis, CA 95616; [2]USDA-ARS Western Regional Research Center, Albany CA 94710

Project Goals: The project aims at using a systems-based approach to develop new breeding tools for perennial grasses and apply these tools towards the improvement of switchgrass (*Panicum virgatum* L.). Our objectives are: (1) Accelerate conventional breeding using the fast generation of doubled haploid lines (developing a CENH3-based method in switchgrass); (2) Use the model perennial grass *Brachypodium sylvaticum* to identify combinations of transgenes that confer tolerance to multiple abiotic stresses; (3) Develop a gene containment system to minimize gene flow from transgenic switchgrass; (4) Create transgenic switchgrass plants containing the best combinations of transgenes identified in objective 2 and the gene containment system from objective 3; (5) Evaluate the best transgenic switchgrass plants from objective 4 in field trials.

B. sylvaticum possesses all the traits necessary to serve as a model perennial grass. Despite its perennial nature, some accessions go from seed to seed in 3-5 months with no vernalization. It is self-fertile. It possesses a small genome size of approximately 350 Mb and a chromosome number of 2n=2x=18. In addition, the close relationship between *B. sylvaticum* and *B. distachyon* allows leveraging the resources developed for *B. distachyon*. Unlike *B. distachyon*, *B. sylvaticum* flowers open, the stigmas exert before the anthers and the anthers shed copious amounts of pollen. These traits and its perennial nature are useful for studies that involve male/female sterility and many crosses.

We have developed a highly efficient transformation system, based on our *B. distachyon* transformation protocol. We have almost completed the transformation of *B. sylvaticum* with constructs comprising 20 genes listed in our proposal (Peleg et al., 2011; Reguera et al., 2013). These genes have been shown to be associated with the enhanced tolerance of monocots to abiotic stress. We have developed a highly efficient transformation system for *B. sylvaticum*, modifying our *B. distachyon* transformation protocol (average transformation efficiency = 55%). In most cases we used the stress-inducible SARK promoter (Delatorre et al., 2012). In some cases (i.e. WRKY47, HB4, NHX1, HSR1), where the constitutive expression is sought we also transformed under the control of the maize ubiquitin promoter sequence (Ubi). In a few cases, we will express the gene-of- interest under more restrictive promoter, in order to avoid secondary effects that could influence growth. In these cases (i.e., WRKY47 and the kinases OsK1 and OsK24) we will use a chemically-induced promoter that allows for tight control of gene expression. In addition, we have made constructs containing a combination of genes (i.e. SARK::IPT- Ubi::HSR1::Ubi::NHX1) in order to simultaneously overexpress genes associated with drought + heat tolerance + salt tolerance. We will make additional combination constructs after we have characterized the single gene constructs. We generated a large number of independent T0 lines for these constructs and are now in the process of generating T1 and T2 generations.

Enhanced stress tolerance of the transgenic plants will be assessed in the greenhouse. Wild type and transgenic plants will be grown in controlled greenhouse conditions and plants will be characterized

for response to a combination of water deficit, heat and salinity stress. Leaf material will be sampled periodically for the measurement of chlorophyll, starch, total and reducing sugars, and protein contents. Net assimilation rates, stomatal conductance and quantum efficiency will be obtained from analysis of gas exchange, during and after stress. Expression analysis of the different transgenes will be performed by quantitative PCR (endogenous transcription elongation factor (*TEF*) will be use as control).

To aid our planned RNA-Seq studies of the *B. sylvaticum* transgenics, we will sequence the genome of *B. sylvaticum*. In preparation, we have inbred the target line, Ain-1, through seven generations of single seed descent to decrease heterozygosity. We are now extracting DNA for sequencing at the UC Davis genomics center. In addition, we have created mapping populations and have harvested F_2 seeds. We will use genotype by sequence to create a high resolution genetic map based on ~500 F_2 individuals. Also, since *B. sylvaticum* is perennial, we can maintain the F_2 lines as an immortal, fully mapped segregating population. Interestingly, F_1 plants from independent crosses between Ain-1 and three other accessions, all exhibited a high degree of hybrid vigor.

Peleg Z, Reguera M, Walia H, Blumwald E. (2011) Cytokinin mediated source-sink modifications improve drought tolerance and increases grain yield in rice under water stress, Plant Biotechnol. J.9:747-758.

Reguera M, Peleg Z, Abdel-Tawab YM, Tumimbang EB, Delatorre CA, Blumwald E (2013). Stress-Induced CK Synthesis Increases Drought Tolerance through the Coordinated Regulation of Carbon and Nitrogen Assimilation in Rice. Plant Physiol. 163:1609-1622.

Delatorre CA, Cohen Y, Liu L, Peleg Z, Blumwald E. (2012). The regulation of the SARK Promoter activity by hormones and environmental signals. Plant Sci. 193:39-47.

107. Engineering Anaerobic Gut Fungi for Lignocellulose Breakdown

Kevin V. Solomon,[1] John K. Henske,[1] Charles H. Haitjema,[1] Diego Borges-Rivera,[2] Dawn A. Thompson,[2] Aviv Regev,[2] **Michelle O'Malley[1]*** (momalley@engineering.ucsb.edu)

[1] Department of Chemical Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5080
[2] Broad Institute of MIT and Harvard, Cambridge, MA 02124

http://omalleylab.weebly.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 (fungal 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. Recently, however, advances in sequencing technologies have made it possible to explore the dynamic metabolic networks within gut fungi for the first time. Our approach combines next-generation sequencing with physiological characterization to establish the critical knowledge base to understand lignocellulose breakdown by gut fungi. This project will (1) enable exploration of novel isolates for desirable enzymatic properties, (2) construct metabolic models to describe biomass degradation, and (3) develop new methods to metabolically engineer gut fungi for bioprocessing.

To initiate this project, we isolated a panel of novel gut fungi from sheep, goat, giraffe, and elephants at the Santa Barbara Zoo. To date, four unique strains from the *Piromyces, Neocallimastix, and Anaeromyces* genera have been obtained through roll tube isolation. Proliferation of the fungal isolates was monitored via fermentation gas production, and cellulosomes from each species were isolated through cellulose-precipitation. All of these isolates exhibited high enzymatic reactivity against a range of cellulosic and lignocellulosic substrates (filter paper, Avicel, reed canary grass), which was repressed in the presence of simple sugars. Within isolated cellulosomes, striking similarities are observed for certain dockerin-fused glycosyl hydrolases, and these proteins are not secreted from fungi when simple sugars are present, supporting the hypothesis that these enzymes are catabolically regulated.

Our subsequent goal is to enumerate novel biomass-degrading enzymes within these isolates and characterize their coordinated expression during biomass breakdown. Towards this goal, we have analyzed the transcriptome of the fungal isolate *Piromyces sp finn* via RNAseq under several growth conditions. This isolate exhibited high enzymatic reactivity against a range of cellulosic and lignocellulosic substrates, which was repressed in the presence of simple sugars. Through strand-specific RNAseq and use of the TRINITY assembly platform, we were able to assemble hundreds of novel cellulase genes *de novo* from >27,000 transcripts without the need for reference genomic sequence information. The *Piromyces sp finn* transcriptome is particularly rich in GH6 and GH43 enzymes, and we find that 27 of 54 diverse glycosyl hydrolase families are transcriptionally repressed during growth on glucose relative to reed canary grass. Within the majority of these transcripts, dockerin-tagged elements of fungal cellulosomes are abundant, and 30% of dockerin-containing transcripts are repressed in the presence of glucose. This suggests that catalytic components of fungal cellulosomes are highly regulated in response to simple sugars, which also agrees with proteomic data. We will further discuss the transcriptional regulation patterns observed for important enzyme families under catabolic regulatory conditions, and connect these regulation patterns to protein expression and lignocellulosic degradation. Additionally, we are collaborating with researchers at the DOE-JGI and PNNL EMSL to sequence the genomes/transcriptomes for other isolates, as well as investigating the dynamics of cellulosome secretion as part of the 2014 Community Science Program.

This Project is supported by the Office of Biological and Environmental Research through the DOE Office of Science

108. A Strategy for Genome-scale Design, Redesign, and Optimization for Ethylene Production in *E. coli*

Sean Lynch[1,2], Carrie Eckert[1], Jianping Yu[1], Pin-Ching Maness[1]* (pinching.maness@nrel.gov), Esteban Toro[3], Adam Arkin[3], and **Ryan Gill**[2]

[1]Biosciences Center, National Renewable Energy Laboratory, Golden, CO; [2]University of Colorado, Boulder; [3]Univeristy of California, Berkeley

Project Goal: This project aims to apply state-of-the-art synthetic and systems biology tools to design and 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 will be selected for, and gene-to-trait mapping will identify key factors for further optimization. Herein, we will focus on the construction of an *E. coli* prototype chassis strain for the production of ethylene by optimization of protein expression, growth medium compositions, and metabolic pathway flux. Moreover, we are evaluating "landing pads" in *E. coli* suited for integration of genes/pathways for biofuel production, as well as 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 most CO_2-emitting process in the chemical industry. Biological ethylene production has the potential to provide a sustainable alternative while mitigating CO_2 emission. The expression of a single gene found in some bacteria and fungi, ethylene-forming enzyme (EFE), can catalyze ethylene formation (1). However, its efficient biotechnological application requires a more in-depth understanding of the interactions between the EFE reaction and other metabolic pathways in the cell, which will be afforded by genome-scale synthetic biology approaches.

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 (2):

$$3\ \alpha\text{-ketoglutarate} + \text{Arginine} + 3O_2 \rightarrow 2C_2H_4 + \text{Succinate} + 7CO_2 + 3H_2O + \text{guanidine} + \text{P5C}^*$$

However, low yield and EFE protein insolubility are key challenges at present. Our initial focus is on optimization of EFE protein expression levels, improvement of its solubility and stability, and analysis of interactions between the EFE reaction and other metabolic pathways by nutrient and genetic alterations. Rates of ethylene production are improved by three-fold when EFE is expressed from a medium vs. high copy number plasmid, yet most of the EFE protein is in the insoluble inclusion bodies based on Western data. To improve solubility, we co-expressed the GroES/EL chaperones and observed a further 7-fold improvement in ethylene productivity, concurrent with an increase in the level of soluble EFE. We are currently exploring the addition of solubility tags to EFE to further improve its solubility. To improve ethylene production based on the above reaction, we explored media composition (rich versus defined), addition of exogenous substrates (α-ketoglutarate [AKG] and arginine), and the incorporation of predicted genetic modifications to improve flux to AKG and

arginine in the TCA cycle (Figure 1), the findings from which will be reported. Lastly, work is ongoing to verify "landing pads" for the safe integration of pathway genes necessary for both ethylene and isobutanol production strains, as well as to develop methods for high-throughput selection of strains with increased production of key intermediates and/or ethylene from pooled libraries.

*P5C: L-Δ^1-pyrroline-5-carboxylate

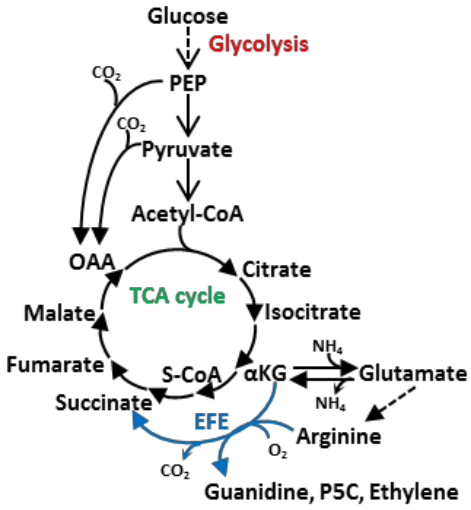


Figure 1. Putative metabolic scheme for ethylene production in *E. coli*. EFE: ethylene-forming enzyme. P5C: L-Δ^1-pyrroline-5-carboxylate.

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. Ibid. 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.

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.*

109. Tracking Combinatorial Engineered (TRACE) libraries at the genome-scale

Ramsey I. Zeitoun[1*] (ramsey.zeitoun@colorado.edu), Andrew D. Garst[1], George D. Degen[2], **Ryan T. Gill[1]**

[1]University of Colorado Boulder; [2]Carnegie Mellon University

Project Goals

The ability to perform genome engineering and editing on practical timescales requires DNA target enrichment and sequencing methods that can be used to quantitatively interrogate libraries containing billions of combinatorially engineered mutants.[1-3] Current library construction methods for genome-engineering operate at throughputs of billions per week, while tracking methods are limited to scales of 100---1000s per week or to operating in a non-combinatorial manner. The result is an inability to complete design-build-test-learn cycles at the overall throughputs required for combinatorial genome engineering (~10^8/week).

The major challenge to the development of combinatorial tracking methods is the necessity to identify mutations located at distant locations throughout a microbial genome, and to do so at high throughput and on a cell specific basis. Current methods for genotyping distant genomic sites are relatively low throughput and can be low resolution. For example, multiplexed allele specific PCR (MASC PCR) reports on the presence or absence of the wild--- type allele in single clones rather than the actual sequence of interest, and is typically limited to throughputs of 100-1000s of clones because of the need to perform individual PCR reactions on each clone.[4-6] Next-gen sequencing approaches are inexpensive, high-throughput, and high-resolution, but suffer from read lengths (several 100 nt) that are too short to allow identification of distantly located mutations, and thus also require separation and tagging strategies that reduce throughput in a manner similar to MASC PCR. Single cell sequencing approaches can be used to solve this problem, but these approaches currently do not operate at the throughputs and depth required.[7] Ideally, combinatorial tracking technology will allow one to interrogate genomes at dozens of different locations and provide quantitative data on individual combinations of mutants across combinatorial libraries containing billions of clones.

Linking PCR offers a possible genotyping solution as this method can be applied to stitch together PCR products from distal chromosomal origins in a single pot reaction. Not only does this significantly reduce the costs associated with colony-based Sanger sequencing by reducing the number of sequencing reactions per clone, but combining this method with emulsification technologies (Linking-emulsification PCR or LE-PCR) offers a potentially useful route to multiplexed genotyping of large populations (Figure 1a). The inefficiency of assembly and amplification in LE-PCR however has thus far limited its application to the study of only two distinct sites.8 To be widely applicable as a technology for tracking targeted genomic libraries, such an approach would ideally enable sampling of many genomic sites. To expand the number of gene targets that can be tracked and improve the overall efficiency of this approach requires an automated de novo primer design to optimize the many criteria necessary for efficient PCR based assembly and subsequent analysis. We thus sought to automated optimization of both the primer design and reaction conditions. This method allows us to achieve assembly of a larger number of genomic sites as proof of concept for multiplexed genotyping of distal chromosomal sites of interest. We then applied these conditions towards Tracking Combinatorial Engineered libraries by coupling the linking chemistry to emulsion PCR and resolving the products using next-generation sequencing technologies.

References

1. Wang, Harris H., et al. "Programming cells by multiplex genome engineering and accelerated evolution." *Nature* 460.7257 (2009): 894-898.
2. Sandoval, Nicholas R., et al. "Strategy for directing combinatorial genome engineering in Escherichia coli." *Proceedings of the National Academy of Sciences* 109.26 (2012): 10540-10545.
3. Warner, Joseph R., et al. "Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides." *Nature biotechnology* 28.8 (2010): 856-862.
4. Cha, Rita S., et al. "Mismatch amplification mutation assay (MAMA): application to the cH-ras gene." *Genome Research* 2.1 (1992): 14-20.
5. Wang, Harris H., and George M. Church. "Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering." *Methods Enzymol* 498 (2011): 409-426.
6. Carr, Peter A., et al. "Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection." *Nucleic acids research* 40.17 (2012): e132---e132.
7. Wang, Daojing, and Steven Bodovitz. "Single cell analysis: the new frontier in 'omics'." *Trends in biotechnology* 28.6 (2010): 281---290.
8. Wetmur, James G., et al. "Molecular haplotyping by linking emulsion PCR: analysis of paraoxonase 1 haplotypes and phenotypes." *Nucleic acids research* 33.8 (2005): 2615-2619.

110. Next Generation Multiplex Genome Engineering

Emily F. Freed* (emily.freed@colorado.edu) and **Ryan T. Gill**

University of Colorado, Boulder

Project Goals: This project is directed at the development of the next-generation of genome-engineering approaches to bridge predictive design with powerful methods for genome-scale search and optimization. Our studies combine iterations of detailed rational design, directed searches of mutational space, and high throughput mapping of protein sequence to activity. Our approach involves two objectives: i) to demonstrate an integrated, genome-scale strategy for "chassis" strain design and construction, and ii) to demonstrate a framework for "genome re-design" built upon multiplex synthetic biology and genome-engineering technologies. We have chosen to develop this platform by engineering i) ethylene production and ii) isobutanol production in *Escherichia coli*. Ethylene and isobutanol are attractive models because both can be directly converted to gasoline (and other advanced liquid biofuels) using well established catalytic routes and both have been produced in *E. coli* at various levels.

We are developing a set of tools that provides improved capabilities for linking genetic alterations to traits. Previously, the Gill laboratory developed trackable multiplex recombineering (TRMR), a technique that allows all genes in the *E. coli* genome to be barcoded and then simultaneously overexpressed or knocked down [1]. Cell populations containing barcoded expression level variants are then subjected to selective pressure and gene variants that result in increased fitness are identified by hybridization of their barcodes to a microarray. To date, TRMR has been used to map genes required for growth in various types of media and to optimize tolerance to acetate, low pH, cellulosic hydrolysate, isobutanol, ethanol, isopentenol, furfural, and various antibiotics [unpublished results, 1, 2]. These studies have given insight into carbon source and vitamin utilization, primary and secondary metabolism, and mechanisms of toxicity under a variety of conditions.

This study focuses on improving the TRMR design. The next generation TRMR library will be similar to the original library in that synthetic DNA (synDNA) cassettes will be used to replace the native promoter and RBS of a gene with a synthetic version of each, thus controlling gene and protein expression. However, the new library will allow for all *E. coli* genes to be expressed at four different levels (off, weak, intermediate, and strong), and will use a bicistronic RBS design (BCD) for more consistent expression levels across different genes [3]. Furthermore, an inducible promoter will be placed in front of each BCD allowing for fine tuning of gene expression to almost any level that is desired just by changing the amount of inducer that is added to the growth medium.

A second change will be the way in which relative fitness data is collected for each allele. In the original TRMR library, relative fitness was determined by hybridizing molecular barcodes to a microarray. In the next version of TRMR, molecular barcodes that are optimized for high throughput sequencing (e.g. Illumina HiSeq or Illumina MiSeq) will be used instead to track alleles. High throughput sequencing allows more quantitative analysis of genotype frequencies, since individual alleles will be tracked at the nucleotide level rather than by relative hybridization intensity (measured in arbitrary fluorescence units). A single run of Illumina HiSeq can generate 10^8-10^9 sequencing reads (a typical microarray signal distribution ranges over about 10^3), allowing for each barcode to be sequenced thousands of times. Another advantage is that barcodes for high throughput sequencing can be designed to be shorter in sequence length than those used for microarray, which reduces the cost of synthesizing the oligo library.

Finally, the new TRMR library is designed to integrate more easily with multiplex automated genome engineering (MAGE) [4]. One of the differences between original TRMR and MAGE is that original TRMR modifies both the promoter and the RBS while MAGE has only been used to modify the RBS. Next generation TRMR uses a synDNA cassette that can easily be modified for direct use in MAGE, which should lead to more consistent results between these two techniques and allow for even more rapid development of *E. coli* strains of interest.

The new TRMR libraries will be used in conjunction with the optimized chassis strains we are also developing to rapidly engineer and screen cell populations that have increased isobutanol or ethylene production.

References

1. Warner, J.R., et al., Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. Nat Biotechnol, 2010. **28**(8): p. 856-62.
2. Sandoval, N.R., et al., Strategy for directing combinatorial genome engineering in Escherichia coli. Proc Natl Acad Sci U S A, 2012. **109**(26): p. 10540-5.
3. Mutalik, V.K., et al., Precise and reliable gene expression via standard transcription and translation initiation elements. Nat Methods, 2013. **10**(4): p. 354-60.
4. Wang, H.H., et al., Programming cells by multiplex genome engineering and accelerated evolution. Nature, 2009. **460**(7257): p. 894-8.

This work is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

111. CRISPR-Assisted Rational Protein Engineering (CARPE) For Biofuel Production

*Gur Pines (gur.pines@colorado.edu), Andrew D. Garst and **Ryan T. Gill**

University of Colorado, Boulder

Project Goals: We are developing tools for rationally-based protein engineering which will allow multiple modifications from the single protein to the whole pathway levels. One of the main obstacles in high throughput genome modification via recombineering is the low efficiency of successful recombination events. This leads to low abundance of cells that have incorporated the desired DNA library relative to the whole population, and impedes library size and downstream processes such as screening and selection. Here, we use the CRISPR system to induce cell death of the unwanted, non-recombinant population and thus dramatically increase the levels of the modified library cells. To test this technology, we selected the DXS pathway that results (with added genes) in the production of lycopene and isopentenol. Lycopene is relatively easy to screen for, and isopentenol is considered to be a 'second generation' biofuel, with higher energy density and lower water miscibility than ethanol. We aim at engineering this pathway at the open reading frame level to increase metabolite flux and ultimately the level of the final product, isopentenol.

The hunt for better biofuels for industrial manufacturing via bacterial production requires the ability to perform state of the art genome design, engineering and screening for the desired product. Previously, our group demonstrated the ability to modify individually the expression levels of every gene in the *E. coli* genome[1]. This method, termed trackable multiplex recombineering (TRMR), produced a library of about 8000 genomically-modified cells (~4000 over-expressed genes and ~4000 knocked down genes). This library was later screened under different conditions, which enabled deeper understanding of gene products' activities and resulted in better performing strains under these selections. TRMR allowed modification of protein expression for two levels (overexpressed and knocked down) but did not enable the modification of the open reading frame (ORF). Here, we aim at producing large libraries of ORF modifications and engineering whole metabolic pathways for the optimal production of biofuels.

A major difficulty in producing such libraries, which are carefully designed (in contrast to random mutagenesis), is the insertion efficiency of the desired mutations into the target cells. Recombineering, the canonical method for genome modifications in *E. coli*, uses recombinant genes from the Lambda phage to facilitate the insertion of foreign DNA into the host genome. However, this process suffers from low efficiencies that are usually being overcome by either adding an antibiotic resistance gene followed by selection (as in TRMR), or by recursively inducing recombination events (i.e., by MAGE [2]). Our strategy to increase the recombineering efficiency involves the use of the CRISPR system to remove all non-recombinant cells from the population. CRISPR is a recently discovered RNA-based, adaptive defense mechanism of bacteria and archaea against invading phages and plasmids[3]. This system underwent massive engineering to enable sequence-directed double strand breaks using two plasmids; one plasmid coding for the CRISPR-associated nuclease Cas9 and the second plasmid coding for the sequence-specific guide RNA (gRNA) that guides Cas9 to its unique location[4]. Our method utilizes the CRISPR system's ability to induce DNA breaks, and consequently cell death, in a sequence-dependent manner. We produce DNA recombineering cassettes that, in addition to the desired mutation within the ORF, include a mutation in a common location outside of the gene's ORF which is targeted by the CRISPR machinery. This approach of linking desired mutations with the avoidance

from CRISPR- mediated death enables dramatic enrichment of the engineered cells within the total population.

To demonstrate our method, we selected the DXS pathway. This pathway results in the production of isopentenyl pyrophosphate (IPP) which results in the biosynthesis of terpenes and terpenoids. Interestingly, IPP can be the precursor of lycopene or isopentenol, given the addition of the required genes. While lycopene renders the bacterial colonies red, and hence is easily screenable, isopentenol is considered to be a 'second generation' biofuel with higher energy density and lower water miscibility than ethanol. In this project three genes were selected for engineering: 1) DXS, the first and the rate-limiting enzyme of the pathway, 2) IspB, which diverts the metabolic flux from the DXS pathway, and 3) NudF, which has been shown to convert IPP to isopentenol in both *E. coli* and *B. subtilis* [5 6]. Both DXS and IspB mutants will be screened for increased lycopene production with a new image analysis tool we developed for colony color quantification. NudF activity will be assayed directly by measuring isopentenol levels by GC/MS and indirectly by isopentenol auxotrophic cells that will serve as biosensors. This method will provide the ability to rationally engineer large mutational libraries into the *E. coli* genome with high accuracy and efficiency. Furthermore, this approach will result in the first proof-of-concept strain that produces high yield of isopentenol.

References:

1. Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. & Gill, R. T. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat Biotechnol* **28**, 856-862, doi:10.1038/nbt.1653 (2010).
2. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-898, doi:10.1038/nature08187 (2009).
3. Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual review of genetics* **45**, 273-297, doi:10.1146/annurev-genet-110410-132430 (2011).
4. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183, doi:10.1016/j.cell.2013.02.022 (2013).
5. Withers, S. T., Gottlieb, S. S., Lieu, B., Newman, J. D. & Keasling, J. D. Identification of isopentenol biosynthetic genes from Bacillus subtilis by a screening method based on isoprenoid precursor toxicity. *Applied and environmental microbiology* **73**, 6277-6283, doi:10.1128/AEM.00861-07 (2007).
6. Zheng, Y. *et al.* Metabolic engineering of Escherichia coli for high-specificity production of isoprenol and prenol as next generation of biofuels. *Biotechnology for biofuels* **6**, 57, doi:10.1186/1754-6834-6-57 (2013).

This work is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

112. Transcriptome sequencing and RNA-seq mRNA expression profiling in the facultative CAM model species *Mesembryanthemum crystallinum*.

Won C. Yim[1], Bernard W.M. Wone[1], Richard L. Tillett[1], Bahay G. Bilgi[1], Rebecca L. Albion[1], Karen A. Schlauch[1], HengfuYin[2], Gerald A. Tuskan[2], Xiaohan Yang[2] and **John C. Cushman[1]*** (jcushman@unr.edu).

[1]Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV;
[2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN.

http://cambiodesign.org

Project goals: The goal of this research 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). Two major objectives will be pursued to achieve this goal: 1) Define the genetic basis of key CAM modules in the facultative CAM species *via* network modeling using 'omics (transcriptome and metabolome) technologies; 2) characterize the transcriptional regulation of 'carboxylation', 'decarboxylation', and 'inverse stomatal control' modules of CAM. The use of the facultative CAM model species, the common ice plant (*Mesembryanthemum crystallinum*), facilitates the identification of genes recruited into CAM function and likely those needed to successfully engineer CAM into C_3 host species.

The salt-tolerant succulent common or crystalline ice plant (*Mesembryanthemum crystallinum*; Caryophylalles, Aizoaceae), can switch from C_3 photosynthesis to CAM under saline or water- deficit stress conditions (Cushman et al., 2008). To help define the transcriptome of this facultative CAM model, next-generation sequencing was performed. Roche 454 pyrosequencing of RNA from salt- and drought-stressed leaves, roots, mature flowers, seed pods, and seeds produced >4 M sequences. Illumina HiSeq sequencing of well-watered and water-deficit stress treated leaf tissues generated an additional 241.8 Gbp of sequence. Reads from each platform were assembled separately, using newbler for 454 reads, and Trinity for Illumina reads. The assemblies were combined using GICL for a final assembly of 31,238 contigs, with a mean length of 1,601 bp, and N50 value of 2,278 bp. Of these, 28,576 contigs passed contamination screening. 22,971 contigs showed sequence homology to known plant genes in the Phytozome database. 15,751 contigs had complete ORF. 37,427 terms were assigned in our contigs with putative Gene Ontology *via* protein motifs identified by InterProScan. 804 potential transcription factors were classified. Ice plant contigs with significant homology were found for >90% the Greencut2, eUCO, and AVPO reference sets of conserved genes. Broader gene family clustering identified 9,188 families of orthologous genes in OrthoMCL-DB into which 21,343 ice plant contigs clustered. 7,275 of these families included orthologs identified in all angiosperms at OrthoMCL. This assembled reference transcriptome will facilitate RNA-seq gene expression analysis of C_3/CAM photosynthesis and circadian regulation and will serve as an essential resource for genome annotation for in-progress genome sequencing.

In order to explore the circadian clock-regulated mechanisms that control the expression of CAM in the common ice plant, Illumina-based RNA-seq digital gene expression profiling was performed on well-watered and water-deficit stress treated leaf tissue. Samples were collected in parallel every 4 h over a 72 h time course under both 24 h light/dark (diel entrainment) and 48 h light/light (zeitgeber or free running) conditions in order to capture the full repertoire of circadian clock-controlled transcriptional outputs. cDNA libraries were constructed for 114 ice plant samples and Illumina HiSeq2000 was performed using a total of 19 flow-cell channels, with 6 samples in each. Approximately 2,241,015,571 bp of raw data were

generated yielding a total of 2,415,294,564 trimmed sequence reads. All trimmed, cleansed reads were assembled *de novo* using the Trinity program. The complete read dataset assembled into 109,902 contigs, which ranged in size from 201 bp to 15,797 bp with a mean contig length of 1228 bp. RSEM and Bowtie were then used to assign reads to multiple genes and count relative transcript abundances.

Read counts were mapped, normalized and differentially expressed genes were detected using the DESeq package. Circadian clock output differences between the C_3 photosynthesis and CAM states will be

required for CAM using comparative genomics by comparing gene expression patterns of known CAM components across C_3, C_4, and CAM species. These results will also aid in future ice plant genome annotation efforts.

References

Cushman JC, Tillett RL, Wood JA, Branco JA, Schlauch KA (2008) Large-scale mRNA expression profiling in the common ice plant, *Mesembryanthemum crystallinum*, performing C_3 photosynthesis and Crassulacean acid metabolism (CAM). J. Exp. Bot. 59:1875-1894.

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–AC05–00OR22725.

113. Engineering Crassulacean Acid Metabolism (CAM) to Improve Water-use Efficiency of Bioenergy Feedstocks

Anne M. Borland[1,3], James Hartwell[2], David J. Weston[3], Karen A. Schlauch[4], Timothy J. Tschaplinski[3], Gerald A. Tuskan[3], Xiaohan Yang[3], and **John C. Cushman[4]*** (jcushman@unr.edu).

[1]School of Biology, Newcastle University, Newcastle upon Tyne; [2]Department of Plant Sciences, Institute of Integrative Biology, University of Liverpool; [3]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [4]Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV.

http://cambiodesign.org

Project goals: The long-term goal of this research 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). Four major objectives will be pursued to enhance photosynthetic performance and WUE in *Arabidopsis* and *Populus*: 1) Define the genetic basis of key CAM modules in both eudicot and monocot CAM species *via* network modeling using 'omics (transcriptome and metabolome) technologies; 2) characterize the regulation of 'carboxylation', 'decarboxylation', and 'inverse stomatal control' modules of CAM using comparative genomics, network and molecular dynamics modeling, and loss-of-function testing; 3) deploy advanced genome engineering technologies to enable stacking of a large number of transgenes into a single genomic locus to improve transgene persistence and transfer fully functional 'carboxylation' and 'decarboxylation' modules from CAM species to C_3 species that can accommodate overnight malic acid storage in the vacuole; and 4) analyze the effects of these transgenic modules on 'stomatal control', CO_2 assimilation and transpiration rates, biomass yield, and WUE in *Arabidopsis* and *Populus*.

Increasing demands for food, feed, fiber, and fuels due to future human population growth, coupled with decreasing arable land area; predicted increasing severity and frequency of extreme weather conditions including higher temperatures and drought; and overreliance on groundwater for crop irrigation indicate the need for novel strategies to improve WUE of plants. One potential strategy is to move CAM into C_3 plants (Borland et al., 2014). CAM is a temporally controlled plant inorganic CCM that maximizes WUE by shifting all or part of the CO_2 uptake to the nighttime, when evapotranspiration rates are reduced compared with the daytime. CAM is distinguished by two major features: (i) 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, and (ii) an inverse stomatal behavior, in which stomata are closed during all or part of the day and are open at night. The organic acids accumulated overnight are subsequently decarboxylated during the day to release CO_2 and concentrate it around ribulose-1- 5-bisphosphate carboxylase/oxygenase (RUBISCO), favoring carboxylase activity and carbohydrate production *via* the C_3 Calvin–Benson cycle.

A fundamental requirement for engineered CAM is to define the minimal set of genes and proteins required for its efficient establishment and operation. Genomic sequences and transcriptome atlases have become available from cycling, facultative, or obligate CAM species sampled from diverse phylogenetic origins including monocot orchids (*Phalaenopsis),* pineapple (*Ananas comosus*), several *Agave* species (*A. Americana*, *A. deserti*, *A. sisalana*, and *A. tequilana*), and core eudicots including the common ice plant (*Mesembryanthemum crystallinum*), *Kalanchoë fedtschenkoi*, *K. laxiflora*, *Sedum album,* and *Opuntia*

ficus-indica, a widely cultivated member of the cactus family. Comparative transcriptomic and genomic approaches are used to discern CAM gene function by comparing gene expression patterns of known CAM components across C_3, C_4, and CAM species. Co-expression network modeling incorporating transcriptional data, functional genomics annotation, and genetics information is also used to discover genes comprising functional CAM modules. In addition, loss-of-function studies of individual enzymes, regulatory proteins, or transcription factors are used to provide critical insights into the basic genetic requirements for CAM. The above information is then combined to guide the design and empirical testing of minimal functional modules for carboxylation and decarboxylation, malate influx and efflux into and out of the vacuole, stomatal control, and anatomical requirements for CAM. A set of genes with coordinate function, rather than gene-by-gene testing, will be used to accelerate the empirical testing process (DePaoli et al., 2014).

Initial CAM biodesign functional testing efforts will target the genetic model *Arabidopsis* owing to its rapid growth rate and ease of transformation. With regard to bioenergy feedstocks, fast-growing woody plants within the *Populus* genus, which are used extensively in the timber, pulp, and paper industries, and more recently as a bioenergy crop, will be targeted. CAM modules will be expressed under the control of circadian clock-controlled, drought-inducible promoters to ensure proper temporal expression of the CAM gene sets and promote maximal productivity. Resulting plants will be tested for transgene expression, biochemical signatures of CAM, CO_2 assimilation, stomatal conductance and transpiration rates, leaf carbon balance, level/mode of CAM activity, biomass productivity and quality, and integrated WUE. The effective transfer of CAM photosynthetic machinery into the important bioenergy crop *Populus* could significantly increase WUE for biofuels production in water-limited environments. If successful, the basic design principles outlined here can be extended to increase significantly the WUE of other bioenergy crops. Thus, this research is expected to have broad potential for ensuring sustainable biofuel feedstock production and for expanding production into semi-arid land areas.

References

Borland AM, Hartwell J, Weston D, Schlauch KA, Tschaplinski TJ, Tuskan GA, Yang X, Cushman JC. (2014) Engineering crassulacean acid metabolism to improve water-use efficiency. Trends Plant Sci. *In press.*

DePaoli HC, Borland AM, Tuskan GA, Cushman JC, Yang X. (2014) Synthetic biology as it relates to CAM photosynthesis: Challenges and opportunities. J. Exp. Bot. *In press.*

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–AC05–00OR22725.

114. Comparative Genomics in Support of CAM Engineering

Xiaohan Yang[1]* (yangx@ornl.gov), Hengfu Yin[1], Susanna F. Boxall[2], James Hartwell[2], David Weston[1], Paul E Abraham[1], Timothy J. Tschaplinski[1], Sara Jawdy[1], Priya Ranjan[1], Gerald A. Tuskan[1], Robert L. Hettich[1], Stan D. Wullschleger[1], Bernard Wone[3], Won Cheol Yim[3], Jungmin Ha[3], **John C. Cushman**[3], Anne M. Borland[1,4], Karen Schlauch[3], Rachel Brenchley[2], Richard Gregory[2], Neil Hall[2], Tanja Magoc[5], Steven L Salzberg[6], Stephen M. Gross[7], Axel Visel[7], Zhong Wang[7], June Simpson[8]

[1]Oak Ridge National Laboratory, Oak Ridge, TN; [2]University of Liverpool, Liverpool; [3]University of Nevada-Reno, Reno, NV; [4]Newcastle University, Newcastle upon Tyne; [5]University of Maryland, College Park, MD; [6]Johh Hopkins University School of Medicine, Baltimore, MD; [7]Lawrence Berkeley National Laboratory, Berkeley, CA; [8]CINVESTAV, Irapuato, Guanajuato.

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). To achieve this long-term goal, four major objectives are being pursued: 1) define the genetic basis of key CAM modules; 2) characterize the regulation of 'carboxylation', 'decarboxylation', and 'inverse stomatal control' modules of CAM; 3) deploy advanced genome engineering technologies to enable stacking of a large number of transgenes into a single genomic locus to transfer fully functional 'carboxylation' and 'decarboxylation' modules from CAM species to C_3 species that can accommodate overnight malic acid storage in the vacuole; and 4) analyze the effects of these transgenic modules on 'stomatal control', CO_2 assimilation and transpiration rates, biomass yield, and WUE in *Arabidopsis* and *Populus*. CAM biodesign research is multidisciplinary by design and multi-institutional in composition.

Improved crop water-use efficiency (WUE) is critical for the long-term sustainability of agricultural production systems in the face of predicted future warmer and drier climates. Crassulacean acid metabolism (CAM) is a specialized mode of photosynthesis that enhances WUE through an inverse day/night pattern of stomatal closure/opening and improves photosynthetic efficiency by concentrating CO_2 around RUBISCO. CAM has evolved multiple times from C_3 photosynthesis and ~6.5% of higher plant species in more than 35 families have acquired CAM via parallel or convergent evolution. There are three fundamental questions to be answered to understand the molecular basis and evolutionary mechanism of CAM: 1) how many genes are needed for the C_3-to-CAM transition? 2) what are these genes? and 3) when did they evolve? To address these questions, we used an approach integrating co-expression gene network and protein sequence evolution, with a focus on the difference between CAM and non-CAM species as well as the conservation among different CAM species. Our network analysis based on *Agave* RNA- seq data identified co-expression gene modules associated with the CAM pathway. Through comparative genomics analysis of 15 species selected from diverse lineages, including CAM (e.g., *Agave*), C_3 (e.g., *Arabidopsis*, *Oryza*, *Populus*), C_4 (e.g., *Setaria*, *Sorghum*, *Zea*), and non-vascular plants (e.g., *Physcomitrella*, *Selaginella*), we found that the core components of CAM machinery have an ancient origin traceable to non-vascular plant lineages, and regulatory proteins were essential to the C_3-to-C_4 transition. Furthermore, our comparison of multiple CAM plants (i.e., *Agave*, *Kalanchoe*, *Mesembryanthemum*) revealed that parallel or

convergent evolution resulted in regulatory machinery that transitioned from the C_3 to CAM pathways. This research establishes a framework for CAM comparative genomics studies and provides new knowledge to inform genetic improvement in WUE and photosynthetic efficiency in crop plants under water-limiting conditions using synthetic biology approaches.

***Acknowledgement**: This research was 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 U.S. Department of Energy under Contract Number DE-AC05-00OR22725.*

115. Engineering Robust Hosts for Microbial Biofuel Production

Tim Tomko[1], Will Turner[1], Jessica Lindle[1], **Mary J. Dunlop*** (mjdunlop@uvm.edu) [1]

University of Vermont, School of Engineering, Burlington, VT 05405

Project Goals: The goal of this project is to enhance microbial synthesis of next-generation biofuels by developing tools for improving microbial tolerance of biofuel production conditions. Research is organized around three objectives: (1) Identify novel biofuel tolerance mechanisms from microorganisms that naturally thrive in hydrocarbon-rich environments. (2) Engineer a synthetic feedback loop that responds to biofuel production. To optimize biofuel production yields, cells must balance several competing sources of stress. We are designing and constructing a novel feedback loop that senses biofuel production and turns on export pumps in response. (3) Integrate multiple tolerance strategies in a biofuel production strain. In addition to having the potential to greatly enhance biofuel yields, this work will advance understanding of how multiple tolerance mechanisms interact within a cell.

Abstract:
Microbial biofuel synthesis is a cost effective and environmentally sustainable way of producing replacements for gasoline, diesel, and jet fuel from lignocellulosic biomass. In a typical production process, biomass is deconstructed into sugars that are metabolized by a microbe engineered to convert sugar into biofuel. In this project, we develop engineering tools for increasing the robustness of a biofuel production host.

A major challenge when using microorganisms to produce bulk chemicals like biofuels is that the production targets are often toxic to cells[1]. Biofuel-like compounds are known to reduce cell viability through damage to the cell membrane and interference with essential physiological processes[1,2]. Thus, cells must trade off biofuel production and survival, reducing potential yields. In addition, residual chemicals that remain after biomass pretreatment can inhibit cell growth, further reducing yields. Recent studies have shown that strains engineered to increase tolerance can improve biofuel production yields[1].

Recent work by the PI has indicated that microorganisms that survive in oil-rich environments are a valuable source of tolerance mechanisms[3]. We are using a transgenic screening approach, building libraries from the genomic DNA of microbes that have been isolated from environments near natural oil seeps and in the vicinity of oil spills, and screening for biofuel tolerance in *Escherichia coli*. We are focusing specifically on libraries from *Pseudomonas* species and the alkane-metabolizing microbe *Alkanivorax borkumensis*.

Expression of transport proteins called efflux pumps can increase tolerance to biofuels and pretreatment chemicals by pumping toxins out of the cell and improving fuel yields. However, overexpression of efflux pumps can compromise the cell, creating a trade-off between chemical toxicity and pump toxicity[4,5]. Research has suggested that certain combinations of efflux pumps can confer additional tolerance compared to the individual pumps themselves. But, the functional form of the combination of the tolerance provided by each pump and the toxicity due to their simultaneous activity is unknown. We approach this problem using a combination of mathematical modeling and experiments. Using differential equations, we developed a growth model incorporating the trade-offs between toxicity of inhibitors and tolerance provided by efflux pumps to describe the dynamics of bacterial growth under these conditions. By analyzing each inhibitor and efflux pump independently

through a series of biological experiments and mathematical models, we determine the functional form of their combined effect on growth rate, with the long-term goal of increasing biofuel yields. Modeling predictions are tested experimentally to quantify the combined effect of expressing multiple tolerance mechanisms.

Funding Statement
This research was supported by the University of Vermont and the Office of Science (Biological and Environmental Research), US Department of Energy.

References

1. Dunlop, M.J. "Engineering Microbes for Tolerance to Next-Generation Biofuels." *Biotechnology for Biofuels* 4:32, 2011.
2. Nicolaou, S.A., S.M. Gaida, E.T. Papoutsakis "A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation." *Metabolic Engineering*, 12(4), 2010.
3. Dunlop, M.J., Z.Y. Dossani, H.L. Szmidt, H.C. Chu, T.S. Lee, J.D. Keasling, M. Hadi, A. Mukhopadhyay. "Engineering Microbial Biofuel Tolerance and Export Using Efflux Pumps." *Molecular Systems Biology* 7:487, 2011.
4. Dunlop, M.J., J.D. Keasling, and A. Mukhopadhyay, *A model for improving microbial biofuel production using a synthetic feedback loop.* Systems & Synthetic Biology, 2010. **4**(2): p. 95-104.
5. Harrison, M.E., M.J. Dunlop. "Synthetic Feedback Loop Model for Increasing Microbial Biofuel Production Using a Biosensor." *Frontiers in Microbiology* 3:360, 2012.

116. Systems Approaches for Engineering Cyanobacteria for Biofuel Production

Matthew Long[1]* (mrlong2@wisc.edu) and **Jennifer L. Reed**[1]

[1]Dept. of Chemical and Biological Engineering, Univ. of Wisconsin-Madison

Project Goals: Cyanobacteria offer a promising route for directly converting solar energy and CO_2 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 *S.* 7002. Experiments will subsequently be performed to construct and analyze *Synechococcus* 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 approaches cannot predict the effects that these changes will have on other parts of metabolism, and generally will not suggest alterations to more distant pathways.

Computational models of cellular metabolic and regulatory networks can be used to guide and accelerate these metabolic engineering efforts by integrating and analyzing experimental data, and identifying genetic manipulations that would increase product yields. In the process of developing metabolically engineered strains, genetic manipulations proposed by computational strain design algorithms depend on the metabolic state of parental strains. One such algorithm is RELATCH (for relative change), which has been shown to accurately predict the effects of gene deletions and environmental shifts on metabolic fluxes [1]. However, RELATCH requires knowledge of both fluxes and gene expression in the parental strain (e.g., from ^{13}C metabolic flux analysis) to predict fluxes in knockout mutants. While gene expression is easily measured, intracellular flux measurements are harder to generate and are not widely available, particularly for cyanobacteria. As such, alternative methods for obtaining knowledge of fluxes through metabolism are needed to evaluate and improve engineered strains.

A number of experimental measurements can be made to evaluate the metabolic state of a cell, such as enzyme activity, gene expression, metabolite concentrations, protein concentrations, and cellular uptake and secretion rates. The integrated analysis of these various datasets can be used to help estimate metabolic fluxes in cells and identify potential bottlenecks in biofuel production. Here we have developed a novel constraint-based modeling method for calculating the flux distribution in a

parental strain using experimental data from multiple gene deletion strains. By using growth rates, extracellular fluxes, and gene expression data from multiple knockout strains, we are able to accurately estimate the parental strain flux distribution, and thus use RELATCH to predict fluxes in new mutant strains with greater accuracy. Future work will be to apply these methods to identify and alter fluxes in cyanobacterial strains engineered for biofuel production.

References:

1. Kim, J. and J.L. Reed, RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations. Genome Biol, 2012. **13**(9): p. R78.

This work was supported by the Office of Science (BER), U.S. Department of Energy (DE-SC008103).

117. Application of Next-Generation Sequencing to Engineering mRNA Turnover in Cyanobacteria

Jeff C. Cameron[1], Gina C. Gordon[1,2], Erin Zess[1], Ryan Clarke[1], Matt B. Begemann[1,2], Andrew L. Markley[1], **Brian F. Pfleger[1,2]**

[1]Department of Chemical and Biological Engineering, University of Wisconsin, Madison, WI
[2]Microbiology Doctoral Training Program, University of Wisconsin, Madison, WI

Project Goals: The goal of our project is to understand mRNA stability in cyanobacteria such that tools can be developed to predictably control gene expression in these exciting phototrophs.

Demand for sustainable transportation fuels and commodity chemicals has motivated biotechnologists to investigate production processes that start from sunlight and CO_2. Biotechnology has long been used as a means to produce chemicals but the range of targets has dramatically increased since the advent of recombinant DNA technology. Modern biotechnology, and specifically synthetic biology, has advanced to the point where metabolic pathways can be designed from scratch and integrated into the genome of hosts with advantageous inherent traits. Furthermore, the dramatic increase in DNA sequencing and synthesis capabilities over the last decade has led to an explosion of systems biology and synthetic biology methods for analyzing cells and building complex regulatory circuits. The remaining fundamental challenge in assembling novel metabolic pathways is designing DNA sequences de novo that encode the necessary structural and regulatory components for optimizing pathway function.

Messenger RNA are key, labile intermediates in the path to synthesizing proteins inside cells. Among the many modes of regulating gene expression, mRNA turnover is the least understood process, particularly with respect to individual genes and sequence elements. Algorithms have been developed for identifying promoters, and predicting transcription and translation rates from genomic sequence, but a predictive model of mRNA turnover does not yet exist. This is in part because of the many modes by which mRNA molecules can be degraded and the multitude of RNA processing enzymes inside cells. Sequence and/or structural elements that can slow or accelerate mRNA processing by specific enzymes have been identified but their ability to confer the same traits to other gene sequences has not been straightforward. Instead, combinatorial methods of altering mRNA stability have been developed and shown to be a powerful strategy for optimizing metabolic pathways when facile screens or selections are available. In order to design optimal gene expression cassettes, additional knowledge of how a particular mRNA sequence decays is needed. Past efforts using DNA microarrays provided data that examined rates of mRNA decay on a global scale but lacked the spatial resolution to determine which parts of a transcript were recycled first, second, or last. We hypothesize that next-generation RNA sequencing protocols will provide this level of resolution and enable us to examine how each mRNA in a bacterium is turned over.

In this talk, we will describe our efforts to develop a genetic engineering toolbox for *Synechococcus* sp. PCC7002 and outline our plans to use RNA-sequencing techniques to study mRNA stability in this industrially promising cyanobacteria. Specifically, we will discuss a novel counterselection method (1) that has greatly facilitated segregation and enabled scarless mutagenesis in PCC 7002. We will also discuss transcriptional and translational control elements that have been characterized in the course of our NSF- sponsored "Cyanobacteria photobiorefineries" project.

References

1. Begemann MB, Zess EK, Walters EM, Schmitt EF, Markley AL, Pfleger BF. An Organic Acid Based Counter Selection System for Cyanobacteria. PLoS One. 2013 Oct 1;8(10):e76594.

Funding statement

The work presented in this talk was funded by DOE-BER through an Early Career award (DE-SC0010329) and by the National Science Foundation (EFRI-1240268, GEO-1215871).

118. Isotopically nonstationary ^{13}C flux analysis of isobutyraldehyde production in *Synechococcus elongatus*

Lara J. Jazmin[1], Yao Xu[2], Adeola Adebiyi[1], Carl H. Johnson[2,3], **Jamey D. Young**[1,3]*
(j.d.young@vanderbilt.edu)

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

Project Goals: This project aims to apply ^{13}C flux analysis to identify strategies that will increase photosynthetic biofuel production in engineered cyanobacteria.

Recent studies have demonstrated the feasibility of converting energy from sunlight and carbon from CO_2 directly into biofuels using photosynthetic microorganisms. Despite the advances made in cyanobacterial biofuels production, the productivities achieved by cyanobacterial fermentations are currently too low for industrial feasibility and few tools are available that specifically address the challenges of redirecting and enhancing metabolic flux in photosynthetic microbes.

Our group is developing novel approaches that use isotope tracers and metabolic flux analysis (MFA) to quantitatively assess *in vivo* metabolic phenotypes of photoautotrophic hosts. Although ^{13}C is the preferred isotope tracer for mapping central carbon metabolism in heterotrophic hosts, photoautotrophs assimilate carbon solely from CO_2 and therefore produce a uniform steady-state ^{13}C-labeling pattern that is insensitive to fluxes. However, transient measurements of isotope incorporation following a step change from unlabeled to labeled CO_2 can be used to map photoautotrophic fluxes by applying newly developed techniques of isotopically nonstationary MFA (INST-MFA). We have recently developed a novel software package called INCA to facilitate model generation and computational solution of INST-MFA models, which is now publicly available to the scientific community [1]. We have also established experimental protocols for performing $^{13}CO_2$ labeling experiments and mass isotopomer analysis that are required for INST-MFA of autotrophic hosts [2, 3].

To establish proof-of-concept, we first applied ^{13}C INST-MFA to map fluxes in the model cyanobacterium *Synechocystis* sp. PCC 6803 growing under photoautotrophic conditions [4]. Comparison of the INST-MFA flux map to theoretical values predicted by a linear programming model revealed inefficiencies in photosynthesis due to oxidative pentose phosphate pathway and malic enzyme activity. Our ongoing work involves extending the ^{13}C INST-MFA approach to examine engineered strains of *Synechococcus elongatus* PCC 7942, with the goal of identifying novel genetic targets that control production of isobutyraldehyde (IBA, a direct precursor of isobutanol). Quantification of photosynthetic carbon fluxes in IBA-producing cyanobacteria is expected to pinpoint pathway bottlenecks that can be subsequently removed in further rounds of metabolic engineering, thus leading to maximal productivity by redirecting flux into biofuel- producing pathways.

References

1. Young, J.D., INCA: A computational platform for isotopically nonstationary metabolic flux analysis. Bioinformatics, 2014. **in press**.
2. Jazmin, L.J., et al., Isotopically nonstationary MFA (INST-MFA) of autotrophic metabolism. Methods Mol Biol, 2014. **1090**: p. 181-210.

3. Young, J.D., D.K. Allen, and J.A. Morgan, Isotopomer measurement techniques in metabolic flux analysis II: mass spectrometry. Methods Mol Biol, 2014. **1083**: p. 85-108.
4. Young, J.D., et al., Mapping photoautotrophic metabolism with isotopically nonstationary ^{13}C flux analysis. Metab Eng, 2011. **13**(6): p. 656-65.

This work was supported by DOE DE-SC0008118. LJJ was supported by DOE SCGF DE-AC05-06OR23100.

119. Analysis of Three Ruminal-associated Bacteria Reveal Different Mechanisms for Cellulose Degradation

Melissa R. Christopherson[1], Anthony P. Neumann[1], Caryn S. Wadler[1], Meagan C. Burnet[2], John Dawson[3], Andrew Cunningham[1], David M. Stevenson[1,4], Shanti Bramhacharya[1], Christina Kenziorski[3], Paul J. Weimer[1,4], Steven J. Callister[2], and **Garret Suen[1,*] (gsuen@wisc.edu)**

[1]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI; [2]Department of Energy Pacific Northwest National Laboratory, Richland, WA; [3]Department of Biosatistics, University of Wisconsin-Madison, Madison, WI; [4]US Dairy Forage Research Center, USDA-ARS, Madison, WI

Project Goals: The goals of this project are to uncover the mechanism through which three ruminal associated bacteria degrade cellulose. These include *Cellulomonas gilvus, Fibrobacter succinogenes,* and *Ruminococcus albus,* all of which cellulose degradation strategies different from canonical approaches such as cellulosomes that are used by other well-understood microbes. We use a combination of cellulose-degradation assays, whole-genome sequencing, RNA-seq, and proteomics to demonstrate their ability to degrade cellulose using entirely different strategies. These data are important for advancing our understanding of cellulose degradation in natural systems and will have impact for enhancing the production of bioenergy products such as ethanol. Given the importance of these bacteria in their natural system (i.e. in ruminants such as cows), understanding their biology will also aid in efforts to improve digestion and optimize animal production.

The ability of ruminants such as cows to degrade and utilize plant biomass is predicated on the symbiotic relationship they have with their gut microbes. Through human domestication efforts, the ruminal microbial ecosystem has been selected for optimal fermentation of plant polysaccharides. This digestive process involves two steps: the degradation of plant polysaccharides such as cellulose and hemicelluloses into simple sugars, followed by their fermentation into short chain fatty acids like acetate, butyrate, and propionate. This model can thus be useful for informing on our ability to produce biofuels as many of the microbes may possess novel enzymes of interest or serve as a chassis for the direct fermentation of produces from plant polysaccharides. We have been studying the cellulolytic and fermentative properties of three ruminal-associated microbes: *Cellulomonas gilvus*, *Fibrobacter succinogenes*, and *Ruminococcus albus.* Our first step was to sequence their genomes and determine their cellulolytic potential. Importantly, while all bacteria have cellulases and other carbohydrate-degrading genes, none possessed scaffoldins and dockerins characteristic of cellulosome-utilizing bacteria. This suggested that all three bacteria do not use cellulosomes and must use other strategies to degrade cellulose. We tested this for each bacterium using a combination of cellulase assays, fermentation assays, RNA-seq, and proteomics.

Cellulomonas gilvus ATCC13127. This bacterium is unique in that it is the only reported bacterium capable of degrading cellulose both aerobically and anaerobically. Analysis of this bacterium under both conditions revealed that it is capable of fermenting small amounts of ethanol. Since all of the predicted cellulases in its genome have signal peptides, we posited that it excretes its cellulases when degrading cellulose under both oxic and anoxic conditions. An analysis of the proteins from spent media confirmed this hypothesis and further revealed a specific set of cellulases excreted based on condition. Our preliminary RNA-seq data also confirm these findings, suggesting that this bacterium may utilize an aerobic cellulose-degrading strategy regardless of oxygen availability.

Fibrobacter succinogenes S85. As one of only two species that define an entire phylum, *F. succinogenes* is a unique cellulose-degrading bacterium best known for its presence in ruminants and the hindgut of termites. Like other strictly anaerobic cellulose degraders, it requires attachment to cellulose fibers for degradation. An analysis of its genome shows atypical cellulases, as none contain carbohydrate-binding modules in families 1, 2, or 3, which are associated with strict anaerobes. Moreover, while this bacterium contains genes for the degradation of hemicelluloses like xylan, it cannot utilize the degraded byproducts. To gain a better understanding of its cellulose degrading mechanism, we conducted RNA-seq experiments on *F. succinogenes* grown in a chemostat on either cellulose or cellobiose. We found a number of highly expressed cellulases and a unique set of Fibro-slime proteins that have been implicated in cellulose-binding. A second set of Fibro-slime proteins were expressed under both conditions, suggesting that these may play other roles in its physiology. These data were further confirmed by proteomics of both spent media and cell pellets. These data indicate that this bacterium utilizes a cell-surface attached approach to degrade cellulose that is different from other anaerobic cellulose degrading bacteria.

Ruminococcus albus 7. This entire family of bacterium is well-known for its cellulolytic abilities, particularly within the rumen ecosystem. *Ruminococcus albus* 7 is no exception in that it is highly cellulolytic with the ability of producing ethanol *in vitro*. An analysis of its genome shows that it does not contain the cellulosome machinery characteristic of other species in this genus like scaffoldins and dockerins. We performed an RNA-seq analysis of its cellulolytic capabilities in a chemostat on both cellulose and cellobiose. Our analysis revealed a number of highly-expressed cellulases along with many CBM37s, which are specific to this bacterium and may be involved in cell-surface attachment to cellulose fibers. Surprisingly, we found that the tryptophan operon was the most highly expressed set of genes when grown on cellulose. This is similar to what has been reported in the proteome of *Clostridium phytofermentens*, a related bacterium that also does not utilize a cellulosome. These data suggest that this bacterium thus utilizes a completely novel cellulose degradation approach likely involving CBM37s that may be embedded in a glycocalyx-like structure on its cell membrane which it uses to attach to cellulose fibers.

Taken together, these analyses are beginning to shed light on our understanding of cellulose degradation in nature. Moreover, this diversity of approaches suggests that the use of cellulosomes is not the necessarily widespread and does not represent the full extent through which microbes degrade cellulose. Finally, given that these microbes are fermenters, capable of generating ethanol in many cases, they may be useful as models for understanding the fermentation of cellulose to ethanol.

This work was funded by the US Department of Energy's (DOE) Office of Biological and Environmental Research (OBER) Early Career Research Program Award DE–SC0008104 to GS. Metaproteomics analysis were supported by the U.S. DOE OBER Pan-omics program at Pacific Northwest National Laboratory (PNNL) and performed in the Environmental Molecular Sciences Laboratory, a U.S. DOE OBER national scientific user facility on the PNNL campus. PNNL is a multiprogram national laboratory operated by Battelle for the DOE under contract DEAC05-76RL01830.

120. Analysis of the Transcriptional Response of Synechococcus sp. PCC 7002 to Specific Growth Conditions from a Compendium of RNA-seq Data

Ryan S. McClure[1], Lye Meng Markille[1], Ronald C. Taylor[1], Jason E. McDermott[1], Marcus Ludwig[2], Donald A. Bryant[2], **Allan Konopka[1]**, and **Alex S. Beliaev[1]**

[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352
[2]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, State College, PA 16802

Project Goals: The regulatory machinery of cyanobacteria, which evolved to provide ecophysiological advantages, plays a major role in both short- and long-term adaptations to environmental perturbations. The challenges of extreme environments require coordinated adjustment of photosynthetic efficiency, carbon processing rates, and "downstream" carbon partitioning between biomass and metabolite storage pools. Understanding how regulatory controls dynamically integrate the external inputs to produce intracellular adaptations will inform strategies to overcome productivity constraints and optimize metabolism for biofuels production. Availability of a well-established genetic system for *Synechococcus* sp. PCC 7002, in conjunction with high-throughput 'omics approaches, provides the scale and resolution for a comprehensive analysis of transcriptional regulation in cyanobacteria.

Synechococcus sp. PCC 7002 (hereafter *Synechococcus* 7002) is a fast-growing unicellular cyanobacterium that can be found in the brackish or saline water of tidal estuaries, a dynamic environment where strict gene regulatory mechanisms are necessary for survival. Such regulation takes place with fewer transcription factors than in proteobacterial systems, and in the presence of numerous antisense transcripts, suggesting post-transcriptional methods of regulation. High resolution quantitative analyses of RNA *via* next-generation sequencing offers the ability to analyze changes, both in specific gene products and the organism's transcriptome as a whole, when exposed to different stimuli. Such analyses represent an alternative to genome analysis alone when identifying regulons as co-expressed genes can be identified regardless of genomic context. Deep sequencing technologies also provide the ability to analyze changes in both known open reading frames (ORFs) as well as unannotated RNA transcripts such as regulatory small RNAs (sRNAs). To that end, widespread occurrence of non-coding antisense transcripts (asRNA) in the genomes of cyanobacteria suggests a prevalence of post-transcriptional regulation. As previously postulated, the regulation of antisense transcription is likely to be tailored to its mode of action, while the co-expression patterns between asRNAs and their targets might indicate the mechanism of action.

In this study, we combined RNA-seq analysis from 41 different growth conditions to determine how *Synechococcus* 7002 responds on a global level using cluster analysis and context likelihood of relatedness (CLR) approaches. Clusters of co-expressed genes were generated and, in conjunction with functional enrichment analysis, were used to generate an overview of regulatory and metabolomic pathways in *Synechococcus* 7002. As RNA-seq also allows for the determination of 5' untranslated regions (UTRs) these were compared within clusters of co- expressed genes to identify possibly homology indicating common transcription factors or sRNAs regulating such gene clusters. In addition, several hundred UTRs were examined to gain insight into the specific transcriptional regulatory mechanisms of *Synechococcus* 7002. Several UTRs were found to be altered based on growth conditions. Comparing N-limitation to C-limited conditions showed UTR changes of at least 30 nucleotides in 45 genes and similar changes were found when comparing other growth conditions. Our analysis was not

limited to protein coding genes as we also identified over 450 instances of unannotated transcription either within an intergenic region or opposite a known protein-coding gene. Through a stepwise application of known sRNA characteristics we describe 24 sRNAs that are between 30-250 nucleotides in length and contain a predicted Rho Independent Terminator (RIT). Expression of several of these sRNAs is regulated based on specific environmental conditions and through an analysis of homologous base-pairing and concordant expression patterns determined from RNA-seq data we propose several mRNA targets of these sRNAs. For example, a sRNA at 1152089-1152190 shows a 2 fold increase in expression when cells are grown at 0.D. 0.4 compared to O.D. 0.1.

The sequence of this sRNA displays homologous basepairing to SYNPCC7002_A0916, a hypothetical protein that shows a 2 fold *decrease* in expression under these conditions. Such putative interactions between RNA strands suggests that sRNA may be a negative regulator of SYNPCC7002_A0916. These studies are the first to combine a large number of growth conditions to gain a global perspective on the transcriptome of *Synechococcus* 7002 as it responds to varying environmental stimuli. Knowledge gained regarding regulatory patterns of this marine organism will be invaluable both from an environmental perspective as well as in the context of bioenergy and biotechnology applications using cyanobacteria.

121. Characterization of Redox Signaling Pathways in Cyanobacteria during Nutrient Limitation

Aaron T. Wright* (email: aaron.wright@pnnl.gov), Natalie C. Sadler, Eric A. Hill, Matthew R. Melnicki, William B. Chrisler, Charles Ansong, Erika M. Zink, Richard D. Smith, **Allan E. Konopka,** and **Alexander S. Beliaev**

Pacific Northwest National Laboratory, Biological Sciences Division, Richland, Washington, 99352

Project Goals: The goal of the PNNL BSFA is to determine the natural design principles of microbial photoautotrophic systems involved in solar energy conversion to biofuels. Significant knowledge gaps exist in our understanding of regulatory events within and between metabolic subsystems that include intracellular signals to which transcriptional regulation is responding. We suggest that "redox sensing," as a means to maintain redox homeostasis in photosynthesizing cells, is likely an equally important mechanism. To investigate this hypothesis, we have developed and applied a chemical probe approach for live- cell capture, characterization, and imaging of proteins that undergo dynamic fluctuation in redox status during environmental changes. It is anticipated that the identification of redox-sensitive dithiol linkages and their modulation by thioredoxin (Trx), peroxiredoxin, and other regulators will provide key inputs for understanding the control points of flux distribution.

The primary challenge in identifying redox-regulated dithiol sensors *in vivo* is that cysteine residues are easily oxidized to dithiol linkages following cellular lysis. This eliminates the ability to experimentally measure redox regulation within native physiological settings, and therefore requires the exogenous addition of chemical or biological reductants. To overcome the dilemma lysis presents, we developed an approach for employing cell permeable probe reagents that react in live cells (*in vivo*) with thiols and permit real-time imaging and mass spectrometric characterization of probe targets. For this approach, we synthesized click-chemistry enabled chemical probes for fluorescent and mass spectrometric identification of redox regulated dithiol sensors. We have used our probes *in vivo* in both *Synechococcus* sp. PCC 7002 and *Cyanothece* sp. ATCC 5144 to identify proteins that undergo disulfide exchange in response to changes in cellular conditions. Probes were synthesized with three chemical elements: a moiety to impart cell permeability, an iodoacetamide or maleimide group for irreversibly labeling cysteines, and a reporter tag for detection and isolation of probe-labeled proteins. We exploited the multimodal bio-compatible click chemistry (CC) reaction to create "tag-free" probes for profiling proteins in living systems. Probe-labeled proteins were visualized by addition of a complementary azido-tetramethylrhodamine for fluorescent SDS-PAGE, Alexafluor-488 for confocal microscopy, or biotin-TEV-azide tag for enrichment and mass spectrometric analysis (LC-MS).

The maleimido and iodoacetamide probes were added simultaneously *in vitro* or *in vivo* to *Synechococcus* sp. PCC 7002 grown in a turbidostat under maximal growth rate conditions. Cells were also removed and placed in the dark for three hours and then labeled *in vitro* or *in vivo*. Following probe-labeling cells were lysed, and probe-labeled proteins were attached to biotin azide via CC. The probe-labeled proteins were then enriched on streptavidin, digested with trypsin, and the peptides analyzed by LC-MS. Critically, in the cells that were labeled post-lysis (*in vitro*) we found no changes in redox regulation of dithiol sensors, demonstrating that lysis rapidly oxidizes biological samples. However, when we analyzed the *in vivo* labeled cells, we identified redox-regulated proteins that were statistically different between the light and dark conditions. In a follow-up study using a carbon-limited chemostat we identified redox changes *in vivo* within 30 seconds

following the addition of CO_2. A time-course study revealed remarkable changes from 30 seconds to 60 minutes post CO_2 addition. The global proteome analysis informed us that none of the proteins abundance changed over the 60 minute time-course. Lastly, we identified the cysteine sites of probe labeling for a majority of the proteins undergoing dynamic redox fluctuations; a finding that allows for a more targeted approach in determining actual protein thiol function.

Similar approach was used to elucidate the protein redox dynamics during H_2 production by the diazotrophic *Cyanothece* sp. ATCC 5144, a cyanobacterium that can evolve H_2 using the N_2-ase pathway. Remarkably, the observed fluctuations of the protein redox status correlated with H_2 evolution dynamics. At the same time, the measured bulk ROS dissipated, thus strengthening the hypothesis that H_2 production is a mechanism, by which *Cyanothece* 51142 alleviates the detrimental effects of ROS on PSII and photosynthetic rates. Importantly, the "redox" changes cannot be measured by traditional transcriptomic or proteomic measurements because these post translational redox events occur within a time-frame that protein and mRNA content has not changed. Probe-identified proteins from both probe types map well onto multi-subunit supra- molecular complexes involving photosynthetic pathways associated with efficient collection of excitation energy (light harvesting), electron transfer reactions linked to formation of electrochemical gradients, carbon dioxide sequestration (dark reactions), and ATP synthesis. Additional redox-dependent pathways include those involving chaperone activity, transcriptional regulation, and antioxidant proteins linked to protein repair. Together, these results provide quantitative information regarding redox- dependent switches associated with photosynthetic regulation, and provide a systems biology tool capable to providing high-throughput information necessary for predictive metabolic modeling. Finally, cell permeable probe approaches represent the only existing methods for identifying and imaging live cell redox regulation, and they will be critical to informing the predictive models of metabolism needed for bioenergy applications.

122. Development of Genome-Scale Dynamic Modeling Framework for Simulating Photoautotrophic Growth of Cyanobacteria

Hyun-Seob Song*(email: hyunseob.song@pnnl.gov), Jim K. Fredrickson, **Allan E. Konopka, and Alexander S. Beliaev**

Pacific Northwest National Laboratory, Biological Sciences Division, Richland, Washington, 99352

Project Goals: The PNNL BSFA conducts fundamental research of microbial photoautotrophs with specific emphasis on photosynthetic energy conversion, reductant partitioning, and central carbon metabolism. As a model system, we utilize unicellular prokaryotic organism *Synechococcus* sp. PCC 7002, which exhibits one of the fastest growth rates known among cyanobacteria and which is also remarkably tolerant to high light intensities. Understanding the genetic and physiological bases of these properties could provide fundamental new insights that are broadly applicable to the optimization of other biological systems for biofuels development. To that end, we are interrogating fluxes through central metabolic pathways to define the major constraining factors (*i.e.* metabolic and regulatory controls) governing carbon partitioning through the metabolic subsystems of cyanobacteria that can be manipulated to increase productivity of specific molecules that are either precursors or fuel molecules themselves.

Systems-level analysis of cellular metabolism has a level of complexity that requires a simulation model to integrate critical information on key genetic and metabolic mechanisms governing fluxes and predict the most likely outcome of metabolic engineering. Flux balance analysis (FBA) has contributed to a number of important successes in metabolic engineering and has been used for genome-scale metabolic reconstructions. The aim of our work is to develop an accurate dynamic model as a tool for investigating the fundamental mechanism driving metabolic shift in prokaryotic photoautotrophs which will explicitly incorporate regulation of metabolism (and thus can use global transcriptomic or proteomic datasets to refine the model in a sophisticated fashion). From a practical perspective, industrial-scale processes have rarely operated at steady-state; the capacity to capture culture dynamics in time or after experimentally imposed perturbations will also give these simulation models practical benefits.

We have focused our efforts on the development of a framework that can facilitate the use of dynamic metabolic models for different organisms and applications. This framework features the capability to account for dynamic regulation thus allowing the accurate prediction of growth rate changes and flux distributions in response to environmental variations. Dynamic modeling approach is particularly useful not only for providing fundamental understanding of metabolic properties but also for exploring the rational strategies of improving the productivity of biofuels and chemicals potentially produced therefrom. However, development of a large- scale dynamic modeling framework faces several critical barriers, including: (*i*) the difficulty in estimating a large number of parameters, and (*ii*) the lack of efficient algorithms for the identification of relevant metabolic pathways from genome-scale networks. With regard to the first issue, our framework avoids the over-parameterization problem by accounting for dynamics of slowly reacting metabolites (such as extracellular metabolites) only, while neglecting fast metabolites (mostly intracellular). This quasi steady-state approximation results in the model with only a few parameters, which are readily identified from collected experimental data. In contrast to constraint-based approaches that consider only a single optimal pathway, our approach analyzes metabolism in terms of diverse pathways options (termed elementary modes). As addressed in (*ii*), extraction of elementary modes from genome-scale networks poses a challenge due to their combinatorial explosion in a complex, large-scale network. Therefore, we developed a novel algorithm based on a new optimization concept that enables selective sequential computation, i.e., one pathway at a time. The developed

algorithm performs alternate implementation of integer and linear programming, which led to fast and numerically stable computation.

As the proof-of-principle, our team has been developing a genome-scale network-based model of *Synechococcus* sp. PCC 7002 (hereafter, *Synechococcus* 7002with the goal of determining the metabolic characteristics of this organism under carbon, nitrogen and light- limiting conditions, and carrying out dynamic simulation of its metabolic behaviors. Using our previously developed framework, we incorporated gene and protein expression data into the genome-scale network of *Synechococcus* 7002 (*i*Syp708) available in the literature, and obtained an initial estimation of flux distributions under each of three growth conditions. This framework is based on the flux minimization principle; flux distribution is estimated by suppressing one flux over another according to the associated gene (or protein) expression level. Then, we applied the alternate integer and linear programming to *i*Syp708 to selectively extract elementary modes that are close to initially estimated flux distributions. The analysis of elementary modes obtained as such provides condition-specific metabolic characteristics of *Synechococcus* 7002 under different environmental stresses. Identification of kinetic parameters is in progress for the dynamic simulation of temporal metabolic shift of *Synechococcus* 7002 subject to variations of environmental conditions.

123. Defining Determinants and Dynamics of Cellulose Microfibril Biosynthesis, Assembly and Degradation

Jocelyn Rose[1*] (jr286@cornell.edu) Thomas Brutnell,[2] Seth DeBolt,[3] Detlef Smilgies,[1] Larry Walker,[1]

[1]Cornell University, Ithaca, New York; [2]Donald Danforth Plant Science Center, St. Louis, Missouri; [3]University of Kentucky

Project Goals: The goals of this multidisciplinary project are to: (1) Establish platforms through reverse and forward genetics to identify and manipulate candidate genes that influence cellulose microfibril synthesis and structure; (2) Characterize the effects of altered candidate gene expression on cellulose microfibril synthesis and structure, and develop a mechanistic model for microfibril crystallization; (3) Determine the consequences of altering microfibril architecture on digestibility and integrate this information with nano-scale observations of enzymatic hydrolysis.

The central paradigm for converting plant biomass into soluble sugars for subsequent conversion to transportation fuels involves the enzymatic depolymerization of lignocellulosic plant cell walls by microbial enzymes. Despite decades of intensive research, this is still a relatively inefficient process, due largely to the recalcitrance and enormous complexity of the substrate. A major obstacle is still insufficient understanding of the detailed structure and biosynthesis of major wall components, including cellulose. For example, although cellulose is generally depicted as rigid, insoluble, uniformly crystalline microfibrils that are resistant to enzymatic degradation, the *in vivo* structures of plant cellulose microfibrils are surprisingly complex.

Crystallinity is frequently disrupted, for example by dislocations and areas containing chain ends, resulting in “amorphous” disordered regions. Importantly, microfibril structure and the relative proportions of crystalline and non-crystalline disordered surface regions vary substantially and yet the molecular mechanisms by which plants regulate microfibril crystallinity, and other aspects of microfibril architecture, are still entirely unknown. This obviously has a profound effect on susceptibility to enzymatic hydrolysis and so this is a critical area of research in order to characterize and optimize cellulosic biomass degradation.

The entire field of cell wall assembly, as distinct from polysaccharide biosynthesis, and the degree to which they are coupled, are relatively unexplored, despite the great potential for major advances in addressing the hurdle of biomass recalcitrance. Our overarching hypothesis is that identification of the molecular machinery that determine microfibril polymerization, deposition and structure will allow the design of more effective degradative systems, and the generation of cellulosic materials with enhanced and predictable bioconversion characteristics.

We believe that the most effective way to address this long standing and highly complex question is to adopt a broad ‘systems approach’. Accordingly, we have assembled a multi- disciplinary collaborative team with collective expertise in plant biology and molecular genetics, polymer structure and chemistry, enzyme biochemistry and biochemical engineering. Our team is using a spectrum of cutting edge technologies, including plant functional genomics, chemical genetics, live cell imaging, advanced microscopy, high energy X-ray spectroscopy and nanotechnology, to study the molecular determinants of cellulose microfibril structure.

Specifically we are coupling with an analytical pipeline to characterize the effects of altering microfibril architecture on bioconversion potential, with the goal of generating predictive models to help guide the identification, development and implementation of new feedstocks. We are using *Arabidopsis thaliana* and *Brachypodium distachyon* as model dicotyledon and grass species, respectively.

This project is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

124. Next Generation Protein Interactomes for Plant Systems Biology and Biomass Feedstocks Research

Shelly A. Trigg[1,3]* (swanamaker@salk.edu), Andrew MacWilliams,[1] Joseph Nery,[1] Anna Bartlett,[1] Rosa Castanon,[1] Adeline Goubil,[1] Joseph Feeney,[1] Ronan O'Malley,[1,2] Mary Galli,[1] and **Joseph R. Ecker**[1,2,4]

[1]Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla CA; [2]Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla CA; [3]Biological Sciences Department, University of California, San Diego; and [4]Howard Hughes Medical Institute

Project Goals: Develop an *en masse* yeast two-hybrid screening system using next generation sequencing to rapidly identify all protein-protein interactions in biofuels feedstocks, and generate comprehensive interactome maps that represent the findings.

In order to keep up with global energy demands, it is imperitive we acquire more knowledge of biofuel feedstocks for improving their cultivation and energy yield. Knowledge of protein-protein interaction (PPI) networks that promote robust plant growth or that are perturbed by pathogens causing disease could progress strategies for improving cultivation. However, current technologies available for obtaining PPI data are insufficient and unrealistic for non-model organisms because of time, cost, and sensitivity constraints. Even the largest high quality PPI map for the model plant *Arabidopsis thaliana* (Arabidopsis Interactome 1 or AI-1), that we generated, contains only 2% of all potential interactions, and took upwards of 5 years and $8 million to finish[1]. To address this problem, we are developing a next-generation sequencing integrated yeast two-hybrid (Y2H) system that will greatly improve the rate at which PPI data can be obtained and will be applicable to virtually any cell from which RNA can be extracted.

Our system enables *en masse* pooling and massively paralleled sequencing for the identification of interacting proteins by exploiting Cre-lox recombination. Screening of Y2H plasmids containing mutant loxP sequences in a yeast strain expressing the reporter gene for Cre Recombinase has shown only interacting proteins can induce Cre-mediated recombination of plasmids. The newly formed double mutant loxP makes an irreversible linkage of each protein's corresponding coding sequence, and has allowed us to identify protein interactions using Illumina paired-end sequencing. Preliminary testing of a positive and random reference set (PRS/RRS) consisting of ~300 open reading frames (ORFs) in a 1 ORF vs 1 ORF screen has shown a sensitivity and reproducibility similar to our previous Y2H screen used to make AI-1. Our next generation sequencing and analysis pipeline has shown a 98% overlap with interactions detected by Sanger sequencing. *En masse* Y2H screening has been attempted on the PRS/RRS and also on a subset of ~4,500 Arabidopsis ORFs that we thoroughly assayed when creating AI-1. Since all expected interactions have not yet been identified because the inherently low percentage of informative DNA (< 0.003%), we are working to improve assay sensitivity by enriching for double-mutant-loxP-sequence-containing DNA fragments with an inverse PCR strategy and/or an in-solution RNA probe capture approach. To more stringently select for diploids containing only hybrid plasmid, we added an antibiotic resistance marker to the plasmids that becomes functional only when plasmids are Cre-recombined. Once we are confident that our assay is detecting all interacting pairs that occurred, we will move forward in testing ORF libraries from feedstocks by shotgun cloning into our Y2H plasmids and carrying out our *en masse* screening pipeline. The ability to rapidly construct large PPI networks will yield deeper insight into a variety of molecular processes and pathways that will potentially allow improvement of feedstock productivity and sustainability.

References

1. Arabidopsis Interactome Mapping Consortium, Science 333, 6042 (2011).

We gratefully acknowledge the U.S. DOE Office of Biological and Environmental Research in the U.S. DOE Office of Science for funding this project; DOE-DE-SC0007078.

125. Development of *Aspergillus niger* as a host for hyperproduction of thermophilic cellulases

Pallavi A. Phatale[1,2,] * (pphatale@lbl.gov), Jed Lynn [1,2], Jinxiang Zhang [1,3], Saori Campen [1,2], **Blake A. Simmons**[1,3], **John Gladden** [1,3], **Scott Baker** [1,2], **Jon Magnuson** [1,2]

[1] Joint Bio Energy Institute, Emeryville, CA 94608; [2] Pacific Northwest National Laboratory, Richland, WA 99352; [3] Sandia National Laboratory, Livermore, CA 94551-0969

https://confluence.jbei.org/display/JBEI/Fungal+Biotechnology

Enzymatic saccharification of plant biomass involves the depolymerization of recalcitrant cell wall structures composed primarily of cellulose, hemicellulose and lignin. In order to efficiently liberate fermentable sugars from the plant cell wall, the biomass must be pretreated to reduce recalcitrance using methods such as dilute acid, ammonia fiber expansion (AFEX), or the use of certain ionic liquids (ILs). Specific ILs have proven to be excellent pretreatment reagents, but unfortunately the common classes of IL's, such as 1-ethyl-3-methylimidazolium acetate [C_2mim][OAc], can inhibit downstream saccharification by commercially available cellulase enzyme mixtures. To overcome this problem, researchers at JBEI have leveraged an apparent correlation between IL-tolerance and thermostability of certain cellulase enzymes to develop an IL-tolerant cellulase cocktail, called JTherm. The JTherm cocktail is composed of two recombinant enzymes, a β-glucosidase (BG) and a cellobiohydrolase (CBH) derived from thermophilic bacteria, and a mixture of native thermophilic endoglucanases produced by a bacterial consortia. In order to produce these IL-tolerant enzymes in a more efficient manner, we are attempting to develop *Aspergillus niger* into a high-titer expression system. *A. niger* was selected as the target host because of its proven ability to secrete high-titers of extracellular protein, but the biological mechanisms at a fundamental level remain unknown. We have therefore sought to generate a fundamental understanding of heterologous fungal secretion pathways so that we can translate this information into strategies to engineer *A. niger* to produce high-titers of heterologous enzymes for further enzyme optimization and mixture development. To initiate this work, we transformed *A. niger* with several IL-tolerant enzymes, many of which expressed well and had similar properties compared to enzymes expressed in *E. coli*. We then used these strains to initiate a variety of "–omics" approaches that will enable us to deepen our understanding of enzyme secretion and develop enzyme hyperproduction strains: 1) utilize a forward genetics approach to generate hyperproduction mutants of heterologous IL-tolerant enzymes followed by high-throughput sequencing to identify the mutations; 2) sequence the genomes of existing industrial enzyme hyper-production strains to identify the genomic alterations responsible for their hyper-production phenotypes; 3) develop a reverse genetics platform to stack traits from hyper-production strains/mutants into a single production host; 4) utilize transcriptomic and proteomic analysis to determine the systemic response of *A. niger* to heterologous enzyme production to aid in strain engineering and to identify genetic "parts" that can be used to construct efficient expression constructs. The primary goal of this genome-scale analysis is to gain sufficient information to enable the development of bioengineering strategies that efficiently and selectively increase the expression of heterologous enzymes, in particular enzymes that are used to develop technologies for the conversion of lignocellulosic biomass to fungible advanced biofuels.

126. Biochemical and Structural Studies of Enzymes Involved in Fatty Acid Biosynthesis

Pouya Javidpour[1,2], Jose H. Pereira[1,2], Ee-Been Goh[1,2], Ryan P. McAndrew[1,2], Suzanne M. Ma[1,2,&], Gregory D. Friedland[1,3,#], Jay D. Keasling[1,2,4], Swapnil R. Chhabra[1,2], Paul D. Adams[1,2,4*] (PDAdams@lbl.gov) and Harry R. Beller[1,5]

[1]Joint BioEnergy Institute, Emeryville, California, USA; [2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA; [3]Biomass Science and Conversion Technology Department, Sandia National Laboratories, Livermore, California, USA; [4]Department of Bioengineering, University of California, Berkeley, Berkeley, California, USA; [5]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA; [&]Present address: Amyris, Inc., Emeryville, California, USA; [#]Present address: Nodality, Inc., South San Francisco, California, USA.

Project Goals: Direct replacement of petroleum-derived gasoline and diesel fuels may be possible through metabolic engineering of microbial fatty acid biosynthetic pathways. We have therefore structurally and biochemically characterized two key enzymes with the goal of aiding the production of fatty acid derived biofuels.

a) Micrococcus luteus is a Gram-positive bacterium that produces branched alkenes, which are potential biofuels, by the condensation of fatty-acid thioesters,. In an effort to better understand the control of the formation of branched fatty acids the structure of FabH was determined [1]. This enzyme catalyzes the initial step of fatty-acid biosynthesis, which is the condensation of malonyl-ACP with an acyl-CoA. Analysis of the structure provides insights into substrate selectivity with respect to length and branching of the acyl-CoA. The most structurally divergent region of FabH is a loop region located at the dimer interface, which is involved in the formation of the acyl- binding channel, limiting the substrate-channel size. A phenylalanine residue that is positioned near the catalytic triad appears to play a major role in branched-substrate selectivity. In addition to structural studies, transcriptional studies were also performed, focusing on an increase in the ratio of *anteiso*- to *iso*-branched alkenes observed at different stages of bacterial growth.

b) In general, anaerobic culture conditions would be preferred in the fermentative production of biofuels, to avoid maintaining dissolved oxygen levels and to maximize the proportion of reducing equivalents directed to biofuel. One concern with such fermentative growth conditions is increased NADH levels, which can adversely affect cell physiology. Bioinformatic and crystallographic analyses were used to identify four potential NADH-dependent variants of *Escherichia coli* FabG, an essential reductase involved in fatty acid biosynthesis [2]. Assays of cofactor preference in the variants showed up to a 35-fold preference for NADH (over NADPH), which was observed with the enzyme from *Cupriavidus taiwanensis*. In addition, use of these NADH-dependent variants in fatty acid- and methyl ketone-overproducing *E. coli* host strains under anaerobic conditions led to increased free fatty acid and methyl ketone titer relative to the controls. With further engineering, these variants could serve as a starting point for establishing a microbial host strain for production of fatty acid-derived biofuels under anaerobic conditions.

References

[1] Pereira JH, Goh EB, Keasling JD, Beller HR, Adams PD: Structure of FabH and Factors Affecting the Distribution of Branched Fatty Acids in *Micrococcus luteus*. *Acta Cryst*. 2012, **D68**:1320-1328
[2] Javidpour P, Pereira JH, Goh EB, McAndrew RP, Ma SM, Friedland GD, Keasling JD, Chhabra SR, Adams PD, Beller HR: Biochemical and Structural Studies of NADH- Dependent FabG Used to Increase the Bacterial Production of Fatty Acids under Anaerobic Conditions. *Appl Environ Microbiol*. 2014, **80**:497-505.

Funding

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.

127. SND1 Transcription Factor–Directed Quantitative Functional Hierarchical Genetic Regulatory Network in Wood Formation in Populus trichocarpa

Ying-Chung Lin[1]* (ylin14@ncsu.edu), Wei Li[1]* (wli24@ncsu.edu), Quanzi Li[14]* (qli2@ncsu.edu), Ying-Hsuan Sun[2], Sapna Kumari[3], Hairong Wei[3], Sermsawat Tunlaya-Anukit[3], Ronald R. Sederoff[1], **Vincent L. Chiang**[1].

[1]Forest Biotechnology Group, Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, North Carolina 27695
[2]Department of Forestry, National Chung Hsing University, Taichung 40227, Taiwan
[3]School of Forest Resources and Environmental Science, Michigan Technological University, Houghton, Michigan 49931
[4]College of Forestry, Shandong Agricultural University, Taian, Shandong 271018, China

Wood is an essential renewable raw material for industrial products and energy. However, knowledge of the genetic regulation of wood formation is limited. We developed a genome-wide high-throughput system for the discovery and validation of specific transcription factor (TF)– directed hierarchical gene regulatory networks (hGRNs) in wood formation. This system depends on a new robust procedure for isolation and transfection of Populus trichocarpa stem differentiating xylem protoplasts. We overexpressed Secondary Wall-Associated NAC Domain 1s (Ptr-SND1-B1), a TF gene affecting wood formation, in these protoplasts and identified differentially expressed genes by RNA sequencing. Direct Ptr-SND1-B1–DNA interactions were then inferred by integration of timecourse RNA sequencing data and top-down Graphical Gaussian Modeling–based algorithms. These Ptr-SND1-B1-DNA interactions were verified to function in differentiating xylem by anti-PtrSND1-B1 antibody-based chromatin immunoprecipitation (97% accuracy) and in stable transgenic P. trichocarpa (90% accuracy). In this way, we established a Ptr-SND1-B1–directed quantitative hGRN involving 76 direct targets, including eight TF and 61 enzyme-coding genes previously unidentified as targets. The network can be extended to the third layer from the second-layer TFs by computation or by overexpression of a second-layer TF to identify a new group of direct targets (third layer). This approach would allow the sequential establishment, one two-layered hGRN at a time, of all layers involved in a more comprehensive hGRN. Our approach may be particularly useful to study hGRNs in complex processes in plant species resistant to stable genetic transformation and where mutants are unavailable.

This work was supported by the Office of Science (Biological and Environmental Research), Department of Energy Grant DE-SC000691 (to V.L.C.).

128. Quantitative Site-Specific Redox Proteomics on Protein Thiols and Broad Light/Dark Modulation of Thiol Oxidation in Cyanobacteria

Jia Guo[1], Dian Su[1], Amelia Y. Nguyen[3], Matthew J. Gaffrey[1], Yi Qu[1], Ronald J. Moore[1], David W. Koppenaal[2], Mary S. Lipton[1], Richard D. Smith[1,2], Himadri B. Pakrasi[3]and **Wei-Jun Qian**[1*] **(**weijun.qian@pnnl.gov)

[1]Biological Sciences Division and [2]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, [3]Department of Biology, Washington University, St. Louis, MO

Project Goals: One of main objectives of this early career research project is to develop novel proteomic approaches that will enable quantitative measurements of site-specific regulatory protein posttranslational modifications (PTMs). The ability to effectively and quantitatively characterize site-specific PTMs is essential for understanding the regulation of cellular signaling and protein functions, as well as for enabling a systems biology approach to study organisms important for bioenergy or environmental applications. Our developments have been primarily focused on three important classes of PTMs: (1) reversible redox modifications on cysteinyl thiols, (2) proteolytic processing and protein N-terminal modifications,[1] and (3) glycosylation[2]. All three classes of modifications are ubiquitous in both prokaryotic and eukaryotic cells and their importance for cellular regulation and signaling have increasingly been recognized.

Quantitative redox proteomics: Functional cysteinyl residues in proteins serve as "redox switches" through reversible oxidation, which is recognized as a fundamental mechanism of redox regulation in almost all organisms. We have developed a novel quantitative redox proteomics approach for measuring different types of reversible modifications on individual cysteine thiols to study redox regulation in metabolism or stress conditions of different organisms.[3]

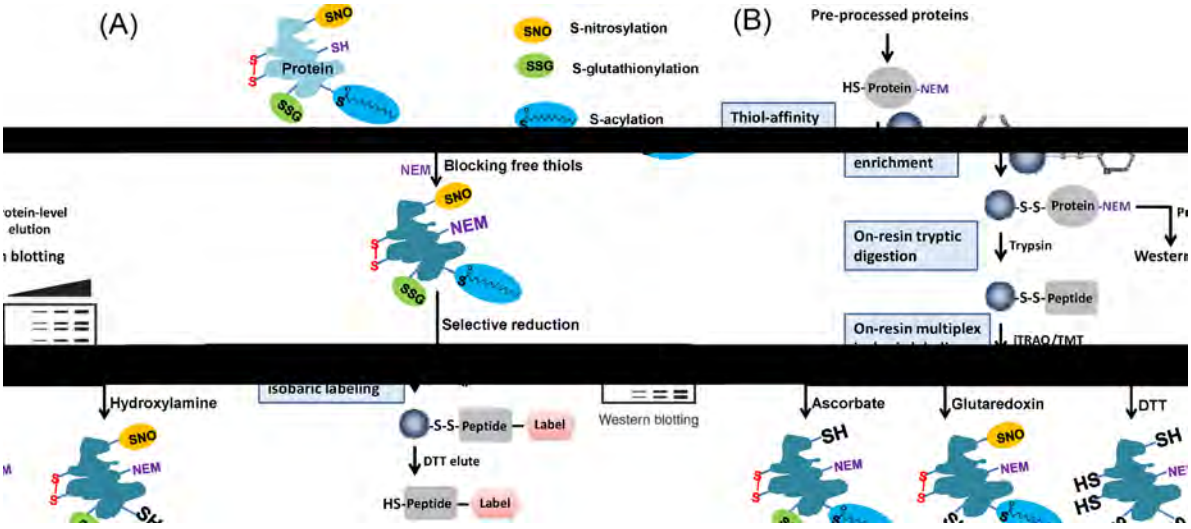


Figure 1. (A) Scheme of sample pre-processing for different reversible cysteine modifications. Different types of modifications are selectively reduced by different reagents. (B) Enrichment method for quantitative analysis of reversible cysteine modifications. (Guo, et al., Nat. Protoc. 2014)

Figure 1 illustrates the general principle of selective enrichment and quantification of site-specific

redox modifications. Briefly, thiol specific modifications can be reduced by specific reagents and the converted free thiols can be captured and enriched by a thiol-specific resin and their dynamics can be quantified by isobaric labeling and LC-MS/MS. We have applied this approach to profile SNO, SSG, and total thiol oxidations in multiple organisms.

Proteome-wide light/dark modulation of thiol-oxidation in cyanobacteria. By applying the redox proteomics approach to profiling the in vivo dynamics of thiol oxidation modulated by light/dark in Synechocystis sp. PCC 6803, an oxygenic photosynthetic prokaryote, we observed redox dynamics for ~2,200 cysteine sites from 1,060 proteins under different conditions (light, dark, and in the presence of a photosystem II inhibitor DCMU) (Fig. 2A and 2B). The results revealed broad proteome-wide changes in thiol oxidation in many key biological processes, including photosynthesis, carbon fixation, and glycolysis. Moreover, the redox sensitivity data enabled prediction of potential functional cysteine sites for proteins of interest (Fig. 2C).

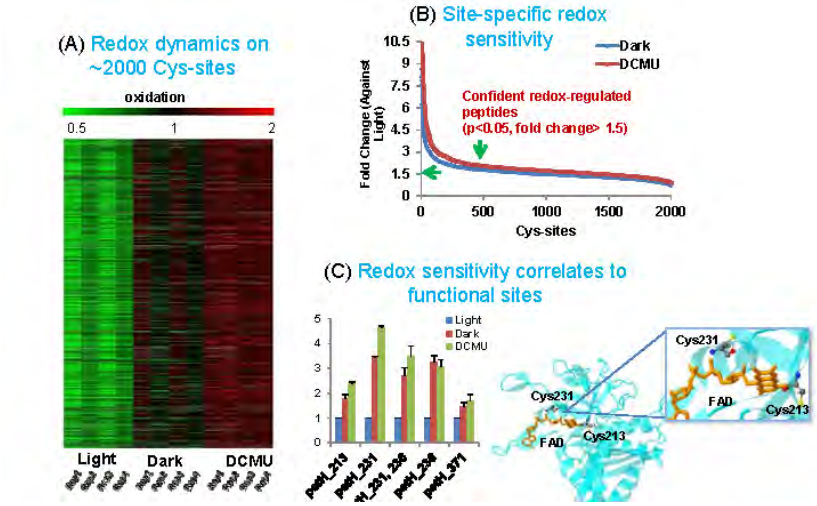


Figure 2. (A) Heatmap of the relative levels of oxidation of ~2000 identified Cys-sites under light, dark or DCMU conditions. (B) The redox sensitivity of individual Cys-sites. (C) The correlation of site-specific redox sensitivity with functional Cys sites.

Taken together, our results not only demonstrate the effectiveness of redox proteomics for profiling site-specific thiol modifications under physiological conditions, but also provide significant novel insights into the broad redox regulation of photosynthetic organisms.

This work is supported by the DOE Early Career Research Award under the Office of Biological and Environmental Research (BER) to WJQ and by the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences in the DOE Office of Science grant number DE-FG02-99ER20350 to HBP. This work uses capabilities developed under the BER-supported Pan-omics project at PNNL and in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the DOE. AYN has been supported by an NSF Graduate Research Fellowship.

References:

1. Kim, J. S.; Dai, Z.; Aryal, U. K.; Moore, R. J.; Camp, D. G., 2nd; Baker, S. E.; Smith, R. D.; Qian, W. J., Resin-assisted enrichment of N-terminal peptides for characterizing proteolytic processing. *Anal Chem* **2013,** *85* (14), 6826-32.
2. Wang, L.; Aryal, U. K.; Dai, Z.; Mason, A. C.; Monroe, M. E.; Tian, Z. X.; Zhou, J. Y.; Su, D.; Weitz, K. K.; Liu, T.; Camp, D. G., 2nd; Smith, R. D.; Baker, S. E.; Qian, W. J., Mapping N-linked glycosylation sites in the secretome and whole cells of Aspergillus niger using hydrazide chemistry and mass spectrometry. *J. Proteome Res.* **2012,** *11* (1), 143-56.
3. Guo, J.; Gaffrey, M. J.; Su, D.; Liu, T.; Camp, D. G., 2nd; Smith, R. D.; Qian, W. J., Resin-assisted enrichment of thiols as a general strategy for proteomic profiling of cysteine-based reversible modifications. *Nat Protoc* **2014,** *9* (1), 64-75.

129. Development of Quantum Dot Probes for Studies of Synergy Between Components of the Wood-Degrading Fungal Enzymes

Simon A. McManus[1*], Kevin D. Welsher[1], Venkata Giridhar Poosarla[2], **Ming Tien**[2], **B. Tracy Nixon**[2], and **Haw Yang**[1]

[1]Department of Chemistry, Princeton University, Princeton, NJ
[2]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA

*smcmanus@princeton.edu

http://www.princeton.edu/~yanglab/research/energy-biol.shtml

Project Goals: Our team is focused on developing a comprehensive understanding for the actions of cellulose degrading enzymes. Our team's approaches to uncover mechanistic insights into cellulose degradation involve 1) Expression and purification of cellulases from bacterial and fungal sources 2) One-to-one conjugation of expressed enzymes to custom-designed fluorescent quantum dots, which provide constant, strong fluorescence for the imaging of cellulases on the single enzyme level over long timescales needed to observe their predicted slow, processive action 3) Real-time multi-scale 3D imaging of cellulases in complex with cellulose using a novel multi-resolution microscopy, allowing images of mobile cellulases in three dimensions with very high localization precision. These experiments will allow the generation of models for the mechanisms of different cellulases and determination of their synergistic effects, paving the way for the engineering of more efficient cellulases for biomass conversion.

As the negative environmental impacts of fossil fuel use become increasingly apparent, it is imperative that other renewal fuel sources with low carbon footprints are developed. Plant biomass, particularly lignocellulose, is an abundant source of bioenergy that could potentially be harnessed without any net carbon emissions and utilized without competing with existing food crops. However, current methods for breaking down lignocellulose into biofuel precursor sugars are costly and inefficient. Fungal and bacterial species, on the other hand, have evolved effective means to breakdown lignocellulose using a plethora of cell wall-degrading enzymes. Through a better understanding of these enzymes it may be possible to develop schemes to utilize, and even engineer, microorganisms and their enzymes to efficiently convert lignocellulose into a viable fuel source. Study of the enzymes that breakdown lignocellulose, including cellulases, has been hampered by the factors such as the heterogeneous nature of their substrate and the fact that they carry out interfacial catalysis on a solid substrate. Single-molecule experiments offer a means to characterize these cellulases by observing the movements of signal cellulase enzymes on natural substrates. For example, these experiments will shed light on processivity of these enzymes and give insight into how these enzymes can breakdown these large recalcitrant strands of cellulose. These constitute the first steps in developing these enigmatic enzymes into workhorses for biofuel production.

Team members at Princeton University and Penn State have been developing technologies to allow the study of cellulase enzymes on the single-molecule level. Currently, two fungal cellulases and one bacterial cellulase, all engineered with orthogonal tags to allow conjugation to luminescent quantum dots, have been expressed and purified. These three cellulases cover the three main modes of action for known cellulases, being reducing end exocellulase activity (*T.reesei* Cel7a), non-reducing end exocellulase activity (*T. reesei* Cel6a) and endocellulase activity (*T. fusca* Cel6a). As proper

glycosylation of fungal cellulases has been found to be important for activity and incorrect amounts of glycolyation has been seen when fungal enzymes as expressed in high expression strains, the tagged fungal cellulases in our study have been expressed from their native species *T. reesei*, with proper levels of glycosylation confirmed by mobility assays. To label the cellulases for imaging, we have carried out and optimized synthesis of giant quantum dots (gQDs), previously established to mitigate switching between fluorescent and dark states, known as 'blinking' behavior, seen with conventional quantum dots due to defects on the quantum dots surface (1). The synthesis of gQDs has been carried out through successive additions of inorganic layers to a CdSe core to mitigate these surface effects. Analysis of the spectral properties of these gQDs has revealed that they do not switch between fluorescent and dark states and display constant fluorescence for minutes on the single-molecule level using laser powers needed for single-molecule experiments. The conjugation of one gQD to one cellulase molecule is essential for real-time imaging of cellulases on their cellulose substrate. This one-to- one conjugation is not a trivial matter, as cellulases must be bound to the spherical surface of a quantum dot with the potential of multiple enzymes binding to each quantum dot. We have designed an electrophoretic separation procedure using very low gel concentrations and high voltage to allow the separation of gQD-cellulase conjugates of different stoichiometries. This procedure has been modified to allow the one-to-one conjugation of the different cellulases to their gQD labels. To follow the activity of single cellulase-gQD conjugates in real-time in the context of the substrates, we have developed an entirely new imaging modality. It capitalizes on the real-time capabilities previously developed in the Yang lab (2,3). This multi-resolution imaging system can concurrently image objects with large differences in size and has recently been used to observe the interaction of peptide-coated nanoparticles with living cells as a proof of concept (4). This multiple resolution capacity allows us to follow the movement of gQD- labeled cellulases in real-time, while simultaneously imaging the fluorescently labeled cellulose substrate that it is interacting with. This allows correlation of cellulase-cellulose interaction with cellulase movement and gives direct evidence of cellulase processivity. Single molecule trajectories of many individual cellulase molecules are being collected to gain a better understanding of the mechanisms of these enigmatic enzymes.

References:

1. Chen, Y., Vela, J., Htoon, H., Casson, J.L., Werder, D.J., Bussian, D.A., Klimov, V.I. and Hollingsworth, J.A. *J. Am. Chem. Soc.* (2008) **130** 5026–5027
2. Cang, H., C.M. Wong, C.S. Xu, A.H. Rizvi, and H. Yang *Appl. Phys. Lett.* (2006) **88** 223901.
3. Cang, H., Xu C.S., Montiel, D., and Yang, H. *Opt. Lett.* (2007) **32** 2729-2731.
4. Welsher, K. and Yang, H. *Nat. Nanotech.* (2014) *in press*

130. A Fully Reversible, Highly Energy Efficient Glycolysis with Unique Cofactor Utilization in *C. cellulolyticum*

Daniel Amador-Noguez,[1] Junyoung O. Park[1,2]* (jopark@princeton.edu), Zheyun Zhang,[1] Sara A. Rubin,[3] Wenxin Xu,[1] and **Joshua D. Rabinowitz[1,3]**

[1]Lewis-Sigler Institute for Integrative Genomics; [2]Department of Chemical and Biological Engineering; and [3]Department of Chemistry, Princeton University, Princeton, N.J.

Project Goals: The production of biofuels from cellulosic biomass holds promise as a source of renewable clean energy. Members of the genus Clostridium collectively have the ideal set of the metabolic capabilities for biofuel production from cellulosic biomass: *C. acetobutylicum* rapidly ferments glucose to biofuels (butanol, hydrogen) and *C. cellulolyticum* effectively degrades cellulose. Here we aim to integrate metabolomics, genomics and genetic engineering to dramatically advance understanding of metabolism in *C. acetobutylicum* and *C. cellulolyticum*. In so doing, we will lay basic science groundwork for engineering of an organism that cost-effectively converts cellulose into solvents.

C. cellulolyticum is an obligate anaerobe capable of degrading cellulose into simple sugars and converting them into useful biofuels such as ethanol. However, conversion yields and production rates are too low to allow its use for commercial production of biofuels. For example, compared to the typical biofuel producer *C. acetobutylicum*, *C. cellulolyticum* has a very slow sugar catabolism, even when growing on simple sugars such as glucose or cellobiose. To gain a quantitative understanding of the sugar catabolic pathways in this bacterium, we utilize metabolomic tools, in combination with isotope tracers and quantitative flux modeling.

We report that glycolysis in *C. cellulolyticum* is fully reversible, with all of its reactions working near equilibrium. Such reversibility is achieved by replacing the cofactor ATP with pyrophosphate or GTP as the high-energy phosphate donors/acceptors as well as by using canonically anaplerotic and gluconeogenic reactions for glucose catabolism. This results in a highly energy-efficient sugar catabolism that generates more energy (more ATP equivalents) than canonical glycolytic pathways. The unique glycolysis in *C. cellulolyticum* reflects the evolution of the metabolic pathway to cope with the low energy availability, which can be attributed to anaerobiosis on cellulose.

Funding: DOE-DE-SC0006839

131. An integrated 'omics approach to large-scale quantitative analysis of cellular metabolic regulation

Sean R. Hackett, Vito Zanotelli, Jonathan Goya, David H. Perlman, **Joshua D. Rabinowitz** Department of Chemistry and Lewis-Sigler Institute for Integrative Genomics, Princeton University

Project Goals: Metabolic transformation of plant material into biofuels holds promise as a source of renewable clean energy. While the major ongoing focus of our DOE-funded research regards metabolism in the promising biofuel-producing genus Clostridium, we also have a broad interest in methodologies for understanding metabolism and its regulation. The overarching aim of the present work is to develop a quantitative, genome-scale approach to elucidation of metabolic regulation in biofuel-relevant microbes.

We propose a novel, general, and scalable strategy for discovering metabolic regulation through integrative 'omics. Previously, we and others have attempted to elucidate regulation through measurement of time-dependent changes in metabolite concentrations (e.g., in response to acute nutrient perturbation) and dynamical modeling of the resulting data. This approach has had notable successes, especially for small networks. Larger dynamical models of nonlinear systems such as metabolism, however, are often intractable. Indeed, even in modeling small networks, we found that the most informative analyses often focused on pre- and post-perturbation pseudo-steady states, rather than dynamics. Specifically, we realized that, at steady-states, regulation can be dissected on an enzyme-by-enzyme basis if one knows the enzyme's concentration, output (flux) and inputs (concentration of all relevant substrates, products, and effectors). Accordingly, we hypothesized that large-scale analysis of different metabolic steady states via integrative 'omics could potentially be a tractable and effective method for revealing metabolic regulation.

To explore this possibility, we elected to use *Saccharomyces cerevisiae* (itself an important biofuel producer) as a test organism. Cells were grown in chemostats at 25 different steady states. Concentrations of metabolites were measured by LC-MS-based metabolomics and of metabolic enzymes by LC-MS/MS-based proteomics. To infer fluxes, uptake and excretion rates of the diversity of metabolites were measured, as was detailed biomass composition; together these measurements were sufficient to constrain a genome-scale flux-balanced metabolic model, resulting in reliable determination of many core metabolic fluxes. Full information (flux, enzyme concentration, and all relevant metabolite concentrations) was obtained across all 25 conditions for ~ 50 enzymes.

For these ~ 50 enzymes, we assessed whether variation in flux across the 25 experimental conditions could be explained based on the enzyme and metabolite concentrations using an equation of the Michaelis-Menten form. This allowed us to determine Michaelis-Menten parameters, based not on isolated biochemistry but physiological cellular data. For about one-third of enzymes, the concentrations of the enzyme, substrates, and products alone were sufficient to explain the observed fluxes. For another approximately one-third of enzymes, the observed fluxes could be explained if one also included metabolite concentration data for potential effectors (e.g., fructose-1, 6-bisphosphate activation of pyruvate kinase). For the final one-third of enzymes, we have yet to elucidate the missing regulation and/or there is too much noise in the data to obtain a good fit.

Beyond providing proof of concept for integrative 'omic analysis of metabolic regulation, this work also addresses some bigger picture questions, e.g., how much of metabolic flux control resides in enzyme concentrations versus metabolite concentrations? Our results show that across the tested physiological steady states, enzyme and metabolite concentrations make nearly equal flux-control contributions. The general strategy of analyzing many metabolic steady states via integrative 'omics (of fluxes, enzymes, metabolites) thus holds the potential to address long-standing global questions regarding the nature of

metabolic regulation, as well as to identify specific physiologically-relevant instances of regulation both in well-studied organisms like S. cerevisiae and in less studied ones like Clostridia.

Funding: DOE-DE-SC0006839

132. Recombinant *Bacillus subtilis* that grows on untreated plant biomass

Timothy D. Anderson[1-3], J. Izaak Miller[1-3], Grace Huang[1-3], Henri-Pierre Fierobe[4], and **Robert T. Clubb**[1-3,*]

[1]UCLA-DOE Institute for Genomics and Proteomics, [2]Molecular Biology Institute and [3]Department of Chemistry and Biochemistry, University of California, Los Angeles. 611 Charles E. Young Drive, Los Angeles, California 90095-1570, USA. [4]Laboratoire de Chimie Bactérienne, IFR88-CNRS Marseille, France.

http://www.doe-mbi.ucla.edu/

Project Goals: The UCLA-DOE Institute for Genomics and Proteomics carries out research in bioenergy, structural biology, genomics and proteomics, consistent with the research mission of the United States Department of Energy. Lignocellulosic biomass is a promising feedstock to produce biofuels and other valuable biocommodities. A major obstacle to its commercialization is the high cost of degrading biomass into fermentable sugars. Here we explore the use of microbes to break down biomass. *Bacillus subtilis* was engineered to display a multi-cellulase containing minicellulosome. The complex contains a miniscaffoldin protein that is covalently attached to the cell wall and three non-covalently associated cellulase enzymes derived from *Clostridium cellulolyticum* (Cel48F, Cel9E, and Cel5A). A unique feature of the minicellulosome is that it spontaneously assembles, thus increasing the practicality of the cells. The recombinant bacteria are highly cellulolytic and grow in minimal media containing industrially relevant forms of biomass as the primary nutrient source (corn stover, hatched straw, and switchgrass). Notably, growth did not require dilute acid pretreatment of the biomass and the cells achieved densities approaching those of cells cultured with glucose. An analysis of the sugars released from acid pretreated corn stover indicates that the cells have stable cellulolytic activity that enables them to break down 62.3 ± 2.6% of the biomass. When supplemented with beta-glucosidase, the cells liberated 21% and 33% of the total available glucose and xylose in the biomass, respectively. The cells produce one third as much reducing sugars upon exposure to acid treated corn stover as compared to a multi-cellulase containing cocktail. As the cells display only three types of enzymes, increasing the number of displayed enzymes should lead to more potent cellulolytic microbes. This work has important implications for the efficient conversion of lignocellulose to value-added biocommodities.

References

1. Anderson TD, Robson SA, Jiang XW, Malmirchegini, GR, Fierobe, H-P, Lazazzera BA and Clubb, RT. Assembly of minicellulosomes on the surface of *Bacillus subtilis*. *Applied and Environmental Microbiology 77* 2011; 4849-58.
2. Anderson TD, Miller JI, Fierobe H-P and Clubb RT. Recombinant *Bacillus subtilis* that grows on untreated plant biomass. *Applied and Environmental Microbiology 79* 2013; 867-76.

The UCLA-DOE Institute for Genomics and Proteomics is supported by the Office of Biological and Environmental Research in the DOE Office of Science

133. The small RNome of *Clostridium acetobutylicum* that orchestrates metabolite stress response

Keerthi P. Venkataramanan[1,2,*] (venkatkp@udel.edu), Shawn W. Jones[1,2] Kevin P. McCormick[2,3], Sridhara G. Kunjeti[2,4], Matthew T. Ralston[2,5], Blake C. Meyers[2,4] and **Eleftherios T. Papoutsakis**[1,2]

[1]Department of Chemical and Biomolecular Engineering, University of Delaware. [2]Delaware Biotechnology Institute, University of Delaware. [3]Department of Computer and Information Sciences, University of Delaware, Newark. [4]Department of Plant and Soil Sciences, University of Delaware. [5]Center of Bioinformatics and Computational Biology, University of Delaware, Newark, DE 19701, USA.

Project Goal: The objective of this project is to engage next generation sequencing technology for the identification the differential expression of small non-coding regulatory RNAs (sRNAs). The outcome of this study will result in the understanding the complexity, paradigm and importance of toxic metabolite stress in *C. acetobutylicum*.

Abstract:
Regulatory small non-coding RNAs (sRNA) have been identified in several Gram$^+$ and Gram$^-$ prokaryotes and are emerging as major, previously unrecognized, components of the cellular regulatory network and structural machinery, several possessing their own regulons. Among those, few sRNAs have been reported as being involved in toxic metabolite stress, mostly in Gram$^-$ prokaryotes, but hardly any in Gram$^+$ prokaryotes and still, their role at the systems level remains poorly understood, especially so in metabolite and general stress responses. In the important genus of *Clostridium*, which is of major importance to pathogenesis, human physiology, the carbon cycle and biotechnological applications, very few sRNAs have been so far identified.

Using RNA deep sequencing (RNA-seq) we examined the sRNome of *C. acetobutylicum* in response to the native but toxic metabolites, butanol and butyrate. 50% of the RNA-seq reads mapped to genomic DNA outside annotated ORFs, thus demonstrating the richness and importance of the small RNome. Using the 113 sRNAs we had previously computationally predicted [1] together with annotated mRNAs, we set metrics for reliably identifying sRNAs from RNA-seq data, thus discovering 46 additional sRNAs. Under metabolite stress, these 159 sRNAs displayed distinct expression patterns, a select number of which was verified by Northern analysis. We identified stress-related expression of sRNAs affecting transcriptional (6S & S-box, *solB*) and translational (*tmRNA* & *SRP-RNA*) processes, and 65 likely targets of the RNA chaperone Hfq. Our results support an important role for sRNAs in toxic-metabolite stress response [2].

This is the first study to elucidate the role of sRNAs in clostridial response to metabolite stress and is essential for understanding the complexity of the regulatory network that underlies the metabolite-stress response, whether related to normophysiology, pathogenesis or biotechnological applications and how that network can be engineered for practical applications to produce chemicals and fuels or for remediation processes.

References

1. Chen, Y., et al., *Small RNAs in the genus Clostridium.* MBio, 2011. **2**(1): p. e00340-10.
2. Venkataramanan, K.P., et al., *The Clostridium small RNome that responds to stress: the paradigm and importance of toxic metabolite stress in C. acetobutylicum.* BMC Genomics, 2013. **14**: p. 849.

Acknowledgement: *This work was supported by the genomic sciences program grant from Department of Energy, USA (grant#DESC0007092).*

134. Experimental Systems-Biology Approaches for Clostridia-Based Bioenergy Production: The Metabolite Stress-Response System in Solventogenic Clostridia

E. Terry Papoutsakis[1,2]* (epaps@udel.edu), Blake C. Meyers[2,3], Kelvin H. Lee[1,2], Maciek R. Antoniewicz[1], Costas D. Maranas[4], H. Huang[2], and **Cathy H. Wu**[2,5]

[1]Department of Chemical & Biomolecular Engineering, University of Delaware, Newark, DE; [2]Delaware Biotechnology Institute, Newark, DE; [3]Department of Plant and Soil Sciences, University of Delaware, Newark DE; [4]Department of Chemical Engineering, Pennsylvania State University, University Park, PA; [5]Department of Biochemistry and Molecular & Cellular Biology, Georgetown University Medical Center, Washington DC

http://www.clostress.org/

Project Goals: The objectives of this project are to engage enabling experimental systems-biology approaches to support the development of integrated, predictive models of the metabolic and regulatory networks underlying the metabolite stress response in solventogenic clostridia. Clostridia are Gram$^+$, obligate anaerobic, endospore-forming bacteria of major importance to fermentative biofuel production. Here, we focus on understanding and modeling the stress-response of *Clostridium acetobutylicum* to two important toxic metabolites: butanol and butyrate. Systems-level understanding is expected to lead to better strategies for industrial-strain development, as well as bioprocessing strategies taking advantage of the stress response to achieve superior bioprocessing outcomes.

Solventogenic and other clostridia are of major importance for developing technologies for biofuel production [1]. A major and unique advantage is their ability to utilize a large variety of substrates (hexoses, pentoses, oligosaccharides, xylan, and starches). Among the two sequenced solventogenic clostridia, *C. acetobutylicum* (*Cac*) is the only one that contains a full cellulosome [2] and may thus directly utilize cellulosic material for production of fuels and chemicals.

The toxic-metabolite stress response is a problem of major and general importance not only in clostridial biotechnologies but in all microbial systems of interest to bioenergy production [3]. In this project we investigated the metabolite stress response by collecting extensive transcriptomic (based on both deep sequencing and microarray analyses) and targeted fluxomic and proteomic data. These data have been used to identify differentially expressed genes during stress conditions along with the coupling of omics data integration with building stress models and modeling platforms that can be linked, as an added modeling dimension, to a 2nd generation GSM of this organism resulting in an in-depth systems-level molecular understanding at multiple genome-scale levels of the metabolite stress response.

Using RNA deep sequencing, we identified the role and expression of small non-coding RNAs (sRNAs) in stress response, apart from identifying 56 novel stress related sRNAs [4]. Furthermore, we identified the transcription factors and genetic circuits orchestrating the complex and multilayered response to butanol and butyrate stress [5]. Quantitative proteomic analysis was performed using iTRAQ tags. 566 and 588 unique proteins were identified and quantified from *Cac* grown under butanol and butyrate stress, respectively. 13C-Metabolic flux analysis is a widely used technique for measuring in vivo metabolic fluxes. Here, we have applied three 13C-labeled amino acid tracers, [1-13C]aspartate, [4-13C]aspartate, and [1-13C]serine to elucidate metabolic cycling between amino acid metabolism and central carbon metabolism in *Cac*. Such cycles are difficult to detect using traditional 13C-glucose tracers. We demonstrate, for the first time, that *Cac* has a highly active metabolic cycle between aspartate and

pyruvate. This metabolic cycle allows clostridium to rapidly interconvert several amino acids that are needed for cell growth and the adaptation to metabolite stress.

A genome-scale model, iCAC802, was developed for *Cac*, and includes 802 genes and 1470 reactions. The model was validated by comparison of in silico results to experimental gene deletion and 13C-MFA data. Gene transcription data for *Cac*'s response to butanol and butyrate stress was incorporated as regulation into the model using the E-Flux method. The regulated model exhibited reduction in biomass yield and down-regulation of glucose uptake as observed experimentally under stress conditions. The regulated model could be used for stress-specific metabolic predictions to aid redesign in pursue of a desired phenotype.

This multidimensional platforms and models based approach leads to an in-depth understanding of the metabolite stress response at molecular level to create a more streamlined genome and engineered strains with a better understanding of the complexity of network and regulation at molecular level. This aspect of project will further result in the development of computational and bioinformatics tools and frameworks towards modeling other complex cellular programs. Overall, this project's outcomes aim to become at enabling paradigm for modeling complex programs of organisms and biological systems of importance to DOE's mission on energy and the environment.

References

1. Tracy, B.P., et al., *Clostridia: the importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications.* Curr Opin Biotechnol, 2011.
2. Nolling, J., et al., *Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum.* J Bacteriol, 2001. **183**(16): p. 4823-38.
3. Nicolaou, S.A., S.M. Gaida, and E.T. Papoutsakis, *A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation.* Metab Eng, 2010. **12**(4): p. 307-31.
4. Venkataramanan, K.P., et al., *The Clostridium small RNome that responds to stress: the paradigm and importance of toxic metabolite stress in C. acetobutylicum.* BMC Genomics, 2013. **14**: p. 849.
5. Wang, Q., et al., *Transcription factors and genetic circuits orchestrating the complex, multilayered response of Clostridium acetobutylicum to butanol and butyrate stress.* BMC Syst Biol, 2013. **7**(1): p. 120.

Acknowledgement: *This work was supported by the genomic sciences program grant from Department of Energy, USA (grant#DESC0007092).*

135. A Systems Biology, Whole-Genome Association Analysis of the Molecular Regulation of Biomass Growth and Composition in *Populus deltoides*

Annette M. Fahrenkrog,[1] Leandro G. Neves,[1] Juan J. Acosta,[2] William "Brad" Barbazuk,[1,3,4]
Kirst M[1,2,3]* (mkirst@ufl.edu)

[1]Plant Molecular and Cellular Biology Program, University of Florida, Gainesville; [2]School of Forest Resources and Conservation, University of Florida, Gainesville; [3]Genetics Institute, University of Florida, Gainesville; [4]Department of Biology, University of Florida, Gainesville

Project Goals: Poplars trees are well suited for biofuel production due to their fast growing habit, favorable wood composition and adaptation to a broad range of environments. The availability of a reference genome sequence, ease of vegetative propagation and availability of transformation methods also make poplar an ideal model for the study of wood formation and biomass growth in woody, perennial plants. The objective of this project is to conduct a genome-wide association genetics study to identify genes that regulate bioenergy traits in *Populus deltoides* (eastern cottonwood). Association mapping is being pursued by combining sequence-capture followed by high-throughput sequencing to genotype coding and regulatory sequences in the whole-genome of *P. deltoides*. To identify genetic polymorphisms that regulate biomass productivity and carbon partitioning we are pursuing the following goals: (1) optimizing sequence-capture for unbiased, high- throughput and low cost recovery of target coding and regulatory sequences in P. deltoides; (2) carrying out "whole-genome" genotyping of a *P. deltoides* unstructured population for association mapping; and (3) identifying significant SNP-trait associations with biomass growth and carbon partitioning to define genes and alleles that regulate trait variation.

To develop an optimal platform for sequence capture in *Populus* we initially measured the efficiency of ultra-long oligonucleotides in retrieving specific sequences of the genome of *P. deltoides* for detection of SNP polymorphisms. Oligonucleotides that tile regulatory and coding sequences of all poplar genes previously identified as expressed in the main vegetative tissues were designed. Next, genomic DNA from three *P. deltoides* individuals previously sequenced by DOE's Joint Genome Institute, were hybridized to oligonucleotides in solution, captured and sequenced. Captured fragments were aligned to the reference sequence, and the degree of sequence enrichment and the power to detect known SNPs was evaluated. Oligonucleotide probes that were effective for capture and genotyping included exons and parts of the 5' and 3' untranslated regions (UTRs) of 18,153 genes.

Using the sequence-capture oligonucleotides developed previously, we analyzed a *P. deltoides* association population composed of 579 unrelated individuals. This is a subset of a larger population composed of 815 unrelated genotypes collected from 13 states in Central, South and Eastern US, covering 35 river systems. In total, we identified approximately one million single nucleotide polymorphism (SNP) markers in the whole population, distributed along the genes captured. In addition, we also captured and genotyped regions distributed every 15 Kbp in the genome, to provide a genome-wide view of the genetic diversity of the species, to be used for population genetic studies. In parallel we propagated the association population and established it in greenhouse and field test sites for biomass and lignocellulosics composition measurements.

Currently the field trials are completing one year of age. Wood samples collected from the greenhouse trials have been collected and phenotyped using pyrolysis MBMS. Detailed analysis of these phenotypic measurements is currently in progress.

With the near completion of the experiments planned for the first two goals of this project, we are now initiating the last aim. Specifically, the analysis will begin with the identification of any family and population structure in the population. Analysis of SNP-trait associations that account for covariance due to relatedness among individuals will be carried out using analysis of variance, followed by a combined analysis of the most highly associated SNPs in one single Bayesian model to estimate joint, epistatic effects of multiple loci.

136. Understanding Fundamental Aspects of Butanol Production by *Clostridium beijerinckii*

Yi Wang[1], Holger Janssen[1], Caroline B. Milne[1,2], Andrew T. Magis[1,2], Yong-Su Jin[1], Nathan D. Price[1,2] and **Hans P. Blaschek[1*]**

1 University of Illinois, Urbana IL; 2 Institute for Systems Biology, Seattle WA

http://bioenergy.illinois.edu

Project Goals: The solventogenic clostridia offer a sustainable approach to petroleum- based production of n-butanol, an important chemical feedstock and potential fuel. With the availability of the genome sequence for *Clostridium beijerinckii* 8052, we can now employ the tools of systems biology and omics technologies in order to gain increased insights into the metabolic and regulatory networks relevant to solvent production. Project goals include examination of: 1) the mutations underlying the *C. beijerinckii* BA101 butanol-overproduction phenotype, 2) the molecular basis for the global shift from acidogenesis to solventogenesis, 3) the genetic basis of butanol tolerance in *C. beijerinckii* and 4) RNA-seq technology for single-nucleotide resolution analysis of the transcriptome of this microorganism.

We sequenced four *Clostridium beijerinckii* mutant strains (*C. beijerinckii* BA101, BA105, SA1 and SA2) in combination with our laboratory wild type strain *C. beijerinckii* NCIMB 8052 (UIUC) to examine their genetic differences. Preliminary results show 17 genomic variations (small nucleotide polymorphisms (SNPs), deletions, nucleotide changes, insertions, duplications etc.) in *C. beijerinckii* BA101, 23 genomic variations in *C. beijerinckii* BA105, 15 genomic variations in *C. beijerinckii* SA1 and 55 genomic variations in *C. beijerinckii* SA2. Among these, we discovered two unique polymorphisms for *C. beijerinckii* BA101, seven unique polymorphisms for *C. beijerinckii* BA105 and 23 unique polymorphisms *C. beijerinckii* SA2. Interestingly, *C. beijerinckii* SA1 has no unique genomic variation and shares all 15 alterations with *C. beijerinckii* BA101 and *C. beijerinckii* BA105. These results will lead to specific gene knock-out targets to investigate single genomic variations and their influence on butanol production and/or tolerance.

Based on the mobile group II intron technology, we have developed a Targetron gene knockout system for *C. beijerinckii* (Wang et al., 2013). This system was successfully employed to disrupt acid production pathways in *C. beijerinckii*, leading to *pta* (encoding phosphotransacetylase) and *buk* (encoding butyrate kinase) negative mutants. Compared to the parental strain (*C. beijerinckii* 8052), acetate production in the *pta* mutant was substantially reduced and butyrate production increased. In contrast, acetate and butyrate production in the *buk* mutant was similar to that of the wild type, but solvent production was consistently 20-30% higher and glucose consumption was more rapid and complete. The characterization results suggest that the acid and solvent production of *C. beijerinckii* can be effectively altered by disrupting the acid production pathways. As the gene disruption method we developed does not leave behind an antibiotic marker in the disrupted allele, multiple and high-throughput gene disruptions are possible. Based on this system, we have constructed more than 20 knockout mutants for *C. beijerinckii*, including several double and triple knockout mutants. The characterization of these mutants is currently underway. The results will provide essential information for understanding the basic metabolism of the ABE process in *C. beijerinckii*, and will guide the further improvement of *C. beijerinckii* strains through systems biology and genetic engineering based approaches.

The data collected from metabolic profiling of *C. beijerinckii* 8052 serve to compliment our genome-scale *C. beijerinckii* model (Milne et al., 2011) by adding dynamic information about metabolites in the network in addition to the static stoichiometric equations obtained from known genome-annotations. We observed shifts in metabolism over the course of an 84 hour fermentation, highlighting the transition to saturated fatty acid production to counteract butanol toxicity as butanol concentrations become maximal. Additionally, new information concerning the production and consumption of lactic acid was observed.

We gained additional insight into the physiology of *C. beijerinckii* BA105. This strain demonstrates an acid crash behavior in batch with little butanol being produced when compared to the wild type strain. It accumulates acetate and butyrate and shows significant decreased sporulation; finally, the metabolic switch to acetone-butanol-ethanol (ABE) fermentation does not occur. However, we were able to prevent 'acid-crash' of *C. beijerinckii* BA105 by controlling the fermentation pH. Under controlled conditions, the BA105 strain shows ~50% higher glucose consumption and up to 45% more n-butanol production when compared to the wild type strain. Furthermore, the sporulation capacity of *C. beijerinckii* BA105 is completely restored similar to the wild type. These results lead to the conclusion that *C. beijerinckii* BA105 can be used as a model organism for examining the metabolic switch from acid to solvent production on a genetic and transcriptional level using both non pH-controlled ('acid crash') and pH-controlled (increased butanol production) batch growth experiments.

We also conducted a comprehensive transcriptional analysis using RNA-Seq approach for *C. beijerinckii* BA105 and *C. beijerinckii* 8052 growing on glucose as sole carbon source with pH controlled at 5.5. We examined changes in gene expression in order to evaluate gene targets responsible for higher glucose consumption and higher butanol production. Preliminary results showed significant changes in the phosphotransferase systems (PTS) which are potentially responsible for the increased glucose consumption by *C. beijerinckii* BA105. Furthermore, genes associated with the acid branch (*pta-ack*, *ptb-buk*) demonstrated increased transcript levels and this correlates with the acid accumulation during the growth of the mutant strain. Interestingly, the *sol* operon (*cbei_3832-cbei_3835*) shows decreased transcription in the early growth phase of *C. beijerinckii* BA105, whereas alternative aldehyde-alcohol dehydrogenases are significantly overexpressed, potentially resulting in higher butanol values when compared to the wild type. The combination of acid accumulation and higher butanol titer also led to increased transcript levels of stress genes at the end of the fermentation (e.g., *groES/EL*, *dnaK*). Several sporulation genes were highly repressed in *C. beijerinckii* BA105, although the strain shows restored sporulation when using pH 5.5 controlled conditions.

References

Y. Wang, et al. 2013. Development of a gene knockout system using mobile group II introns (targetron) and genetic disruption of acid production pathways in *Clostridium beijerinckii*," Applied and Environmental Microbiology. 79:5853-5863.

Milne, C.B, et al. 2011. Metabolic network reconstruction and genome-scale model of butanol- producing strain *Clostridium beijerinckii* NCIMB 8052. BMC Systems Biology. 5:130.

This work was funded by the DOE Office of Science through FOA: DE_FOA-0000368 to HPB and a NSF CAREER grant to NDP.

137. A New DIET for Methanosarcina barkeri: Direct Interspecies Electron Transfer in a Genetically Tractable Methanogen

Amelia-Elena Rotaru, Pravin Malla Shrestha, Fanghua Liu, Beatrice Markovaite, Minita Shrestha, Kelly P. Nevin, and **Derek R. Lovley** (dlovley@microbio.umass.edu)

Department of Microbiology, University of Massachusetts, Amherst

http://www.geobacter.org

Project Goals: The long-term goal of our project, which is entitled "Systems Level Analysis of the Function and Adaptive Responses of Methanogenic Consortia", is to develop genome-scale metabolic models of microbial communities that play an important role in the global carbon cycle that can be coupled with the appropriate physical-chemical models to predict how the microbial communities will respond to environmental perturbations, such as climate change. The short-term objective of the current research is to elucidate the mechanisms for direct interspecies electron transfer (DIET), the diversity of methanogens that are capable of DIET, and the prevalence of DIET in soils and sediments that make important contributions to atmospheric methane.

Effective interspecies electron transfer is key to the smooth functioning of methanogenic communities. Promoting interspecies electron transfer to methanogens enhances the anaerobic digestion of wastes and appropriate models of the pathways for interspecies electron transfer are necessary in order to predictively model the response of methanogenic communities to environmental change. For over 40 years interspecies hydrogen transfer has served as the paradigm for anaerobic interspecies electron transfer. However, our recent studies demonstrated that direct interspecies electron transfer (DIET) is possible[1-5] and may be the predominant mechanism for electron exchange in some methanogenic environments[5,6].

We recently reported[5] the surprising finding that methanogens in the genus *Mehthanosaeta* can accept electrons for the reduction of carbon dioxide via DIET. *Methanosaeta* species are considered to produce more methane on earth than any other group of methanogens. They are important contributors to atmospheric methane resulting from methane production in soils and sediments and are often the most active methanogens in digesters converting wastes to methane. *Methanosaeta* were previously considered to be limited to acetate as a substrate for methane production. However, metatranscriptomic analysis of anaerobic waste digester aggregates, as well as transcriptomic, radiotracer, and genetic analysis of defined co-cultures in which *Methanosaeta harundinacea* served as the sole methane- producing partner, revealed that *Methanosaeta* species can function as the primary electron-accepting organism in syntrophic partnerships, by accepting electrons via DIET[5].

In our most recent studies the potential for other methanogens to participate in DIET was evaluated. A number of methanogens that have been reported to exclusively utilize H_2 or formate as electron donors could not participate in DIET, but *Methanosarcina barkeri* could. Like *Methanosaeta* species, *M. barkeri* is able to use acetate as a substrate for methane production, but unlike *Methanosaeta* species, *M. barkeri* can also use H_2 as an electron donor for carbon dioxide reduction. When cultured in media with ethanol as the electron donor, *M. barkeri* formed syntrophic cultures with *Pelobacter carbinolicus* in which the two species exchanged electrons via interspecies H_2 transfer, whereas in co-culture with *Geobacter metallireducens*, *M. barkeri* accepted electrons via DIET. Analysis of the transcriptome of the two syntrophic cultures revealed increased transcript abundance for genes encoding putative filaments with similarity to the electrically conductive[7] type IV pili of *Geobacter* species. Furthermore, there was a

distinctive increase in transcript abundance for genes encoding several outer surface proteins in *M. barkeri* growing via DIET. These results suggest that *M. barkeri* expresses one or more outer-surface proteins to facilitate DIET.

M. barkeri is only the second methanogen found to be capable of DIET and the first methanogen known to have the option of accepting electron either via H_2 transfer or DIET. *Methanosarcina* are often among the most abundant methanogens in methanogenic soils and sediments, landfills, and anaerobic digesters. However, other than the obvious importance of DIET in some anaerobic digesters treating brewery waste[5,6], the prevalence of DIET in other methanogenic environments is unknown. The fact that at least two major genera of methanogens have evolved the capacity for DIET suggests that there are conditions in some soils and sediments in which DIET confers a selective advantage.

Although the importance of electrically conductive pili and outer-surface cytochromes in extracellular electron exchange, including DIET, is well-known for *Geobacter* species[8], it is premature to speculate on the potential extracellular electron contacts that might permit methanogens to accept electrons via DIET. The availability of tools for genetic manipulation of *barkeri*[9] suggests that it may be the ideal candidate for functional analysis of DIET mechanisms in methanogens.

This research was supported by the Office of Science (BER), U. S. Department of Energy, Award No. DE-SC0004485.

1. Summers ZM, Fogarty HE, Leang C, Franks AE, Malvankar NS, Lovley DR. 2010.
2. *Science* 330:1413-1415.
3. Liu F, Rotaru A-E, Shrestha PM, Malvankar NS, Nevin KP, Lovley DR. 2012. Promoting direct interspecies electron transfer with activated carbon. *Energy Environ. Sci.* 5:8982- 8989.
4. Shrestha PM, Rotaru A-E, Summers ZM, Shrestha M, Liu F, Lovley DR. 2013. Transcriptomic and genetic analysis of direct interspecies electron transfer. *Appl. Environ. Microbiol.* 79:2397-2404.
5. Shrestha PM, Rotaru A-E, Aklujkar M, Liu F, Shrestha M, Summers ZM, Malvankar N, Flores DC, Lovley DR. 2013. Syntrophic growth with direct interspecies electron transfer as the sole mechanism for energy exchange. *Enivron. Microbiol. Rep.* 5:904-910.
6. Rotaru A-E, Shrestha PM, Liu F, Shrestha M, Shrestha D, Embree M, Zengler K, Wardman C, Nevin KP, Lovley DR. 2014. A new model for electron flow during anaerobic digestion: direct interspecies electron transfer to *Methanosaeta* for the reduction of carbon dioxide to methane. *Energy Environ. Sci.* 7:408-415.
7. Morita M, Malvankar NS, Franks AE, Summers ZM, Giloteaux L, Rotaru AE, Rotaru C, and Lovley DR. 2011. *mBio* 2:e00159-11.
8. Malvankar NS, Vargas M, Nevin KP, Franks AE, Leang C, Kim BC, Inoue K, Mester T, Covalla SF, Johnson JP, Rotello VM, Tuominen MT and Lovley DR. 2011. *Nature Nanotechnol.* 6:573-579.
9. Malvankar NS, Lovley DR. 2014. Microbial nanowires for bioenergy applications. *Curr. Opin. Biotechnol.* 27:88-95.
10. Kohler PRA, Metcalf MW. 2012. Genetic manipulation of *Methanosarcina* spp.
11. *Frontiers Microbiol*. 3:259.

138. Cell Walls and the Developmental Anatomy of *Brachypodium distachyon* Stem Biomass

Dominick A. Matos[1], Ian P. Whitney[1,*] (iwhitney@umass.edu), Michael J. Harrington[1], **Samuel P. Hazen[1]**

[1]Biology Department, University of Massachusetts Amherst

Project Goals: To understand the spatial and temporal deposition of cell wall polymers within stem tissue.

While many aspects of plant cell wall polymer structure are known, their spatial and temporal distribution within the stem are not well understood. Here, we studied vascular system and fiber development, which has implication for both biofuel feedstock conversion efficiency and crop yield. The subject of this study, *Brachypodium distachyon*, has emerged as a grass model for food and energy crop research. Here, we conducted our investigation using *B. distachyon* by applying various histological approaches and Fourier transform infrared spectroscopy to the stem internode from three key developmental stages. While vascular bundle size and number did not change over time, the size of the interfascicular region increased dramatically, as did cell wall thickness. We also describe internal stem internode anatomy and demonstrate that lignin deposition continues after crystalline cellulose and xylan accumulation ceases. The vascular bundle anatomy of *B. distachyon* appears to be highly similar to domesticated grasses. While the arrangement of bundles within the stem is highly variable across grasses, *B. distachyon* appears to be a suitable model for the rind of large C_4 grass crops. A better understanding of growth and various anatomical and cell wall features of *B. distachyon* will further our understanding of plant biomass accumulation processes.

This work was supported by the Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy grants DE-FG02-08ER64700DE and DE- SC0006621.

139. Systems Level Regulation of Rhythmic Growth Rate and Biomass Accumulation in Grasses

Samuel P Hazen[1,*] (hazen@bio.umass.edu), Benjamin J Cole[2], Dominick Matos[1], Ian Whitney[1], Kirk MacKinnon[1], **Steve A Kay**[2]

[1]Biology Department, University of Massachusetts, Amherst; [2]University of Southern California, Los Angeles.

Project Goals: The overall goal of this project is to understand the transcriptional regulation of biomass accumulation, which occurs during a specific time of day. Our aims are to (1) Identify growth-associated genes by temporal transcriptome analysis of developing *Brachypodium distachyon*. (2) Ascertain the time of maximum plant growth and cues that determine that timing through real-time imaging of *Brachypodium* growth. (3) Identify molecular regulatory circuits for biomass accumulation by screening for transcription factor interactions with promoters of growth-associated genes. (4) Determine the functional role of genes that modulate biomass accumulation.

Plant growth is commonly regulated by external cues such as light, temperature, water availability, and internal cues generated by the circadian clock. Changes in the rate of growth within the course of a day have been observed in the leaves, stems, and roots of numerous species. We examined the influence of diurnal temperature and light changes, and that of the circadian clock on leaf length growth patterns in *Brachypodium distachyon* using high-resolution time-lapse imaging. Pronounced changes in growth rate were observed under combined photocyles and thermocycles or with thermocycles alone. A considerably more rapid growth rate was observed at 28°C than 12°C, irrespective of the presence or absence of light. Despite evidence of circadian clock regulated gene expression, plants exhibited no change in growth rate under conditions of constant light and temperature or under photocycles alone. Therefore, temperature primarily drives oscillations in growth rate, which is not impacted by the circadian clock or by photoreceptor activity. Furthermore, temperature did not affect leaf meristem size or final cell length. Therefore, the nearly five-fold difference in growth rate observed across thermocycles can be attributed to proportionate changes in the rates of cell division and expansion. To better understand relative contributions of diurnal and circadian factors on gene expression changes, we analyzed the *B. distachyon* transcriptome in three conditions: photocyles and thermocycles, thermocycles alone, and constant light and temperature. Overall, we found ~9,000 genes exhibit cyclic expression in any one condition. Furthermore, far fewer genes are cyclic in free-running conditions in *B. distachyon* in comparison to *Arabidopsis thaliana*. Circadian clock regulated genes cycled with lower amplitude and longer periodicity, indicating a looser control of the clock on gene expression. Together, these data suggest fundamental differences between diurnal control of gene expression between cereals and dicots, especially with regard to temperature. Interestingly, the expression of cellulose and lignin genes was lowest in warm conditions; thus, negatively correlated with leaf elongation. Based on coexpression, we have identified candidate genes for the transcriptional regulation of temperature specific growth and biomass accumulation. A better understanding of the growth cues in *B. distachyon* will further our ability to model metabolism and biomass accumulation in grasses.

This work was supported by the Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy grant DE-SC0006621.

140. Creating a Multi-Functional Library of Grass Transcription Factors for the Energy Crop Model System *Brachypodium distachyon*

Joshua H Coomey[1,*] (jcoomey@cns.umass.edu), John Gierer[2], Samuel Deutsch[3], Steven A Kay[4],**Todd C Mockler[2], Samuel P Hazen[1]**

[1]University of Massachusetts Amherst; [2]Donald Danforth Plant Science Center, St. Louis; [3]DOE Joint Genome Institute, Walnut Creek, CA; [4]University of Southern California, Los Angeles

Project Goals: The overall goal of this project is to develop BradiTORFL, a comprehensive *Brachypodium distachyon* Transcription Factor ORF Library.

Comprehensive collections of full-length transcription factor cDNAs (fl-cDNA) have proven to be an extraordinary reagent for advanced research systems such as human, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Arabidopsis thaliana*. One of the seven DOE-JGI flagship plant genome species, *Brachypodium distachyon*, serves as a model for potential energy crops such as switchgrass, sorghum, and *Miscanthus*, as well as for the cereal crops that constitute a large part of the world's diet. We are constructing a complete grass transcription factor collection in an entry vector that will be of value for numerous functional studies. Even with a strong history of resource development that included numerous large-scale multinational projects, the *A. thaliana* collection lacks approximately 200 genes predicted to bind DNA. To overcome these same pitfalls and accelerate reagent development, we have included gene synthesis as a technique to capture fl-cDNAs. Through the analysis of several types of expression profiling data sets, we identified high priority candidates for the regulation of biofuel feedstock relevant traits, such as growth and cell wall biosynthesis and abiotic stress tolerance. The DOE-JGI synthesized 143 unique transcription factors from families that include bHLH, bZIP, CCAAT, GRAS, Homeodomain, HSF, MADS box, MYB, NAM, WRKY, and several classes of zinc fingers. Specific subfamilies include putative Aux/IAA-ARF auxin response factors, bHLH factors predicted to function in light signaling, G-box binding proteins implicated in light signaling, CBF-like genes predicted to function in abiotic stress responses, ethylene- associated factors, GRAS type genes implicated in growth and development, and MYBs predicted to function in cell wall, circadian clock and light signaling, and CCT domain containing genes associated with circadian clock and photoreceptor signaling. The collection will be transferred into multiple destination vectors for downstream applications including protein- DNA and protein-protein interaction platforms in yeast. Genes that yield positive interactions can then be shuttled from their pENTR vector to a variety of other constructs for various purposes including expression *in planta* to further characterize their functions. We are presently evaluating *B. distachyon* protein-DNA interactions in yeast using two approaches. The first is "gene-centric" where a promoter is tested for interactions with all transcription factor proteins in the library. The second approach, developed by the Mockler Lab, is "protein-centric" and interrogates the capacity of each transcription factor protein to interact with a collection of 768 synthetic 250 bp promoters. This collection of synthetic promoters was designed to maximize potential binding motif sequence diversity and all possible 8 nt DNA motifs occur in at least 4 independent promoters. Proof-of-concept experiments demonstrate the utility of this approach and we are currently expanding the analysis to infer the binding specificities for all of the transcription factors synthesized by DOE-JGI in this project.

This work was supported by the Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy grant DE-SC0006621 and the Joint Genome Institute Community Sequencing Program grants CSP-667 and CSP-1431.

141. BdMYB48 Regulates Biofuel Feedstock Attributes in the Model Grass *Brachypodium distachyon*

Pubudu P Handakumbura[1], Kathryn Brow[1], Ian P Whitney[1], Scott J Lee[1], Mike T Veling[1], **Samuel P Hazen**[1,*] (hazen@bio.umass.edu)

Biology Department, University of Massachusetts, Amherst

Project Goals: (1) Identify regulators of grass cell wall biosynthesis and biomass accumulation. (2) Analyze key transcription factor function and genomic binding sites. (3) Determine the effects of regulator perturbation on growth and amenability to deconstruction and cell wall properties.

Plant biomass offers a sustainable low cost alternative to fossil fuels and grasses such as miscanthus, sorghum, and switchgrass can provide ample biofuel feedstocks. While there is a need to better understand the molecular switches governing grass biomass accumulation, few have been characterized. Therefore, we conducted a screen for *Brachypodium distachyon* transcription factor proteins that interact with the regulatory regions of genes that encode cell wall biosynthetic enzymes. *B. distachyon* servers as a suitable model for energy crop research due to its close phylogeny to species such as switchgrass and miscanthus. In addition, it has many attributes characteristic of an ideal model organism, including a completely sequenced genome, self-compatible rapid life cycle, mutant collections, and genetic transformation. Seven of the 14 transcription factor proteins we observed that had a significant affinity with cell wall promoters were MYBs. Based on gene expression and amino acid homology, *BdMYB48* was selected for further characterization. The *BdMYB48* transcript is abundant in *B. distachyon* stem tissue, which accounts for a majority of above ground biomass. In order to functionally characterize *BdMYB48*, gain-of-function mutants (*Ubi::BdMYB48*) were generated by constitutively over expressing the full length coding region under the maize ubiquitin promoter. Similarly, dominant repressor mutants (*Ubi::BdMYB48:CRES*) were generated by over expressing the full-length coding region fused to a dominant repressor. Despite both types of transgenics having no significant changes in flowering time, gain-of-function lines had greater aboveground biomass accumulation while *Ubi::MYB48:CRES* plant biomass was reduced.

Interestingly, cellulose and lignin biosynthesis genes were significantly down regulated with a striking reduction in schlerenchyma fiber cell lignification in the *Ubi::BdMYB48:CRES* plants. Moreover, acetyl bromide soluble lignin content was significantly reduced in *BdMYB48:CRES* plants and modestly increased in *Ubi::BdMYB48* plants. Considering lignin is inversely correlated with bioconversion efficiency phenotypes, ethanol yield was measured after culturing stems with *Clostridium phytofermentans*. As expected, a decrease in ethanol yield was observe for *Ubi::BdMYB48* and a significant increase for *Ubi::BdMYB48:CRES* samples. Overall, these data suggest a cell type specific role for *BdMYB48* in *B. distachyon* secondary wall synthesis.

This work was supported by the Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy grants DE-FG02-08ER64700 and DE-SC0006641.

142. A Bacterial Role in Lignin Decomposition Under Future Rates of Nitrogen Deposition

Zachary Freedman[1]* (zacf@umich.edu) and **Donald R. Zak**[1,2]

[1]School of Natural Resources & Environment and [2]Department of Ecology and Evolution, University of Michigan, Ann Arbor, Michigan, USA 48109

Project Goals: With support from DoE BER, we have been able to study the molecular mechanisms by which experimental N deposition has decreased plant litter decay and increased the accumulation of organic matter in a northern hardwood forest ecosystem located in Michigan, USA. Experimental plots have been exposed of increased NO_3 deposition for 20 years. Using biogeochemical and molecular analyses, we have been able to test our initial hypotheses that ecosystem response to experimental N deposition (i.e., greater soil C storage) is governed by the environmental regulation of microbial gene expression. Here, we present findings that suggests lignin-modifying bacteria, which degrade lignin less efficiently than their fungal counterparts, may be favored under future rates of N deposition.

http://sitemaker.umich.edu/drzak/front_page

Anthropogenic release of biologically available nitrogen (N) has increased dramatically over the last 150 years, which can alter the processes controlling the storage of carbon (C) in terrestrial ecosystems. In a northern hardwood forest ecosystem located in Michigan, USA, nearly 20 years of experimentally increased atmospheric NO_3 deposition has reduced forest floor decay and increased soil C storage. This change occurred concomitantly with compositional changes in Basidiomycete fungi, and *Actinobacteria*, as well as the down-regulation of fungal lignocelluloytic genes.

Recently, laccase-like multicopper oxidases (LMCO) have been discovered among Bacteria. LMCOs participate in lignin decay, wherein lignin is depolymerized to dissolved organic carbon (DOC) with minimal CO_2 production. In this study, we examined how nearly two decades of experimental N deposition has affected the abundance and composition of lignolytic bacteria (*i.e.*, bacteria harboring LMCO genes).

In our experiment, bacterial LMCOs were more abundant in forest floor under experimental N deposition, whereas abundance of Bacteria and fungi were unchanged by this agent of global change. Experimental N deposition also led to less diverse, significantly different bacterial (16S rRNA) and LMCO gene assemblages, with known lignin-modifying bacterial taxa (i.e., Actinobacteria, Bacteroidetes, and Proteobacteria) accounting for the majority of compositional changes. These results suggest experimental N deposition favors bacteria in forest floor that harbor the LMCO gene and represents a plausible mechanism by which anthropogenic N deposition slows decomposition, increases soil C storage, and accelerates the leaching of DOC. In combination, our observations suggest future rates of atmospheric N deposition could fundamentally alter the physiological potential of soil microbial communities.

Funding through DOE's Office of Biological and Environmental Research program.

143. Spatially co-localizing the incompatible oxidative and enzymatic steps during fungal brown rot wood degradation - Early Career Program

Robert A. Blanchette,[1] Shona M. Duncan,[1] Timothy Filley,[2] Joel Jurgens,[1] Justin Kaffenberger[1], Jon Menke[1]* (menk0001@umn.edu), Gerald N. Presley, Jiwei Zhang, and **Jonathan S. Schilling**[1]

1University of Minnesota, Saint Paul; 2Purdue University, West Lafayette, IN

http://schillinglab.cfans.umn.edu

Project Goals: Wood-degrading fungal genomes are increasingly sequenced and annotated, including brown rot functional types starting with *Postia placenta* in 2009. Wood-degrading fungi metabolize wood either by removing lignin to access carbohydrates, or they extract carbohydrates without extracting significant amounts of lignin. The brown rot fungi evolved more recently and at least seven different times from ancestral white rot lineages, suggesting an advantage in efficiency. As the first annotated genome representative of 'typical' brown rot fungi, *P. placenta* is capable of an oxidative hydroxyl radical pretreatment that occurs concurrently with enzymatic saccharification of woody carbohydrates. This consolidation of otherwise incompatible reactions is fundamentally interesting and has great implication on the potential to consolidate harsh pretreatments with saccharification in a single processing step. Therefore, our research goals are as follows:

1) **physically sample wood degraded by the brown rot fungus *P. placenta* in order to map coincident pretreatment and saccharification reactions and to correlate relevant lignocellulose chemistry,**

2) **image pH and porosity at the fungus-plant interface and layer this data with images showing cellulase ingress, and**

3) **map, along the active hyphal front, the co-occurring expression of iron reductases associated with pretreatment and of cellulases used in saccharification.**

Abstract: Enzymatic bioconversion of lignocellulose plant tissues generally requires an initial pretreatment step, followed by saccharification and then fermentation or other downstream processing approaches. Consolidated bioprocessing (CBP) of lignocellulose combines enzymatic sugar release (saccharification) with fermentation, but pretreatments typically remain separate and costly. In nature, lignocellulose-degrading brown rot fungi consolidate pretreatment and saccharification, likely using spatial gradients to partition these incompatible reactions. Our goal is to characterize how this is achieved, in order to better understand the fungus and to potentially apply this approach in a mimicked consolidated approach.

The goal of this research is characterizing this relevant biological system, with objectives (stated above) to 1) physically sample wood degraded by the brown rot fungus *Postia placenta* to map reactions spatially and to correlate with cell wall modifications, 2) produce images of the environmental variables (pH and porosity) affecting cellulase ingress over time during brown rot, and map, along the active hyphal front, the co-occurring expression of iron reductases associated with pretreatment and of cellulase involved in saccharification. These are spatially-focused goals.

Therefore, my respective approaches involve either small-scale, spatially resolved characterization (Obj. 1), or appropriately resolved microscopy (Obj. 2 & 3).

To date, we have completed Objective 1 and published the findings in a single, large paper in *International Biodeterioration and Biodegradation* in 2013. Results show a spatially-segregated 'zone of interest' near the hyphal front, where depolymerization measured by alkali solubility of residues occurs ahead of both lignin modifications and active cellulase. These results include caveats on detection limits, etc., but suggest that Objectives 2 and 3, in progress, will be highly valuable. It also has shifted the hypotheses away from the assumed two-step staggering of reactions to instead assume that lignin may be modified by non-oxidative reactions and that cellulases may not be pivotal as the second step. At present, we have transformed *Piccia pastoris* to produce the brown rot cellulase PpCel5B for immunolabel, and will soon have imaged its ingress into wood cell walls at the resolution possible with the transmission electron microscope (TEM). This has been coupled with the Objective 3 effort to use fluorescence in situ hybridization (FISH) to co-localize expression of those genes assumed pivotal to brown rot, overlaying the depolymerization front we located in Objective 1. The result will be a full picture of temporal progression of brown rot, using a spatial gradient developed in wafers to do so.

References

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.

Funding statement: This work is supported by the Early Career Program and within the Office of Biological and Environmental Research in the DOE Office of Science.

144. The molecular basis for electron flow within metal-reducing biofilms: new insights from genome-scale genetics

Caleb Levar[1], Rachel Snider[1], Lori Zaccharof[1], Janet Rollefson[2], **Daniel R. Bond[1]***

*[1]Biotechnology Institute and Department of Microbiology University of Minnesota, Twin Cities;
[2]Department of Biochemistry, Biophysics, and Molecular Biology, University of Minnesota, Twin Cities.

Project Goals: Electrochemical, spectral, genetic, and biochemical techniques have provided evidence that multiple redox proteins and structural macromolecules outside the cell work together to move electrons long distances between *Geobacter* cells and to metals. This extracellular matrix contains many proteins that contribute to conductivity, in addition to complex polysaccharides and other extracellular macromolecules. Many of these components were likely lost or ignored in previous proteomic and biochemical surveys, and many of these proteins are only present under specific growth conditions. We aim to define this matrix, develop new tools discover the basis for its synthesis and construction, and visualize it in action. The goals of this project are to 1) identify elements crucial to the expression, assembly and function of the extracellular conductive matrix, 2) expand spectral and electrochemical techniques to define the mechanism and route of electron transfer through the matrix, and 3 combine this knowledge of electron transfer proteins and their role in multicellular electron transfer to visualize redox and gene expression gradients in space over time.

When bacteria change the state of metals in the environment, they transport electrons unprecedented distances from intracellular metabolic reactions through their membranes, and ultimately to distant mineral surfaces or other bacteria. This electron movement drives subsurface bioremediation, controls aquifer chemistry, and powers new microbial energy generation applications. Yet a molecular understanding of how this electron transfer is accomplished both at the interface and across long distances by *Geobacteraceae,* who are among of the most predominant bacteria in such systems, remains one of the grand challenges in microbial environmental processes. Unlike key processes spanning anaerobic redox gradients, such as methanogenesis, sulfate reduction, and nitrate reduction, metal reduction still lacks conserved redox proteins that can be used as indicators of function in metagenomic surveys.

In past years of this project, direct measurements of living biofilms using electrochemistry have revealed redox-potential dependent exchange between *Geobacter* redox proteins to be a rate-controlling step at all stages of growth. Direct spectral analysis of living biofilms confirmed that *c*-type cytochromes are a major reservoir of charge in these films, and that these cytochromes experience a bottleneck to oxidation when electrons must be transferred longer distances. Fine-scale immunogold labeling has discovered gradients in cytochrome abundance throughout these films, further suggesting the presence of redox and/or nutrient gradients within this biofilms, where cells distant from the electrode experience different conditions than those close to the acceptor.

Genetics has discovered separate polysaccharide biosynthesis operons, and secretion systems essential for the attachment of *Geobacter* to metals and other cells in the biofilm, indicating that cells must sense their substrate (metals, electrodes, other cells) and utilize different attachment systems accordingly. Genome sequencing of new *Geobacteraceae* representatives, such as those from alkaline environments, has refined our concept of 'core' cytochrome and metabolic genes that may be most essential to this process, and most useful in metagenomic studies.

A key finding of transposon-based genetic screens combined with proteomic analysis has been the separation of conductive biofilm development into stages. For example, disruption of type II secretion (*gsp*) proteins in *G. sulfurreducens* does not affect the cell's ability to attach to surfaces, and mutants transfer electrons to electrodes at wild type rates. However, once the electrode surface is covered, *G. sulfurreducens gsp*$^-$ mutants are unable to attach to each other, or form conductive biofilms, showing that proteins required for interfacial electron transfer can be separated from those required for long-range interactions. Mutants in multiheme *c*-type cytochromes, such as *omcS* and *pgcA,* are able to attach to surfaces and transfer electrons to electrodes at wild-type rates, and are still able to form thick interconnected biofilms, but fail to exhibit the cell-cell electron transfer needed to support growth distant from the electron acceptor. This indicates that cell-cell adhesion can be separated from conductivity. To more accurately measure electron transfer between cells, we have fabricated Interdigitated Electrode Assemblies, containing 10 µm electrodes separated by 15 µm gaps, and measured electron transfer at known redox potentials defined by biopotentiostats. In general, over a hundred mutants have been characterized to have defects in one of these 4 stages of biofilm development (attachment, electron transfer, cell-cell attachment, and cell-cell electron transfer), supporting the hypothesis that electron transfer across distances requires multiple cooperative adaptations.

Recently, we have adapted saturation mutagenesis and adaptive evolution approaches to demonstrate that the pathway of electron transfer out of *G. sulfurreducens* is dependent upon the electron acceptor in unexpected ways. For example strains of *Shewanella* possess only one mechanism for electrons to exit the quinone pool and enter metal- reduction pathways. However, sequencing of Tn-seq libraries grown with a variety of electron acceptors, combined with analysis of new mutants in putative inner membrane quinone oxidases, has shown that different respiratory pathways are involved in reduction of soluble acceptors and electrodes than insoluble Fe(III) oxides. This data led to the discovery that mutants in the highly conserved multiheme inner membrane protein *imcH* are unable to respire any soluble metal acceptor, or electrodes, but can reduce solid metals. Thus, we have constructed new saturation mutagenesis libraries in markerless *imcH* backgrounds to identify this second pathway. Complementary analyses with parallel evolved *imcH* suppressor mutants have also been resequened and SNPs enabling this second pathway identified with breseq. As we untangle pathways to different metals, new targets for expression and metagenomic analyses are emerging which could for the first time indicate the type of metal acceptor being used by *Geobacteraceae* in the environment.

This project is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

145. Regulation of Neutral Lipid Compartmentalization in Vegetative Plant Tissues

Kent D. Chapman[1] (PI, chapman@unt.edu), John M. Dyer[2] (Co-PI, john.dyer@ars.usda.gov), Robert T. Mullen[3]*(Co-PI, rtmullen@uoguelph.ca)

[1]University of North Texas, Center for Plant Lipid Research, Denton, TX; [2]USDA-ARS, US Arid-Land Agricultural Research Center, Maricopa, AZ; [3]University of Guelph, Department of Molecular and Cellular Biology, Guelph, Ontario, Canada.

Project Goal: The overall goal of our research program is to identify, characterize and manipulate the cellular machinery that influences the accumulation and compartmentalization of neutral lipids in vegetative tissues of plants.

Our research has focused on identifying the subcellular machinery that plants use to compartmentalize neutral lipids in cells of vegetative tissues. Virtually all cells in plants (and other organisms) synthesize triacylglycerols (TAGs) and deposit them in cytosolic lipid droplets (LDs), yet little is known about the molecular mechanisms responsible for the biogenesis and function of LDs, especially in cells of non-seed tissues. While seed tissues accumulate large amounts of TAGs in specialized, oleosin protein-coated LDs, most non-seed tissues do not. The energy density of TAGs is approximately twice that of carbohydrate on a mass basis, and as such, the seed oils of plants are a major source of calories for animal and human nutrition. By contrast, TAGs normally accumulate to a very small percentage in vegetative tissues like leaves and stems. In recent years, the use of increased amounts of seed TAGs as a bioenergy feedstock has placed energy and food demands at increasing conflict. Hence, the mechanisms in vegetative tissues that limit the accumulation TAGs represent key targets to expand the total TAG/energy content of plants, thereby reducing the competition between oilseeds for food and biomass for energy uses.

Our previously-funded research uncovered a gene in the plant *Arabidopsis thaliana* that, when disrupted, resulted in the hyperaccumulation of TAG-containing LDs in leaves. This gene is a homologue of human comparative gene identification-58 (CGI-58), which when mutated causes Chanarin-Dorfman syndrome, a neutral lipid storage disorder that also results in an increase in cellular accumulation of LDs in human tissues, particularly in cell types that typically lack high amounts of TAG such as skin and blood cells. Follow-up work this year with *Arabidopsis* has pointed to a central role for CGI-58 in regulating the turnover of cellular fatty acids, thus influencing the availability of lipids for membrane production, energy storage, and signaling pathways. We are now focused on other genes known to be involved in LD biogenesis in mammalian cells and testing their ability to modulate LD formation in *Arabidopsis*, including several endogenous *Arabidopsis* homologues of mammalian proteins (e.g., seipins), or other genes encoding foreign proteins that are not found in plants (e.g., the mammalian fat-inducing transmembrane protein 2, FIT2). Furthermore, our recent proteomics analysis of non-seed LDs, derived from avocado tissues, revealed a new class of LD-associated proteins that are highly conserved in plants, and we are currently investigating the ability of these proteins to promote the formation and/or regulation of LDs in leaves of *Arabidopsis*. Collectively, this research has important bioenergy applications, but also will help to unravel the complex machinery in eukaryotes that is deployed for the metabolism and maintenance of neutral lipids in LDs. Our results are also expected to stimulate new ideas about the dynamic interplay between lipid storage, mobilization and signaling in eukaryotic systems -- an understanding that will be important to achieving the broader BER goal of "sustainable and affordable production of renewable biofuels in an environmentally conscientious manner".

This research is supported by the U.S. Department of Energy, Office of Science, Division of Biological and Environmental Research (BER; DE-SC0000797).

Publications supported by award (2013):

1) **Chapman KD, Dyer JM, Mullen RT.** (2013) Commentary: why don't plant leaves get fat? *Plant Science* 207:128-34. doi: 10.1016/j.plantsci.2013.03.003.
2) **Park S, Gidda SK, James CN, Horn PJ, Khuu N, Seay DC, Keereetaweep J, Chapman KD, Mullen RT, Dyer JM.** (2013) The α/β hydrolase CGI-58 and peroxisomal transport protein PXA1 coregulate lipid homeostasis and signaling in *Arabidopsis*. *Plant Cell* 25(5):1726-39.
3) **Horn PJ, Silva JE, Anderson D, Fuchs J, Borisjuk L, Nazarenus TJ, Shulaev V, Cahoon EB, Chapman KD** (2013) Imaging heterogeneity of membrane and storage lipids in transgenic *Camelina sativa* seeds with altered fatty acid profiles. *Plant Journal*, Oct;76(1):138-50. doi: 10.1111/tpj.12278. Epub 2013 Aug 5.
4) **Horn PJ, James CN, Gidda SK, Kilaru A, Dyer JM, Mullen RT, Ohlrogge JB, Chapman KD.** (2013) Identification of a new class of lipid droplet-associated proteins in plants. *Plant Physiology*, 162(4):1926-36. doi: 10.1104/pp.113.222455.
5) **Gidda SK, Watt S, Collins-Silva J, Kilaru A, Arondel V, Yurchenko O, Horn PJ, James CN, Shintani D, Ohlrogge JB, Chapman KD, Mullen RT, Dyer JM** (2013) Lipid droplet-associated proteins (LDAPs) are involved in the compartmentalization of lipophilic compounds in plant cells. *Plant Signaling and Behavior*, 8(11). pii: e27141.
6) **Chapman, K.D., Dyer, J.M., and Mullen, R.T.** (2013). Deciphering the role of CGI-58 in lipid regulation: more than one way to trim the fat? *ASBMB Today* 12(11):26-27.
7) **Park S, Keereetaweep J, James CN, Gidda S, Chapman KD, Mullen RT, Dyer JM** (2013) CGI-58, a key regulator of lipid homeostasis and signaling in plants, also regulates polyamine metabolism. *Plant Signaling and Behavior, in press*.

146. Development of Crucial Tools for Lignin Research

Claudia Cardenas,[1] Utku Avci,[1] Yiki Tobimatsu,[2] Ruili Gao,[2] Alden Voelker,[2] Matt Regner,[2] Fachuang Lu,[2] **Fang Chen**,[3] **Steven Decker**,[4] **John Ralph**,[2] and **Michael G. Hahn**[1]* (hahn@ccrc.uga.edu)

[1]Complex Carbohydrate Research Center, University of Georgia, Athens, GA; [2]Department of Biochemistry and the Wisconsin Energy Institute, University of Wisconsin, Madison, WI; [3]Department of Biological Sciences, University of North Texas, Denton, TX; [4]National Renewable Energy Laboratory, Golden, CO

Project Goals: Research into aspects of lignification is currently stymied by lack of access to powerful modern methods to answer fundamental and practical questions. The three primary objectives are to: a) Develop a set of monoclonal antibodies to specific structures in lignins – for structural and localization studies; b) Develop a robust and flexible system for producing polymer-supported lignin monomers and oligolignols – for antibody screening, reactivity determination, elucidation of cross-coupling propensities, and beyond; and c) Develop fluorescent-tagged monolignols – to aid in lignin localization studies and to help elucidate monolignol transport mechanisms. These developments will provide researchers with an improved arsenal of tools to delve into today's most pressing and recalcitrant problems in lignin research, in projects ranging from very fundamental to those applied to improving pulp and papermaking or efficient biomass conversion to biofuels.

Abstract: We report here on progress in four goals for the project. First, in order to produce protein-bond phenolics for use in both raising antibodies and for screening, we have developed robust methodology for attaching phenolic compounds (including lignin models of interest here) to proteins (1). Along the way, we developed an improved method for derivatizing proteins such as the standard BSA as its cBSA (amines converted to acids with a small linker) derivative – cBSA is expensive commercially and only about 90% converted – we seem to be able to do it basically 100%, cleanly, and cheaply from BSA.

Second, a range of 'normal' and transgenic lignins with characteristics of interest have been isolated and used for both raising antibodies and for screening. A few of these will be noted below.

Third, we have synthesized and tested improved ('second-round') fluorescence-tagged monolignols, *p*-hydroxycinnamyl alcohols γ-linked to fluorogenic dyes such as aminocoumarin or nitrobenzofuran derivatives, as photoprobes for *in vitro* and *in vivo* studies of cell wall lignification. To illustrate the use of the probes for analysis of monolignol-protein interactions, we successfully monitored the complexation of aminocoumarin-tagged monolignols with horseradish apoperoxidase by Förster resonance energy transfer (FRET); the hindered binding affinity of this peroxidase- related protein towards syringyl substrates was first demonstrated experimentally by this method. In addition, to test imaging approaches, fluorescence-tagged monolignols were fed into various plant systems such as *Arabidopsis*, pine, and maize, and the localization of incorporated probes was readily visualized by fluorescence microscopy. The methods appear to be excellent for 'lighting up' actively lignifying zones, and appear to be useful in monolignol transport studies, e.g., using protoplasts. The paper was accepted as the featured cover article in *The Plant Journal* (2). [As it is now published, this aspect will not be highlighted on the poster here]

The main topic of this poster is to report on new lignin-directed monoclonal antibodies. Mice were immunized with solubilized purified lignin polymers from aspen/poplar, including from a high-syringyl transgenic line; the lignins are therefore referred to a S-L (a syringyl-only lignin) and GS-L (a guaiacyl/syringyl lignin as typically found in all monocots and dicots (including hardwoods such as

poplar, and *Arabidopsis*). Spenocytes harvested from the immunized mice were used to generate hybridoma lines that secrete S-lignin-directed monoclonal antibodies. Here we report the initial characterization of two of these antibodies, termed GLIMs for **G**eorgia **LI**gnin **M**onoclonals, namely GLIM6 and GLIM10. These antibodies were tested against several plant lignin preparations. Interestingly, GLIM6 recognizes a lignin substructure present in aspen and poplar, but not in corn, pine, or vanilla seed (a catechyl lignin) lignins. GLIM10 binds to aspen, poplar and corn lignin, suggesting a more ubiquitous S-lignin epitope is being recognized.

GLIM6 and GLIM10 were also used for immunolocalizations in w.t. and mutant *Arabidopsis* lines carrying mutations in the lignin biosynthetic pathway. GLIM6 immunoflorescence labeling was observed in *C4H::F5H* (S-enriched lignin phenotype), whereas no labeling was displayed on *fah1-2* (G-enriched lignin phenotype); on the other hand, GLIM10 recognizes a lignin substructure present in both G- and S-enriched lignin phenotypes. GLIM6 and GLIM10 also exhibit different labeling patterns on inter-fascicular fibers and vascular bundles as a function of plant development. Lastly, the effects of enzyme and alkaline pretreatment of the sections on GLIM6 and GLIM10 labeling patterns were studied. GLIM6 and GLIM10 labeling intensities increase after xylanase M1, pectin methylesterase or alkaline treatment, but no effect was noticed after polygalacturonase I and II treatments. The immunolocalization patterns observed with GLIM6 and GLIM10 show some striking (and never before noted) features that will require explanation and additional research.

References

1. Gao, R., Zhu, Y., Hahn, M. G., and Ralph, J. Flexible method for conjugation of lignin model compounds to carrier proteins. (2014) *J. Agr. Food Chem.*, submitted
2. Tobimatsu, Y., Wagner, A., Donaldson, L., Mitra, P., Niculaes, C., Dima, O., Kim, J.-I., Anderson, N. A., Loque, D., Boerjan, W., Chapple, C., and Ralph, J. Visualization of plant cell wall lignification using fluorescence-tagged monolignols. (2013) *Plant J.* **76**(3), 357-366 [COVER ARTICLE]

This work was funded by the US Department of Energy, the Office of Science (DE- SC0006930).

147. Fluorescent probe selective for hydrogen atom abstraction based on polyethylene glycol cleavage

Vitaliy Timokhin[1], Sarah Baumann[2], Peter Kitin[2], **Carl Houtman[3]**, John Ralph[1], **Kenneth E. Hammel[2, 3]**, **Christopher G. Hunt[3]*** (cghunt@fs.fed.us)

1 Great Lakes Bioenergy Research Center, Madison, WI; 2 Dept of Bacteriology, University of Wisconsin- Madison, 3 USDA Forest Products Laboratory, Madison WI.

http://www.bact.wisc.edu/faculty.php?init=KEH&show=PEP

Project Goals:

Problem. Few organisms can break down lignocellulose, the primary structural component of terrestrial plants, and those with this ability employ unique, efficient biochemical mechanisms to deconstruct this polymeric composite. These mechanisms are relevant to US national energy needs because they may inspire the design of new technologies for biofuels production from abundant forms of lignocellulosic biomass such as wood. Research has shown that the only efficient lignocellulose degraders in nature are certain filamentous fungi that inhabit plant litter and wood, and that these fungi initiate lignocellulose deconstruction by disrupting the lignin that otherwise serves as a barrier to enzymes that can release the substrate's energy-rich sugars. There is also evidence that the agents responsible for lignin disruption are small molecules that oxidize this recalcitrant polymer. However, little is known about the specific identity of these oxidants, their quantity, or the processes that fungi use to generate them.

Approach. This project has the following goals and methods:

1) Develop new chemical sensors that resemble lignin structurally and are immobilized on small micrometer-scale beads that can be added to specimens of wood. When these sensors encounter a lignin-degrading oxidant, they will be cleaved to produce a fluorescent moiety that can be detected by fluorescence microscopy.
2) Develop methods to grow lignocellulose-degrading fungi on wood specimens that contain sensor beads.
3) Use fluorescence microscopy, in conjunction with these sensors, to quantify oxidants and obtain spatial maps of their distribution in wood undergoing deconstruction by several fungi whose genomes have been sequenced under the auspices of the US Department of Energy because they are candidates for bioenergy research. (1)
4) Develop mathematical models that account for the extents and spatial distributions of lignin-degrading oxidants in the biodegrading lignocellulosic specimens. This approach is expected to reveal which fungi employ the most efficient deconstruction mechanisms. (2)
5) Identify likely processes for lignin disruption by correlating the expression of fungal genes that may encode oxidant production with the oxidation of bead-linked sensors in the wood. This approach is expected to pinpoint fungal enzymes and metabolites that have key roles in disruption of the lignin barrier.

Impact. This research on biological processes involved in the efficient natural deconstruction of lignocellulose is likely to identify fungal enzymes and metabolites that have potential utility in biofuels development.

Abstract

In nature, basidiomycete fungi are the major agents of lignocellulose recycling, using enzymes to produce reactive oxygen species that in many cases operate by hydrogen atom abstraction. In an attempt to understand the natural deconstruction of lignocellulose, we have developed a probe specific for this reaction. This probe carries two fluorescent dyes linked by a short polyethylene glycol (PEG) chain. Fluorescence FRET between these two dyes is lost when the PEG is depolymerized by a hydrogen-abstracting oxidant. That is, cleavage is observable by a loss of FRET between a fluorescent donor and acceptor on either end of the PEG molecule. The activity of these strong hydrogen abstracting species can thus be mapped by visualizing the loss of FRET. This work presents data on the basic structure of the probe and its response to various oxidants, along with preliminary data on its ability to report oxidative activity in real culture systems.

We have shown by NMR analysis that PEG is cleaved by H-abstraction oxidants. To capitalize on this observation, we have placed fluorescein as a FRET donor and TAMRA as an acceptor on either end of a PEG with with a degree of polymerization of 7, which separates the dyes at most by 4.2 nm, which is within the FRET radius of this pair. We have further shown that FRET occurs in solution in this system and that oxidation reduces the occurrence of FRET, resulting in an increase in the donor signal relative to the acceptor. Finally, we have preliminary data showing a decline in FRET after this probe was placed on wood and exposed to the white rot fungus *Phanerochaete chrysosporium* for eight days. These results suggest that our system is capable of reporting oxidative activity on real lignocellulosic substrates.

Areas of improvement that we will explore for oxidation probes of this type are better pH stability, photostability and oxidation stability, as well as better solubility and a longer FRET radius for the donor and acceptor fluorophores. We are also attempting to anchor the probe to a solid substrate (3-µm porous silica beads) to prevent diffusion of the probe or its ingestion by the fungi, which should allow more reliable measurements of the spatial distribution of oxidants on lignocellulose.

References

1. Kitin, P.; Hunt, C. G.; Houtman, C.; Timokhin, V.; Hammel, K. E. submitted, Estimation of diffusible extracellular oxidants produced by three disparate lignonolytic fungi. Fungal Genetics and Biology.
2. Hunt, C. G.; Houtman, C. J.; Jones, D. C.; Kitin, P.; Korripally, P.; Hammel, K. E. 2013, Spatial mapping of extracellular oxidant production by a white rot basidiomycete on wood reveals details of ligninolytic mechanism. Environmental Microbiology 15: 956-966.

This work was supported by grant DE-SC0006929 (K.E.H, C.G.H., and C.J.H.) from the U.S. Department of Energy, Office of Biological and Environmental Research.

148. Use of Systems Biology Approaches to Develop Advanced Biofuel-Synthesizing Cyanobacterial Strains

Thomas J. Mueller[1], Bertram M. Berla[2] (bertramberla@wustl.edu), Jingjie Yu[3], Ronald Stebegg[3], Louis A. Sherman[4], Costas D. Maranas[1] and **Himadri B. Pakrasi**[2,3]

[1]Department of Chemical Engineering, Pennsylvania State University, University Park, PA 16802; [2]Department of Energy, Environmental, and Chemical Engineering, Washington University, St. Louis, MO 63130; [3]Department of Biology, Washington University, St. Louis, MO 63130; [4]Department of Biological Sciences, Purdue University, West Lafayette, IN 47907

http://maranas.che.psu.edu/
http://pages.wustl.edu/photo.synth.bio

Project Goals: This project aims to develop new tools for biofuel production in photosynthetic bacterial hosts. Our studies include the identification of a novel, fast- growing, mixotrophic, transformable cyanobacterium. This strain has been sequenced and will be made available to the community. In addition, we have developed genome-scale models for a family of cyanobacteria to assess their metabolic repertoire. Furthermore, we developed a method for rapid construction of metabolic models using multiple annotation sources and a metabolic model of a related organism. This method will allow rapid annotation and screening of potential phenotypes based on the newly available genome sequences of many organisms.

A new, fast-growing, mixotrophic, transformable cyanobacterium

Photosynthetic microbes are of considerable interest in biotechnological applications due to their ability to use sunlight to convert CO_2 into fuels and other useful chemicals. Important and desirable traits for production organisms include fast growth and amenability to genetic manipulation. Synechococcus elongatus UTEX 2973, a novel cyanobacterial strain showed rapid growth rates, with a doubling time of less than two hours under optimal conditions, while other commonly used cyanobacteria (Synechocystis sp. PCC 6803, Synechococcus sp. PCC 7942 and Synechococcus sp. PCC 7002) in biofuel production have considerably longer doubling times. We have determined that this strain can grow mixo- and photoheterotrophically in the presence of 30 mM fructose. Synechococcus elongatus UTEX 2973 can be readily transformed by conjugation and fully segregated mutants can be generated more quickly than with other model cyanobacteria such as Synechocystis sp. PCC 6803.

The genome of this strain was sequenced and compared to its close relatives Synechococcus elongatus PCC 6301 and PCC 7942. Significant differences were found, including a large deleted genomic region that may be related to the unique attributes of this strain. Future aims include using Synechococcus UTEX 2973 for long-chain alkane overproduction and gene function studies, as well as developing a genome-scale metabolic model for this organism.

Genome-scale model development

Genome-scale models allow researchers to both analyze an organism's metabolism and make predictions about how genetic engineering might change that metabolism. However, the reconstruction of quality genome-scale metabolic models of organisms with limited annotation resources remains a challenge that often requires a time-consuming manual approach. To mitigate this challenge, we developed a workflow that combines annotation information from multiple sources: the Universal

Protein Resource (Uniprot); NCBI Protein Clusters; Rapid Annotations using Subsystems Technology (RAST); and a previously developed reference model for the cyanobacterial genus Cyanothece (the iCyt773 model for Cyanothece ATCC 51142) to create genome-scale metabolic reconstructions of new sequenced strains with limited manual effort. Models were created for five Cyanothece strains, namely Cyanothece sp. PCC 7424, 7425, 7822, 8801 and 8802. All five models include fully traced photosynthesis reactions and respiratory chains, as well as mass and charge-balanced reactions and gene-protein-reaction (GPR) associations. Meeting these stringent criteria for model quality makes the models far more useful for phenotype prediction and for guiding metabolic engineering.

Upon examination, the reactions shared between these five models match the known phylogenetic relationships between the organisms. These models also allow for the assessment of the bio-production potential of the modeled species. The non-fermentative pathway for alcohol production is found only in Cyanothece 7424, 8801, and 8802, while the fermentative pathway for butanol production exists in varying levels of completion within the five models. The models also highlight other metabolic differences, such as in arginine catabolism.

The workflow that we have developed expedites construction of curated metabolic models for organisms that, while not yet developed as model systems, have sequenced genomes, reviewed gene annotations, and are related to an organism with a curated metabolic model. Models created from this workflow can be used to develop strategies for targeted metabolite overproduction or to gain insight into the metabolic differences between organisms.

Alkane production by Synechocystis 6803
Nearly all known cyanobacteria produce n-alkanes in the C_{17} range. While initial reports of these compounds date back to the 1960's, the genes responsible were only identified recently, and the biological function of these compounds remains a mystery. Although these compounds are produced at relatively low levels by biotechnology standards (~0.1% dw), they make up a significant portion of the cell membranes of cyanobacteria, and are produced during all phases of cell growth and under a wide variety of growth conditions at relatively constant cellular concentration, suggesting that they play a critical role in normal cellular function.

Supported by funding from the Office of Science (BER), U. S. Department of Energy

Publications

1. Saha, R., Verseput, A. T., Berla, B. M., Mueller, T. M., Pakrasi, H. B. and Maranas, C. D. (2012) Reconstruction and comparison of the metabolic potential of cyanobacteria *Cyanothece* sp. ATCC 51142 and *Synechocystis* sp. PCC 6803. PLOS One 7: e48285.
2. Berla, B. M. and Pakrasi, H. B. (2012) Up-regulation of plasmid-encoded genes during stationary phase in *Synechocystis* sp. PCC 6803, a cyanobacterium. Appl. Environ. Microbiol., 78: 5448-5451.
3. Landry, B. P., Stöckel, J. and Pakrasi, H. B. (2013) Use of degradation tags to control protein levels in the cyanobacterium *Synechocystis* sp. PCC 6803. Appl. Environ. Microbiol., 79: 2833-2835.
4. Mueller, T. J., Berla, B. M., Pakrasi, H. B. and Maranas, C. D. (2013) Rapid construction of metabolic models for a family of Cyanobacteria using a multiple source annotation workflow. BMC Systems Biology, 7(1): 142.

149. Determining Eukaryotic Microalgal TAG Yield Using a Commercial Serum Triglyceride Determination Kit

Carrie Goodson,[1] Jan Jaworski,[2] Jia Li,[2] Taylor Weiss,[1] Tuya Wulan,[1] and **Ursula Goodenough**[1]***** (goodenough@wustl.edu),

[1]Dept. Biology, Washington University, St. Louis MO; [2]Danforth Plant Science Center, St. Louis MO

Project Goals: In response to stress, eukaryotic microalgae produce triacylglycerides (TAGs) that are stored in lipid bodies; the TAG is an excellent substrate for biodiesel production. Assays that have been developed for assessing microalgal TAG content are time-consuming and require equipment and technical expertise that is often not available to groups interested in pursuing the development of this technology. We have optimized a rapid and quantitative colorimetric TAG assay for two microalgae, the green *Chlamydomonas reinhardtii* and the red *Cyanidioschyzon merolae*, using an inexpensive and commercially available kit developed to monitor serum TAG levels in human blood samples.

TAG quantitation in *C. reinhardtii* commonly utilizes a version of the procedure developed in the Benning lab (described in detail in 1): cells are pelleted and extracted with organic solvents; the extract is run on thin-layer chromatography plates; the TAG band is scraped off and converted to fatty-acid methyl esters (FAMEs); and the FAMEs are quantitated using mass-spec (GC-MS) or flame ionization detection (GC-FID) gas chromatography, with peaks then integrated to calculate total TAG yield. An alternative approach, developed in the Hildebrand lab for diatoms (2,3), quantitates the signal from BODIPY 493/503 (Molecular Probes), which fluoresces in a neutral-lipid environment (but see refs xy), using an imaging flow cytometer. The first method has been described as "tedious and time consuming" (4); the second requires determination that, for a given microalgal strain, only TAG is BODIPY-positive; and both require sophisticated equipment operated by experienced users. Research labs without access to such resources usually report their results using versions of Bligh and Dyer assays (5) that measure both polar and neutral lipids and hence give no information about TAG content.

Commercially available kits allow the performance of a colorimetric assay, using a standard bench-top spectrophotometer, that measures the glycerol released after TAGs are digested with TAG lipases. We previously (6) quantitated the TAG content of purified lipid bodies from *C. reinhardtii* using such a kit, and obtained values very similar to those obtained using GC-MS. We report here the use of such a kit to evaluate the TAG content of whole cells. In optimizing the assay, we used *C. reinhardtii* and *C. merolae*, which have very different pigment profiles, and heeded the potential artifacts in the assay described in a paper on *Drosophila* (7).

Take it from here!

1. Boyle, N.R., M.D. Page, B. Liu, I.K. Blaby, D. Casero, J. Kropat, S. Cokus, A. Hong-Hermesdorf, J. Shaw, S.J. Karpowicz, S. Gallaher, S. Johnson, C. Benning, M. Pellegrini, Grossman, and S.S. Merchant. 2012. Three acyltransferases and a nitrogen responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*. J. Biol. Chem. **287**: 15811-15825.
2. Traller, J.C., and M. Hildebrand. 2013. High throughput imaging to the diatom *Cyclotella cruptica* demonstrates substantial cell-to-cell variability in the rate and extent of triacylglycerol accumulation. Algal Res. **2**: 244-252.

3. Trentacoste, E.M.m R.P. Shrestha, S.R. Smith, C. Glé, A.C. Hartmann, M. Hildebrand, and W.H. Gerwick. 2013. Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. Proc. Natl. Acad. Sci. **110**: 19748- 19753.
4. Siaut, M., S. Cuiné, C. Cagnon, B. Fessler, M. Nguyen, P. Carrier, A. Beyly, F. Beisson, C. Triantaphylidès, Y. Li-Beisson, and G. Peltier. 2011. Oil accumulation in the model green alga *Chlamydomonas reinhardtii*: Characterization, variability between common laboratory strains and relationship with starch reserves. BMC Biotechnol. **11**: 7**.**
5. Bligh, E.G. and W.J. Dyer. 1959. A rapid method for total lipid extraction and purification. Can. J. Biochem. Physiol. **37**: 911-917.
6. Wang, Z.T., N. Ullrich, S. Joo, S. Waffenschmidt, and U. Goodenough. 2009. Algal lipid bodies: stress induction, purification, and biochemical characterization in wild-type and starchless *Chlamydomonas reinhardtii*. Eukaryot. Cell **8**: 1856-1868.
7. Al-Anzi, B., and K. Zinn. 2010. Colorimetric measurement of triglycerides cannot provide an accurate measure of stored fat content in *Drosophila*. PLoS One **5**: e12353.

Supported by grant DE-SC0006873 from the DOE Office of Science (BER) to Ursula Goodenough.

150. Ultrastructure and Composition of the *Nannochloropsis gaditana* Cell Wall

Matthew J. Scholz[1], Taylor L. Weiss[2], Jia Jing[3], Robyn Roth[4], **Ursula Goodenough[2]*** (goodenough@wustl.edu), Matthew C. Posewitz[1], and Henri G. Gerken[5]

[1]Department of Chemistry and Geochemistry, Colorado School of Mines, Golden, CO; [2]Department of Biology, Washington University, St. Louis, MO; [3]CAS Key Laboratory of Biofuels, Shandong Key Laboratory of Energy Genetics and BioEnergy Genome Center, Qindao Institute of BioEnergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong China; [4]Department of Cell Biology and Physiology, Washington University School of Medicine, St. Louis, MO; [5]Arizona Center for Algal Technology and Innovation, Arizona State University, Mesa, Arizona

Project Goals: Some of the oleaginous eukaryotic microalgae that best thrive in open-pond growth facilities, like the chlorophyte *Scenedesmus* and the eustigmatophyte *Nannochloropsis*, are endowed with "recalcitrant" cell walls that presumably confer resistance to predation and desiccation during growth but are refractory to breakage and drying and hence to product extraction—notably the extraction of triacylglyceride (TAG). These walls have been shown to contain cross- linked long-chain hydrocarbons, generically called "algaenans," that defy solubilization in the laboratory. We report a comprehensive analysis of *N. gaditana* wall ultrastructure and composition, documenting that its two major components are algaenan and cellulose, and present the first spectral profiles of non-denatured algaenan using attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy. We also identify and characterize *N. gaditana* genes encoding proteins that may participate in the biosynthesis and secretion of these wall components— genes that might be engineered such that expression conferring robustness during growth is switched off at time of harvest.

Quick-freeze deep-etch electron micrographs of native walls or shed mother walls of *N. gaditana* document a two-layered structure (Fig. 1). When walls are instead prepared from linear-phase cells subjected to multiple passes through a French Press, layer 2 is converted from a dense to a loose configuration. Whereas the dense configuration is refractory to cellulase digestion, the loose configuration is completely sensitive, generating pure preparations of layer 1 that have not been subjected to any denaturants. Biochemical and spectral analyses of non-digested and digested wall preparations document that layer 1 is dominated by algaenan, forming a seal around the cell, and that layer 2 is dominated by cellulose. We hypothesize that the native dense layer-2 configuration represents a form of crystalline cellulose that is resistant to most cellulases and provides additional hardiness to the organism.

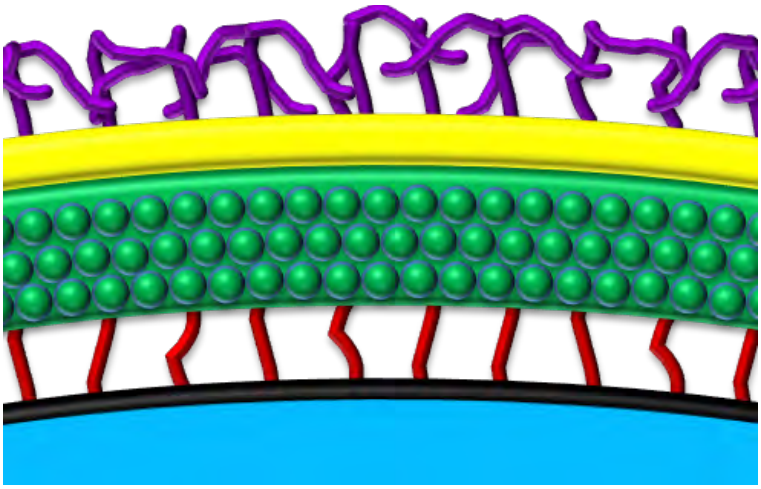


Figure 1

Extensions
Layer 1 (algaenan-based)
Layer 2 (cellulose-based)
Struts
Plasma membrane
Cytoplasm

Spectra of non-denatured layer 1 are consistent with previous analyses of denatured material (1, 2) which concluded that *Nannochloropsis* algaenan consists of straight-chain, aliphatic hydrocarbons that are cross-linked via ether bridging and contain few, if any, branched methyl groups. This is consistent with algaenan synthesis proceeding by the crosslinking of fatty alcohols/aldehydes/acids synthesized by either fatty acid synthases (FASs) or polyketide synthases (PKSs), rather than via terpenoid biosynthetic pathways.

Inspection of the *N. gaditana* genome identified 6 putative polyketide synthase (*PKS*) genes, and domains encoding components of a type II FAS complex. Five of these *PKS* genes are conserved in the *N. oceanica* genome. The *PKS1* gene contains a thioesterase module, indicating free fatty acids as the final product, whereas the other 5 genes do not encode this module. The *PKS4-6* genes encode a fatty acyl-reductase (FAR) domain instead of a thioesterase domain, indicating that fatty aldehydes—or potentially fatty alcohols if a four-electron reduction is catalyzed—are released from the acyl-ACP (acyl- carrier protein) rather than fatty acids. The *PKS2/3* genes contain terminal non-ribosomal peptide synthetase (NRPS) domains; these participate in a myriad of reactions wherein a diversity of nucleophiles, including amino acids, can attack the thioester linkage on ACP to establish a new covalent linkage with the acyl group formerly bound to the PKS ACP. The lack of dehydratase (DH) and enoyl reductase (ER) domains in *PKS2/3* suggests that internal alcohols are retained in the products of these enzymes after keto reductase (KR) reduction, at the β-position of the C_2 extension(s), to the fatty acids that are loaded onto these PKSs by fatty acyl-AMP ligase (FAAL).

Intriguingly, the NRPS modules in *PKS2/3* contain predicted membrane-spanning domains that could allow a growing algaenan chain to exit on the extracellular side of the plasma membrane as fatty acyl groups are polymerized, which could explain how intracellular metabolic precursors are polymerized into longer-chain algaenan precursors that are deposited on the cell exterior. Additionally, *PKS1*, *3*, and *4* are adjacent to genes encoding ABC transporters (pathway genes are frequently clustered in the *Nannochloropsis* genome), which may also play a role in the export of some lipophilic algaenan precursors.

In sum, 6 PKSs have been identified that are promising candidates for roles in algaenan synthesis, and putative mechanisms for the secretion of algaenan precursors have been identified. Future experiments will explore the effects of knocking out or knocking down one or more of these genes. If the mutants display compromised wall structure, then we will attempt to construct strains wherein the gene(s) are expressed during growth but blocked near the time of harvest.

1. Gelin, F., I. Boogers, A.A.M. Noordeloos, J.S.S. Damste, R. Riegman, and J.W. De Leeuw (1997). Resistant biomacromolecules in marine microalgae of the classes eustigmatophyceae and chlorophyceae: Geochemical implications. *Organic Geochemistry*, **26**, 659-675.

2. Gelin, F., J.K. Volkman, C. Largeau, S. Derenne, J.S.S. Damste, and J.W. De Leeuw (1999). Distribution of aliphatic, nonhydrolyzable biopolymers in marine microalgae. *Organic Geochemistry*, **30**, 147-159.

Supported by grant DE-SC0006873 from the DOE Office of Science (BER) to U. G., funding from the Colorado Energy Research Collaboratory to M.C. P., and a New American University Fellowship at ASU AzCATI to H. G.

151. The Path to Triacylglycerol (TAG) Obesity in the *sta6* Strain of *Chlamydomonas reinhardtii*

Ursula Goodenough,[1*] (goodenough@wustl.edu) Ian Blaby,[2] David Casero,[3,4,a], Sean D. Gallaher,[2] Carrie Goodson,[1] Shannon Johnson,[5] Jae-Hyeok Lee,[6] Sabeeha S. Merchant,[2,4] Matteo Pellegrini,[3,4] Robyn Roth,[7] Jannette Rusch,[1] Manmilan Singh,[8] James G. Umen,[9] Taylor L. Weiss,[1] and Tuya Wulan[1]

[1]Department of Biology, Washington University, St. Louis MO; [2]Department of Chemistry and Biochemistry, University of California, Los Angeles CA; [3]Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles CA; [4]Institute of Genomics and Proteomics, University of California, Los Angeles CA; [5]Genome Science, Los Alamos National Laboratory, Los Alamos, New Mexico; [6]Department of Botany, University of British Columbia, Vancouver, BC, Canada; [7]Department of Cell Biology and Physiology, Washington University School of Medicine St. Louis MO; [8]Department of Chemistry, Washington University, St. Louis MO; [9]Danforth Plant Science Center, St. Louis MO 63132

[a]Current address: Department of Pathology and Laboratory Medicine, University of California, Los Angeles, CA 90095.

Project Goals: When the *sta6* (starch-null) strain of the green microalga *Chlamydomonas reinhardtii* is nitrogen-starved in acetate and then "boosted" after 2 days with additional acetate, the cells become "obese" after 8 days, with TAG-filled lipid bodies filling their cytoplasm and chloroplasts (1). To assess the transcriptional correlates of this response, *sta6* and the starch-forming *cw15* strain were subjected to RNA-Seq analysis during the 2 days prior and 2 days post boost. During the 2 hr post boost, ~ 425 genes are up-regulated ≥2-fold and ~875 genes are down-regulated ≥2-fold in each strain. Our results indicate that the boost serves both to avert an autophagy program and to prolong the operation of key biochemical pathways that conserve nitrogen and shuttle carbon from acetate into storage lipid, the outcome being enhanced TAG accumulation, notably in *sta6*. Four genes -- encoding a diacylglycerol acyltransferase (*DGTT2*), a glycerol-3-P dehydrogenase (*GPD3*), and two candidate lipases (Cre03.g155250 and Cre17.g735600) - are selectively up- regulated in *sta6*, and are therefore candidates for future genetic engineering.

The Merchant/Pellegrini and Los Alamos laboratories recently generated and analyzed RNA-Seq transcriptomes of *cw15*, *sta6*, and several complemented *sta6* strains during two days of N-starvation (0→48h-N) (2). In collaboration with these groups, the Goodenough lab generated a second pair of transcriptomes using *cw15* and *sta6*, tracing 0→48h-N gene expression patterns under a different set of culture conditions and taking the time course out to 96h-N, with an intervening acetate boost. Analysis of these data was deeply informed by cross-comparisons with the Blaby et al. (2) data and the Boyle et al. data (3) on an N-starved wild-type strain.

By consolidating these data, it has been possible to identify "robust" biochemical pathways, like starch, fatty-acid, and TAG biosynthesis, wherein patterns of expression of the relevant genes are largely concordant regardless of genetic background or culture conditions, thereby calling attention to the few exceptional cases. Also identified are 21 "sensitive" genes, encoding products operating in several pathways, including the glyoxylate and Calvin Benson cycles, gluconeogenesis, and the pentose phosphate pathway, that are influenced by on-going carbon flux; their expression is coordinated but varies within strains and between conditions, suggesting that they play a role in monitoring and responding to N-depletion in particular biosynthetic/metabolic contexts. Thirteen of these "sensitive" genes are strongly responsive to the cell's acetate status.

The bulk rate of acetate depletion from the medium is not boost-enhanced, but evidence for a spike in acetate uptake is presented, and three candidate acetate permease-encoding genes in the *GPR1_FUN34_YaaH* superfamily are strongly boost-up-regulated.

A cohort of 64 autophagy-related genes is down-regulated by boost. We propose that this is linked to microscopic observations showing that non-boosted cells initiate an autophagocytic response at 48h-N, accompanied by diminished TAG accumulation, that is not initiated in boosted cells.

The four genes whose expression is specifically enhanced in *sta6* encode enzymes expected to play a role in lipid-body formation, where one of the candidate lipases -- Cre03.g155250 – is homologous to *PGD1*, recently shown to participate in TAG biosynthesis (4). Whether they specifically operate in the chloroplast to form chloroplast lipid bodies will be the subject of future experimentation. We further propose that the disruption of starch synthesis in *sta6* creates a glucose-6-P "backflow" that feeds into chloroplast lipid-body formation.

1. Goodson, C., R. Roth, Z.T. Wang, and U. Goodenough. 2011. Structural correlates of cytoplasmic and chloroplast lipid body synthesis in *Chlamydomonas reinhardtii* and stimulation of lipid body production with acetate boost. Eukaryot. Cell **10**: 1592-1606.

2. Blaby, I.K., A.G. Glaesener, T. Mettler, S.T. Fitz-gibbon, S.D. Gallaher, B. Liu, N.R. Boyle, J. Kropat, M. Stitt, S. Johnson, C. Benning, M. Pellegrini, D. Casero, and S.S. Merchant. 2013. Systems-level analysis of nitrogen starvation-induced modifications of carbon metabolism in a *Chlamydomonas reinhardtii* starchless mutant. Plant Cell: (online publication).

3. Boyle, N.R., M.D. Page, B. Liu, I.K. Blaby, D. Casero, J. Kropat, S. Cokus, A. Hong- Hermesdorf, J. Shaw, S.J. Karpowicz, S. Gallaher, S. Johnson, C. Benning, M. Pellegrini, Grossman, and S.S. Merchant. 2012. Three acyltransferases and a nitrogen responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*. J. Biol. Chem. **287**: 15811-15825.

4. Li, X., E.R. Moellering, B. Liu, C. Johnny, M. Fedewa, B.B. Sears, M.-Hao Kuo, and C. Benning. 2012. A galactoglycerolipid lipase is required for triacylglycerol accumulation and survival following nitrogen deprivation in *Chlamydomonas reinhardtii*. Plant Cell **24**: 4670-4686.

Supported by Contract DE-EE0003046 (to U.G., S.J., S.S.M., and M.P.) via the National Alliance for Advanced Biofuels and Bioproducts (NAABB) from the U.S. Department of Energy (DOE) and by Cooperative Agreement No. DE-FC02-02ER6342 from the DOE Office of Science (BER) to the Institute of Genomics and Proteomics, UCLA. Support to U.G. from grant DE-SC0006873 from the DOE Office of Science (BER); to J-H.L. from grant 2013M1A8A1056300 from the Korea CCS R&D Center (KCRC), Korean Ministry of Science; to I.K.B. from training grant T32 ES015457 from the National Institutes of Health; and to S.D.G. and S.S.M. from grant R24 GM092473 from the National Institutes of Health.

152. The Multi-Center Iron Respiratory Chain of *Acidithiobacillus ferrooxidans* Functions as an Ensemble With a Single Macroscopic Redox Potential

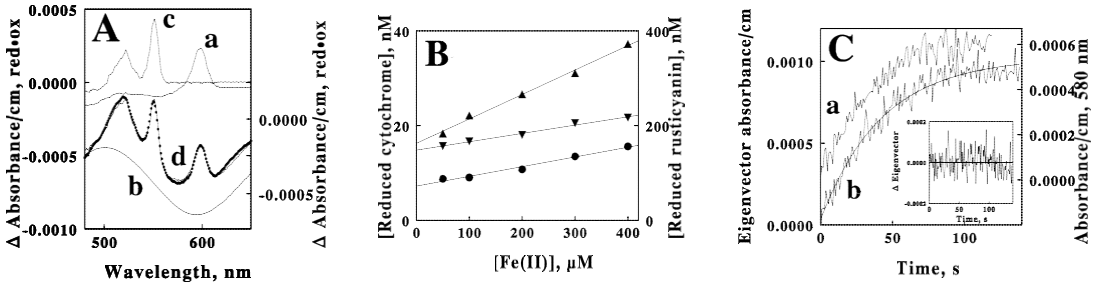
Robert Blake II,[1] *(rblake@xula.edu), TingFeng Li, [1] Bhupal Ban,[1] and Richard Painter[1] [1]Xavier University of Louisiana, New Orleans, LA

URL: www.xula.edu

Project Goals: The practical goal is to develop an integrating cavity absorption meter (ICAM) to directly observe respiratory electron transfer reactions in live bacteria as they exchange electrons with soluble or insoluble iron under physiological conditions. The premise is that accurate UV-visible spectroscopy of electron transfer reactions among colored biomolecules can be conducted in highly turbid suspensions if the live bacteria are irradiated in an isotropic homogeneous field of incident measuring light. We will exploit this ICAM to test the hypothesis that acidophilic bacteria in different phyla express different types of electron transfer proteins to respire on extracellular iron. We will also study the dynamic behavior of the electron transport system at the microbe-mineral interface. This project will provide a new means to examine the extents and rates of biochemical events *in situ* without disrupting the complexity of the live cellular environment.

An experimental beta unit of an integrating cavity absorption meter was obtained from On Line Instrument Systems (Bogart, GA) in which the cuvette is a reflecting cavity completely filled with the absorbing suspension. Recent studies have focused on the reactions exhibited when suspensions of intact *At. ferrooxidans* were mixed with Fe(II) at pH 1.5. The electron transfer pathway for the flow of electrons from iron to molecular oxygen in *At. ferrooxidans* is hypoth- esized to consist of an initial electron transfer from extracellular ferrous ions to a cytochrome *c* located in the outer membrane of this Gram-negative organism. The periplasmic blue copper protein, rusticyanin, then transfers the electron from the cytochrome *c* in the outer membrane to a different periplasmic cytochrome *c*. The final electron transfer is from the periplasmic cytochrome *c* to the terminal oxidase, an *aa3*-type cytochrome that is located in the cytoplasmic membrane and reduces molecular oxygen. All three types of electron transport proteins were identified and readily visible in the difference spectrum represented by the *data points* in Fig. 1A. The reduced cytochrome *c* is represented by the peaks at 520 and 551 nm in the spectrum. Similarly, the reduced cytochrome *a* is represented by the peak at 598 nm in the difference spectrum shown in the figure. Finally, rusticyanin, the blue copper protein, has a broad absor- bance band at around 600 nm in the oxidized state that disappears entirely when the protein is reduced. Curves *a*, *b*, and *c* in Fig. 1A represent the difference spectra of 13 nM cytochrome *a*, 22 nM cytochrome *c*, and 370 nM rusticyanin, respectively. Curve *d* represents the sum of curves *a*, *b*, and *c*. It was evident that we could quantify the concentrations of all three types of electron transfer proteins in the intact bacterium at any concentration of Fe(II).

Figure 1. Absorbance measurements when 1.4×10^8 cells/ml of intact *At. ferrooxidans* were mixed with 300 µM Fe(II) at pH 1.5. **A**, *data points* define the difference spectrum obtained immediately after mixing. Curves representing difference spectra: *a*, cytochrome *a*; *b*, rusti- cyanin; *c*, cytochrome *c*; and *d*, sum of curves *a*, *b* and *c*. **B**, dependences of the concentrations of rusticyanin (*triangles*), cytochromes *c* (*inverted triangles*) and *a* (*circles*) on the initial concentration of Fe(II). **C**, kinetic traces for the aerobic oxidation of reduced cellular components. Curve *b* is the eigenvector absorbance from a global fit of the combined spectral changes from all three components visible in the intact bacterium; curve *a* is the absorbance change at 580 nm, representing primarily the rusticyanin alone. *Inset*, a residual plot of the global fit to the eigenvector absorbance.



Difference spectra such as that shown in Fig. 1A were obtained at different initial concentrations of Fe(II). Fig. 1B shows the dependence of the concentrations of each of the three reduced components on the concentration of ferrous iron. Despite the fact that all three electron transfer proteins exhibit quite different standard reduction potentials *in vitro*, the relative concentrations of the reduced proteins did not vary in the intact organism with the concentration of iron. That is, at sub-saturating concentrations of electrons derived from soluble iron, the three types of proteins maintained the same relative concentrations.

The resting absorbance spectrum of the bacterium observed under air-oxidized conditions was always regenerated from that of the Fe(II)-reduced bacterium initially observed in the presence of Fe(II). More convincing evidence for the ensemble-like behavior of the iron respiratory components came from the oxygen-dependent decay of the reduced proteins back to their resting oxidized states. Fig. 1C shows a comparison of the time courses for the oxidation of the bacter- ium. Curve *a* shows the pseudo-first order increase at 580 nm, a wavelength where the absor- bance change was primarily due to the oxidation of the rusticyanin. Curve *b* shows the identical pseudo-first order increase calculated from a multiwavelength global fit of the absorbance changes to a single exponential function of time over the range of 480 to 660 nm. It was evident that all three electron transfer proteins oxidized at the same rate, regardless of their inherent redox potential or their position in the overall respiratory chain.

This research was supported by the Office of Biological and Environmental Research in the US Department of Energy Office of Science.

153. FizzyQIIME: Feature Selection for Metagenomics

Gregory Ditzler[1], J. Calvin Morrison[1], Christopher Blackwood[2], and **Gail L. Rosen**[1]

[1]Department of Electrical & Computer Engineering, Drexel University, Philadelphia, PA
[2]Department of Biological Sciences, Kent State University, Kent, OH

http://www.ece.drexel.edu/gailr/EESI/index.php
http://github.com/EESI/FizzyQIIME

Project Goals: Develop open-source software tools for variable selection with -omics data formats

Sequences from bacterial communities are collected from whole genome shotgun (WGS), or amplicon sequencing runs, which allows researchers to study the taxonomic composition and func- tion of a sample. Ecologists represent the data in the form of an abundance matrix, which usually holds counts of operational taxonomic units (OTUs), but can also hold counts of genes/metabolic pathway occurrences. It is quite common to have different factors in a metagenomic study, such as environmental pH and salinity values, or a health related status [4, 5]. A natural question to ask about these studies with multiple factors is: "which OTUs are important for differentiating the multiple factors?" Knowing the answer to this question can be useful for understanding which conditions are driving/being affected by differences in composition and function across samples. Answering this question can be addressed using feature selection – sometimes referred to as vari- able selection [1, 2].

Motivation Some of the current software tools for comparative metagenomics provide researchers the ability to investigate and explore bacterial communities using α– & β–diversity. Feature se- lection – a sub-field of machine learning – provides an intuitive solution to performing these com- parisons. In particular, these methods pick which OTUs (or functional features) have the most influence on the condition being studied. For example, our previous work has used information theoretic feature selection to understand the differences between protein family abundances that best discriminate between different age groups in humans [3].

Results We have developed a new Python module for the QIIME software package for microbial ecologists that implements information-theoretic feature selection methods. We demonstrate the software tools capabilities on publicly available data sets.

Availability We have made the software implementation freely available under the GNU GPL. The software can be found at http://github.com/EESI/FizzyQIIME.

References

[1] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

[2] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh. *Feature Extraction: Foundations and Applications*. Springer, 2006.

[3] Y. Lan, A. Kriete, and G. Rosen. Selecting age-related functional characteristics in the human gut microbiome. *BMC Microbiome*, 2013.

[4] J. Qin, R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D. R. Mende, J. Li, J. Xu, S. Li, D. Li, J. Cao, B. Wang, H. Liang, H. Zheng, Y. Xie, J. Tap, P. Lepage, M. Bertalan, J. M. Batto, T. Hansen, D. Le Paslier, A. Lin- neberg, H. B. Nielsen, E. Pelletier, P. Renault, T. Sicheritz-Ponten, K. Turner, H. Zhu, C. Yu, M. Jian, Y. Zhou, Y. Li, X. Zhang, N. Qin, H. Yang, J. Wang, S. Brunak, J. Dore, F. Guarner, K. Kristiansen, O. Pedersen, J. Parkhill, J. Weissenbach, P. Bork, and S. D. Ehrlich. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464:59–65, 2010.

[5] P. Turnbaugh, M. Hamady, T. Yatsunenko, B. Cantarel, A. Duncan, R. Ley, M. Sogin, W. Jones, B. Roe, J. Affourtit, M. Egholm, B. Henrissat, A. Heath, R. Knight, and J. Gordon. A core gut microbiome in obese and lean twins. *Nature*, 475:480–485, 2009.

154. Integrated Pan-omics Measurements for Systems Level Characterization of Biological Systems

Presenter: Joshua N. Adkins*, Kristin E. Burnum, Carrie D. Nicora, Qibin Zhang, Stephen J. Callister, Scott R. Kronewitter, Aaron T. Wright, Ronald J. Moore, Eric Huang, Yi Qu, Zeping Hu, Xiaoli Gao, Weijun Qian, Erin S. Baker, Therese R. W. Clauss, Karl K. Weitz, Daniel J. Orton, Matthew E. Monroe, Anil K. Shukla, Rui Zhao, Roslyn Brown, **Gordon A. Anderson**, **Thomas O. Metz, Mary S. Lipton,** and **Richard D. Smith (PI;** rds@pnnl.gov**)**

Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99352

http://www.pnl.gov/

Project Goals: Achieving a predictive systems level understanding of plants, microbes and microbial communities requires the integration of developments to enable solutions to energy, environment, and climate challenges. We are applying advanced mass spectrometry (MS)-based capabilities for comprehensive molecular characterization (proteomics including post-translational modifications, metabolomics, lipidomics, and glycomics) of biological systems. Automating and speeding key steps in pan-omics sample processing, high resolution separations combined with high mass accuracy MS measurements affords large gains in measurement quality and throughput. Additional advancements include targeted proteomics methods (activity-based protein profiling and multiple reaction monitoring) and elucidation of protein proteoforms through integrated top-down and bottom-up and post-translational modifications measurements. This pan-omics approach provides new insights by elucidating complex phenotypic relationships between environmentally important microorganisms and higher organisms, as well as metabolic activities within microbial communities.

Central to developing a complete understanding of biological systems is the paradigm of pan- omics: the ability to comprehensively characterize the range of biomolecules from individual samples. To that end, we are developing MS-based approaches and technologies, and applying their capabilities in the context of GSP collaborations. One benefit of the pan-omic technology and approaches that we are developing is the ability to analyze and obtain data on a range biomolecular classes from the same sample, eliminating key sources of biological variation while greatly reducing the amount of sample required. In one aspect of these efforts, we have developed and initially evaluated an integrated biomolecule extraction method that uses a chloroform/methanol extraction to isolate proteins, as well as polar and non-polar metabolites, from the same sample. For both a model bacterium and a unicyanobacterial microbial consortium, we have also shown nearly identical results (e.g. reproducibility and proteome coverage) using these analyses compared with a standard proteomics sample workflow.

Important challenges for pan-omics approaches involve both the biological complexity of the systems being studied, which include the microbial communities of environmentally crucial ecosystems. The study of such microbial communities can provide an understanding of the manner in which microbes affect and are affected by their environment.

Examples of resent pan-omics applications and a few of the resulting insights include:

Fungus *L. gongylophorus* dominates during cellulose degradation while symbiotic bacteria play supporting roles. Studies of fungus-growing ant–microbe symbiosis are paradigmatic of organic

complexity generated through symbiotic association. We have demonstrated in-depth profiling of the fungal garden complete with bacteria (including fungus alone, isolated bacteria, and the actual fungal garden) to understand the relationship between the fungus and the bacterial protectors. Proteomics and metabolomic studies have revealed that the fungus *L. gongylophorus* plays a dominant role in breaking down cellulose and other plant polymers, while the bacteria turn the partially digested sugars into a variety of nutrients that support the fungal and ant growth.

Gluconeogensis dominates in phototrophic mats in the early morning. Phototrophic mats rely on photosynthetic organisms for carbon capture, storage and nutrient cycling to nourish heterotrophic community members. As such, these photosynthetic organisms are dependent on storing high energy nutrients during the day that can be used after the sun goes down. A pan- omics based study of phototrophic microbial mats from Yellowstone National Park revealed that during early morning, *Synechococcus* sp are engaged in gluconeogenesis rather than glycolysis. We observe pools of glucose-6-phosphate and the presence of bifunctional fructose 1,6- bisphosphatase which catalyzes the reversed reaction of PFK and phosphoenolpyruvate synthase, which catalyzes the reversed reaction of the other major control point of glycolysis.

Pan-omics reveals community elasticity comes at a price of functional redundancy. For many microbial communities presently of interest, the metagenome is unavailable; presenting a significant challenge since proteomics is often interpreted using genome sequence(s). However, the pan-omics approaches we are developing can still provide insights into the ecology of such microbial communities. In a study of a community isolated from cow rumen grown in an engineered bioreactor in the absence of metagenome data, our pan-omic approach (including proteomic, transcriptomics, and metabolomics studies) revealed redundant patterns and correlations between groups of proteins and metabolites with the community structure. Additionally, we found that after perturbation of the community, the pan-ome for the community returned to a new steady state, but the functional redundancy was decreased.

Revealing the 'true proteome' of periplasmic proteins. An important aspect of our approach is the ability to provide extensive information on the actual proteoforms present, and where use of conventional bottom-up proteomics approaches are generally ineffective. The periplasm of Gram-negative bacteria is a dynamic and physiologically important subcellular compartment where the constant exposure to potential environmental insults amplifies the need to protect function, and thus the proteoforms present are functionally important. Hence, the evaluation of the periplasmic fraction for *Novosphingobium aromaticivorans* revealed a large array of proteoforms for 55 proteins in the periplasm. The proteoforms found included post translational modifications due to signal peptide removal, N-terminal methionine excision, acetylation, glutathionylation, pyroglutamate, and disulfide bond formation.

Acknowledgements: This research is supported by the Office of Biological and Environmental Research (OBER) of the U.S. Department of Energy (DOE). Portions of this research were performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the DOE's OBER. Pacific Northwest National Laboratory is operated for the DOE by Battelle Memorial Institute through Contract No. DE-AC05-76RLO 1830.

155. Modeling metabolic dynamics in 3D: applications to synthetic microbial ecosystems

William R. Harcombe[1,€,*], William Riehl[2,#], Ilija Dukovski[2,*](dukovski@bu.edu), Brian R. Granger[2], Alex Betts[1,&], Alex H. Lang[3], Gracia Bonilla[2], Amrita Kar[2], Nicholas Leiby[1,4], Nathaniel C. Cady[5], Pankaj Mehta[2,3], Christopher J. Marx[1,6,7], **Daniel Segrè**[2,8] (**dsegre@bu.edu**)

http://www.bu.edu/segrelab/

[1]Dept. of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA; [2]Bioinformatics Program, Boston University, Boston, MA; [3]Dept. of Physics, Boston University, Boston, MA; [4]Systems Biology Graduate Program, Harvard University, Cambridge, MA; [5]College of Nanoscale Science & Engineering, University of Albany, Albany, NY; [6]Faculty of Arts and Sciences Center for Systems Biology, Harvard University, Cambridge, MA; [7]Dept. of Biological Sciences, University of Idaho, Moscow, ID; [8]Dept. of Biology and Dept. of Biomedical Engineering, Boston University, Boston, MA; [#]Current address: Lawrence Berkeley National Laboratory, Berkeley, CA; [&]Current address: Dept. of Zoology, University of Oxford, Oxford, United Kingdom; [€]Current address: Dept. of Ecology, Evolution and Behavior, University of Minnesota, St. Paul, MN.

Project goals: The goal of this project is to develop and test a computational tool for modeling metabolic flux and inter-species interactions in spatially structured microbial ecosystems. Our approach, named Computation of Microbial Ecosystems in Time and Space (COMETS), combines dynamic flux balance modeling with diffusion. Experimental testing has confirmed the capacity of COMETS to predict nontrivial results about the spatio-temporal dynamics of small synthetic communities. Our initial 2D version is being expanded to a 3D version, and applied to a systematic large-scale study of pairwise interactions between microbes relevant to biofuel production, bioremediation and biogeochemical cycles.

Although often studied alone in well-mixed flasks, most microbial organisms live in multi- species, structured, highly dynamic consortia [1]. COMETS is a multi-scale modeling framework that computes ecosystem-level metabolic dynamics based on detailed intracellular metabolic stoichiometry, without any a priori assumption on whether and how different species would interact. It implements a dynamic FBA (dFBA, [2]) algorithm on a lattice, making it possible to track the spatio-temporal dynamics of multiple microbial species in complex environments with complete genome scale resolution [3]. As a first test of COMETS we verified that it can quantitatively recapitulate known aspects of single colony growth, such as the linear expansion of colony diameter, under different carbon sources. We than applied COMETS to a previously constructed *E. coli*/*S. enterica* artificial consortium [4], and showed that the simulation correctly predicts the convergence of the system to specific species ratios. More surprisingly, COMETS yielded correct predictions also for a newly engineered three-member consortium that incorporates *Methylobacterium extorquens* AM1 into the *E. coli*/*S. enterica* system (Figure 1). Finally, we investigated the impact that spatial arrangement has on colony interactions, with special attention to a puzzling shadowing effect between colonies that engage in syntrophic interactions, which we refer to as a "metabolic eclipse". We found that COMETS accurately predicts community interactions and dynamics as the natural outcome of intracellular metabolic processes, and suggests unexpected new features of model consortia, with important implications for understanding natural and synthetic microbial communities. From its initial two- dimensional version, COMETS has now been extended to a fully functional three dimensional platform. COMETS in 3D can find extensive applications in the detailed study of bacterial biofilm growing in structured environments and environmental gradients.

COMETS 3D visualization is based on the open source OpenGL 3D library implemented as Java binding (JOGL), thus maintaining the open source nature of the code. Additional new features being added to COMETS include the capacity to model metabolite diffusion through biomass, and the introduction of lag phase. Moreover, we have been able to import in COMETS more than a hundred stoichiometric models from Model SEED [5], paving the way for systematic combinatorial studies of inter-species interactions.

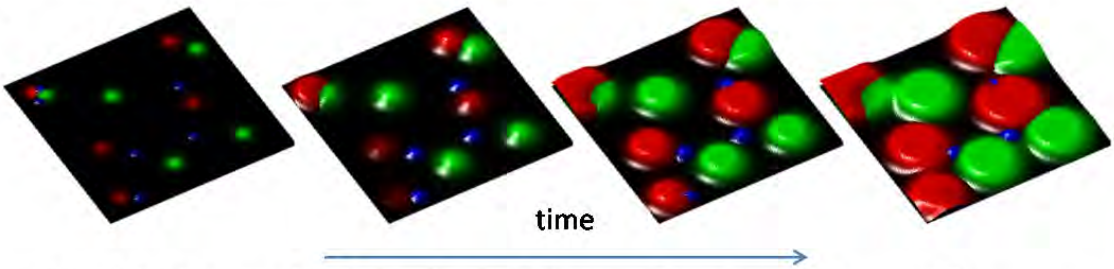


Figure 1: Spatio-temporal simulation of a microbial community composed of three mutually interdependent species (*Salmonella typhimurium, E. coli, Methylobacterium extorquens*), performed using COMETS (Computation of Microbial Ecosystems in Time and Space).

References

[1] V. Mazumdar, S. Amar and D. Segrè: Metabolic proximity in the order of colonization of a microbial community, PLOS ONE (2013), 8(10): e77617.
[2] Mahadevan R., Edwards J.S. and Doyle F.J. Dynamic flux balance of diauxic growth in Escherichia coli. Biophys J. 2002, Sep; 83(3): 1331-40.
[3] W. R. Harcombe, W. Riehl, I. Dukovski, B. R. Granger, A. Betts, A. H. Lang, G. Bonilla, A. Kar, N. Leiby, P. Mehta, C. J. Marx, D. Segrè. Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. Revised manuscript in preparation.
[4] Harcombe, W. (2010). Novel cooperation experimentally evolved between species. Evolution; International Journal of Organic Evolution 64, 2166–2172.
[5] Henry, C.S., DeJongh, M., Best, A. a, Frybarger, P.M., Linsay, B., and Stevens, R.L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. Nature Biotechnology 28, 977–982

Funding: DOE-DE-SC0004962, DE-SC0006731

156. Expanding the toolkit for comparative metagenomics, implementing it in K-Base, and applying it to the study of the effects of experimental warming in Midwestern and Alaskan soils

Konstantinos T. Konstantinidis[1*] (kostas@ce.gatech.edu), Eric R. Johnston[1], Chengwei Luo[1], Luis M. Rodriguez-R[1], **Liyou Wu**[2], Kai Xue[2], **Zhili He**[2], Mengting Maggie Yuan[2], **Yiqi Luo**[2], **Edward A.G. Schuur**[3], James R. Cole[4], **James M. Tiedje**[4], **Jizhong Zhou**[2].

Georgia Institute of Technology, Atlanta, GA 30332, USA [1]; University of Oklahoma, Norman, OK 73019, USA [2]; University of Florida, Gainesville, FL 32611, USA [3]; Michigan State University, East Lansing, MI 48824, USA [4].

http://ieg.ou.edu

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. Towards this goal, we are pursuing the following objectives: (i) To determine the responses of microbial community structure, functions and activities to climate warming, altered precipitation, soil moisture regime and/or clipping in the tundra and temperate grassland ecosystems; (ii) To determine the temperature sensitivity and substrate priming on recalcitrant C decomposition; (iii) To determine microbiological basis underlying temperature sensitivity of recalcitrant C decomposition; 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 the K-Base.

Abstract: Under this project, we have begun investigations on microbial communities from Alaskan tundra permafrost (AK) and Oklahoma temperate grassland (OK) soils, both of which have been experimentally warmed 2 to 4 °C for one and half year above ambient temperature *in-situ*. Our analyses of well-replicated 16S rRNA gene amplicon, meta-transcriptomic, and whole-community shotgun metagenomic datasets from these soils showed small but significant shifts in community composition, gene expression, and functional metabolic potential compared to control (un-warmed) adjacent communities. The specific microbial populations and genes/pathways enriched by warming were different between the two locations. Greater taxonomic composition differences were observed at the OK site relative to AK, presumably resulting from longer generation times due to the less optimal conditions for growth at permafrost soils. Analysis of fragments of rRNA genes recovered in the shotgun-metagenomic data revealed no significant shifts in fungal taxa at both sites, but that the ratio of fungi to bacteria decreased by warming, indicating that the warming treatment is more favorable for bacteria, at least in the short term. The most pronounced bacterial taxon shifts observed at OK site, which were somewhat also observed at the AK site, were increased in abundance of *Actinobacteria* and decreased in *Planctomycetes*, both representing major phyla in soils, particularly in regards to C-cycling. In terms of functions, the communities of AK warmed plots were enriched in metabolic pathways related to labile carbon mobilization and oxidation whereas fewer of these patterns were observed in the OK communities, indicating that soil C is more vulnerable to microbial respiration at AK. The OK microbial communities were instead enriched in genes involved in heat shock response and cellular surface structures, particularly, trans-membrane transporters for glucosides and ferrous iron. These results, which were consistent with independent

physicochemical measurements and process rates determined *in-situ*, were linked with higher primary productivity of the aboveground plant communities stimulated by warming. Collectively, our findings suggest that microbial communities of grassland soils play important roles in mediating feedback responses of the soil ecosystem to climate change and that even short periods of warming induce significant changes in microbial community function and composition.

To enable this research, we have also developed several bioinformatics tools that addressed practical limitations during the comparative analysis of the soil metagenomes such as how to assess the fraction of the community captured by a metagenomic dataset and estimate the sequencing effort required in study design (Nonpareil tool; Rodriguez-R and Konstantinidis, Bioinformatics 2013), how to determine the taxonomic affiliation of a metagenomic sequence (MyTaxa; Luo et al., in revision), how to bin assembled contigs into population genomes based on time-series metagenomes (BinGeR; Luo et al., in preparation), and how to determine differentially present genes between metagenomic datasets (Luo et al., Methods Enzymol. 2013). Furthermore, we have developed new tools for the analysis of high-throughput gene amplicon data, including a tool for frameshift correction and nearest-neighbor classification (FrameBot; Wang et al., mBio2013), and a 16S rRNA classifier for short read data (Cole et al., Nucleic Acids Res. 2014). Altogether, these make up our *Microbial Process Toolkit* for gene, metagenomic and metatranscriptomic data integration, modeling and visualization. We are in the process of implementing our toolkit in K-Base and we will report on these efforts as well.

References

1. C. Luo, L. M. Rodriguez-R, and K. T. Konstantinidis. A User's Guide to Quantitative and Comparative Analysis of Metagenomic Datasets. Methods Enzymol. 531:525-47. 2013.
2. C. Luo, L. M. Rodriguez-R, E. R. Johnston, L. Wu, L. Cheng, K. Xue, Q. Tu, Y. Deng, Z. He, J. Zhou Shi, M. M. Yuan, R. A. Sherry, D. Li, Y. Luo, E. A. G. Schuur, P. Chain, J. M. Tiedje, J. Zhou, and K. T. Konstantinidis. Soil Microbial Community Responses to a Decade of Warming as Revealed by Comparative Metagenomics. Applied & Environ. Micriobiol. *In press*.
3. J. R. Cole, Q. Wang , J. A. Fish, B. Chai , D. M. McGarrell, Y. Sun , C. T. Brown, Porras-Alfaro, C. R. Kuske, and J. M. Tiedje. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 42(1):D633-42. 2014.

***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-SC0004601,*

157. The Network Portal and Gaggle Workspace: New Generation of Tools for Network Biology

Serdar Turkarslan[1]*(sturkarslan@systemsbiology.org), Elisabeth J. Wurtmann[1], Wei-Ju Wu[1], Ning Jiang[1], J. Christopher Bare[1], Karen Foley[1], Pavel Novichkov[2] and **Nitin S. Baliga[1]**

[1] Institute for Systems Biology, Seattle, WA, 98109, USA, [2] Lawrence Berkeley National Labs, Berkeley, CA, 94720, USA

http://networks.systemsbiology.net

Project Goals: The mission of this project is to develop set of tools to democratize access to the inference, storage, exploration, and visualization of gene regulatory networks. In addition, it aims to promote cross-platform data analysis and collaboration among researchers with distinct areas of expertise by extending the Gaggle framework.

Abstract: Recent years have witnessed a bloom of algorithms to infer regulatory network architectures to investigate the regulation of diverse biological processes. Yet the wider scientific community has limited access to tools for network inference, visualization, and analysis because these tasks often require advanced computational knowledge and expensive computing resources. We have designed the Network Portal (http://networks.systemsbiology.net) to serve as a modular database for inferring networks by integrating data (user uploaded and public). The portal is also fully equipped with tools for storage, visualization, and analysis of biological networks. The current release of the database contains networks for 13 prokaryotic organisms from diverse phylogenetic clades (4678 co-regulated gene modules, 3466 regulators and 9291 cis- regulatory motifs). The portal will be rapidly populated with additional networks from diverse organisms as relevant data become available in public repositories and through user input. The portal is fully integrated into the Gaggle framework for interactive exploration of diverse kinds of data using a wide array of desktop and web-based applications. We have also developed a Gaggle Workspace environment for data organization, storage, and analysis with workflow capabilities and a save state function. The Workspace is a data-centric, cloud-based, highly scalable, and multi-platform solution that improves the productivity of performing large-scale data analysis and visualization for biological researchers with and without programming expertise. The Network Portal has already contributed algorithms (e.g., the network inference pipeline) as well as data (e.g., networks) to the DOE KBase. We are also in the process of implementing KBase services into Gaggle Workspace environment.

References: Turkarslan, S., Wurtmann E. J., Wu W. - J., Jiang N., Bare C. J., Foley K., et al. (2013). Network portal: a database for storage, analysis and visualization of biological networks. Nucleic Acids Res. 42(1):D184-90

158. KBase Overview

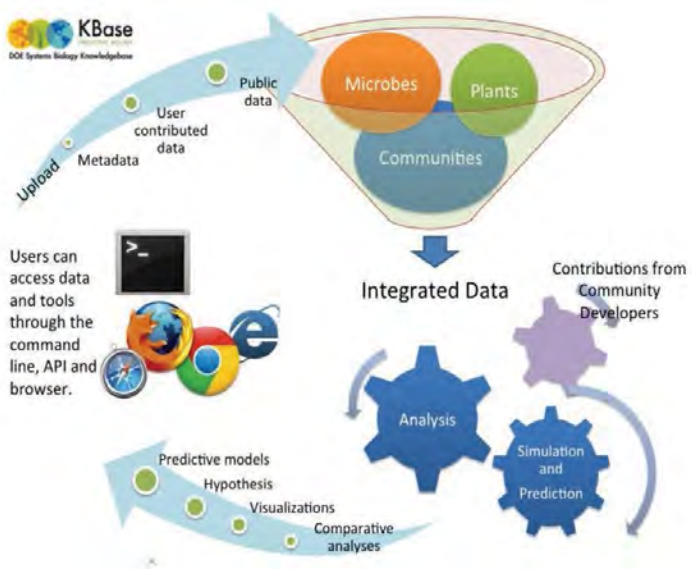
Adam P. Arkin*[1] (aparkin@lbl.gov), Jason Baumohl[1], Aaron Best[2], Jared Bischof[2], Ben Bowen[1], Tom Brettin[2], Tom Brown[2], Shane Canon[1], Stephen Chan[1], John-Marc Chandonia[1], Dylan Chivian[1], Ric Colasanti[2], Neal Conrad[2],, Brian Davison[3], Matt DeJongh[6], Paramvir Dehal[1], Narayan Desai[2], Scott Devoid[2], Terry Disz[2], Meghan Drake[3], Janaka Edirisinghe[2], Gang Fang[7], José Pedro Lopes Faria[2], Mark Gerstein[7], Elizabeth M. Glass[2], Annette Greiner[1], Dan Gunter[1], James Gurtowski[5], Nomi Harris[1], Travis Harrison[2], Fei He[4], Matt Henderson[1], Chris Henry[2], Adina Howe[2], Marcin Joachimiak[1], Kevin Keegan[2], Keith Keller[1], Guruprasad Kora[3], Sunita Kumari[5], Miriam Land[3],, Folker Meyer[2], Steve Moulton[3], Pavel Novichkov[1], Taeyun Oh[8], Gary Olsen[9], Bob Olson[2], Dan Olson[2], Ross Overbeek[2], Tobias Paczian[2], Bruce Parrello[2], Shiran Pasternak[5], Sarah Poon[1], Gavin Price[1], Srivdya Ramakrishnan[5], Priya Ranjan[3], Bill Riehl[1], Pamela Ronald[8], Michael Schatz[5], Lynn Schriml[10], Sam Seaver[2], Michael W. Sneddon[1], Roman Sutormin[1], Mustafa Syed[3], James Thomason[5], Nathan Tintle[6], Will Trimble[2], Daifeng Wang[7], Doreen Ware[5], David Weston[3], Andreas Wilke[2], Fangfang Xia[2], Shinjae Yoo[4], Dantong Yu[4], **Bob Cottingham[3], Sergei Maslov[4], Rick Stevens[2]**

[1] Lawrence Berkeley National Laboratory, Berkeley, CA, [2] Argonne National Laboratory, Argonne, IL, [3] Oak Ridge National Laboratory, Oak Ridge, TN, [4] Brookhaven National Laboratory, Upton, NY, [5] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, [6]Hope College, Holland, MI, [7] Yale University, New Haven, CT, [8] University of California, Davis, CA, [9] University of Illinois at Champaign-Urbana, Champaign, IL, [10]University of Maryland, College Park, MD

http://kbase.us

Project Goals: The KBase project aims to provide the computational capabilities needed to address the grand challenge of systems biology: to predict and ultimately design biological function. KBase enables users to collaboratively integrate the array of heterogeneous datasets, analysis tools and workflows needed to achieve a predictive understanding of biological systems. It incorporates functional genomic and metagenomic data for thousands of organisms, and diverse tools including (meta)genomic assembly, annotation, network inference and modeling, thereby allowing researchers to combine diverse lines of evidence to create increasingly accurate models of the physiology and community dynamics of microbes and plants. KBase will soon allow models to be compared to observations and dynamically revised. A new prototype Narrative interface lets users create a reproducible record of the data, computational steps and thought process leading from hypothesis to result in the form of interactive publications.

The Department of Energy (DOE) Systems Biology Knowledgebase (KBase) is an emerging computational environment that enables researchers to bring together the diverse data, algorithms, analytical tools, and workflows needed to achieve a predictive understanding of biological systems (see figure, KBase Overview). As a project supported by the Office of Biological and Environmental Research within the DOE Office of Science, KBase focuses on microbial and plant systems that support DOE missions in energy production and environmental science. However, the KBase approach to analyzing and modeling DOE-relevant microbes and plants can be applied to organisms from across the tree of life.



KBase Overview. This overview diagram shows how users interact with KBase by uploading biological data, analyzing it with tools developed by both KBase and the community, and using analysis results to drive experiments and a better understanding of biological systems.

KBase is also an open and extensible development environment that invites and trains community members to contribute new tools and data. Tool developers can implement their methods as new KBase services, making their tool accessible to a wide user community and placing a world of biological data at their fingertips for tool validation. Data producers can integrate their data into the KBase data model, so that all of the analysis and visualization tools available in KBase may be applied to interpret the data. By enabling members of the community to integrate and use a wide spectrum of analysis tools and datasets, KBase will serve as a catalyst for biological research, accelerating discovery for DOE missions and providing insights and benefits that can ultimately serve numerous application areas.

Systems biology is driven by the ever-increasing wealth of data resulting from new generations of genomics-based technologies. With the success of genome sequencing, biology began to generate and accumulate data at an exponential rate. In addition to the massive stream of sequencing data, each type of technology that researchers use to analyze a sequenced organism adds another layer of complexity to the challenge of understanding how different biological components work together to form a functional living system. Achieving this systems-level understanding of biology will enable researchers to predict and ultimately design how the system will function under certain conditions. Gaining this predictive understanding, however, requires an unprecedented level of collaboration among researchers in different disciplines around the world. A new collaborative computational environment is needed to bring these researchers together so they can share and integrate large, heterogeneous datasets and readily use this information to develop predictive models that drive scientific discovery.

KBase is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

159. The DOE Systems Biology Knowledgebase: Plant Science Domain

Doreen Ware[1,2*] (ware@cshl.edu), David Weston[3], Jason Baumohl[4], Aaron Best[5], Jared Bischof[6], Ben Bowen[4], Tom Brettin[6], Tom Brown[6], Shane Canon[4], Stephen Chan[4], John- Marc Chandonia[4], Dylan Chivian[4], Ric Colasanti[6], Neal Conrad[6], Brian Davison[4], Matt DeJongh[5], Paramvir Dehal[4], Narayan Desai[6], Scott Devoid[6], Terry Disz[6], Meghan Drake[3], Janaka Edirisinghe[6], Gang Fang[7], José Pedro Lopes Faria[6], Mark Gerstein[7], Elizabeth M. Glass[6], Annette Greiner[4], Dan Gunter[4], James Gurtowski[2], Nomi Harris[4], Travis Harrison[6], Fei He[1], Matt Henderson[4], Chris Henry[6], Adina Howe[6], Marcin Joachimiak[4], Kevin Keegan[6], Keith Keller[4], Guruprasad Kora[3], Sunita Kumari[2], Miriam Land[3], Folker Meyer[6], Steve Moulton[3], Pavel Novichkov[4], Taeyun Oh[8], Gary Olsen[9], Bob Olson[5], Dan Olson[5], Ross Overbeek[5], Tobias Paczian[5], Bruce Parrello[5], Shiran Pasternak[2], Sarah Poon[4], Gavin Price[4], Srivdya Ramakrishnan[2], Priya Ranjan[3], Bill Riehl[4], Pamela Ronald[8], Michael Schatz[2], Lynn Schriml[10], Sam Seaver[6], Michael W. Sneddon[4], Roman Sutormin[4], Mustafa Syed[3], James Thomason[2], Nathan Tintle[5], Will Trimble[6], Daifeng Wang[7], Andreas Wilke[6], Fangfang Xia[6], Shinjae Yoo[1], Dantong Yu[1], **Robert Cottingham[3]**, **Sergei Maslov[1]**, **Rick Stevens[6]**, **Adam P. Arkin[4]**

[1]Brookhaven National Laboratory, Upton, NY, [2]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, [3]Oak Ridge National Laboratory, Oak Ridge, TN, [4]Lawrence Berkeley National Laboratory, Berkeley, CA, [5]Hope College, Holland, MI, [6]Argonne National Laboratory, Argonne, IL, [7] Yale University, New Haven, CT, [8]University of California, Davis, CA, [9] University of Illinois at Champaign-Urbana, Champaign, IL, [10]University of Maryland, College Park, MD

http://kbase.us

Project Goals: The KBase project aims to provide the capabilities needed to address the grand challenge of systems biology: to predict and ultimately design biological function. KBase enables users to collaboratively integrate the array of heterogeneous datasets, analysis tools and workflows needed to achieve a predictive understanding of biological systems. It incorporates functional genomic and metagenomic data for thousands of organisms, and diverse tools for (meta) genomic assembly, annotation, network inference and modeling, allowing researchers to combine diverse lines of evidence to create increasingly accurate models of the physiology and community dynamics of microbes and plants. KBase will soon allow models to be compared to observations and dynamically revised. A new prototype Narrative interface lets users create a reproducible record of the data, computational steps and thought process leading from hypothesis to result in the form of interactive publications.

DOE Systems Biology Knowledgebase (KBase) has two central goals. The scientific goal is to produce predictive models, reference datasets, and analytical tools and to demonstrate their utility in DOE biological research relating to bioenergy, carbon cycle, and the study of subsurface microbial communities. The operational goal is to create the integrated software and hardware infrastructure needed to support the creation, maintenance, and use of predictive models and methods.

The plant team is currently focused on reconstruction and modeling of genotype-to- phenotype relationships in plant species relevant for DOE mission. Our workflows accessible via narrative and command line interfaces provide interactive, data-driven analysis and exploration across multiple experiments and diverse data-types. KBase allows our users to process next generation sequencing data to identify novel genomic variation and to quantify genome-wide expression levels. Users can process

expression data to calculate co-expression networks and to identify and annotate functional modules within those networks. Furthermore, we provide computational tools to carry out Genome-Wide Association analysis to identify SNPs and candidate genes significantly correlated with plant phenotypes. Predicted genotype-to-phenotype relationships can be validated by a variety of public and user-generated networks metabolic models in KBase.

KBase is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

160. Analysis capabilities for microbial communities in KBase

Folker Meyer*[1](folker@anl.gov), Dylan Chivian[2], Jason Baumohl[2], Aaron Best[3], Jared Bischof[1], Ben Bowen[2], Thomas Brettin[1], Tom Brown[1], Shane Canon[1], Stephen Chan[2], John-Marc Chandonia[2], Ric Colasanti[1], Neal Conrad[1], Brian Davison[4], Matt DeJongh[3], Paramvir Dehal[2], Narayan Desai[1], Scott Devoid[1], Terry Disz[1], Meghan Drake[4], Janaka Edirisinghe[1], Gang Fang[7], José Pedro Lopes Faria[1], Mark Gerstein[7], Elizabeth M. Glass[1], Annette Greiner[2], Dan Gunter[2], James Gurtowski[6], Nomi Harris[2], Travis Harrison[1], Fei He[5], Matt Henderson[2], Chris Henry[1], Adina Howe[1], Marcin Joachimiak[2], Kevin Keegan[1], Keith Keller[2], Guruprasad Kora[4], Sunita Kumari[6], Miriam Land[4],, Steve Moulton[4], Pavel Novichkov[2], Taeyun Oh[8], Gary Olsen[9], Bob Olson[1], Dan Olson[1], Ross Overbeek[1], Tobias Paczian[1], Bruce Parrello[1], Shiran Pasternak[6], Sarah Poon[2], Gavin Price[2], Srivdya Ramakrishnan[6], Priya Ranjan[4], Bill Riehl[2], Pamela Ronald[8], Michael Schatz[6], Lynn Schriml[10], Sam Seaver[1], Michael W. Sneddon[2], Roman Sutormin[2], Mustafa Syed[4], James Thomason[6], Nathan Tintle[3], Will Trimble[1], Daifeng Wang[7], Doreen Ware[5,6], David Weston[4], Andreas Wilke[1], Fangfang Xia[1], Shinjae Yoo[5], Dantong Yu[5], **Robert Cottingham[4], Sergei Maslov[5], Rick Stevens[1], Adam P. Arkin[2]**

[1]Argonne National Laboratory, Argonne, IL, [2]Lawrence Berkeley National Laboratory, Berkeley, CA, [3]Hope College, Holland, MI, [4]Oak Ridge National Laboratory, Oak Ridge, TN, [5]Brookhaven National Laboratory, Upton, NY, [6]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, [7]Yale University, New Haven, CT, [8]University of California, Davis, CA, [9]University of Illinois at Champaign-Urbana, Champaign, IL, [10]University of Maryland, College Park, MD

http://kbase.us

Project Goals: The KBase project aims to provide the capabilities needed to address the grand challenge of systems biology: to predict and ultimately design biological function. KBase enables users to collaboratively integrate the array of heterogeneous datasets, analysis tools and workflows needed to achieve a predictive understanding of biological systems. It incorporates functional genomic and metagenomic data for thousands of organisms, and diverse tools for (meta)genomic assembly, annotation, network inference and modeling, allowing researchers to combine diverse lines of evidence to create increasingly accurate models of the physiology and community dynamics of microbes and plants. KBase will soon allow models to be compared to observations and dynamically revised. A new prototype Narrative interface lets users create a reproducible record of the data, computational steps and thought process leading from hypothesis to result in the form of interactive publications.

KBase provides the infrastructure and tooling for in-depth metagenome analysis, facilitating the annotation of microbial communities and the quest for identification of key players in a microbial community or the identification of trends. By automatically transforming microbial communities into abundance profiles KBase is enabling users to drill down so that trends, specific taxa or functions can be identified. Combining metagenomic and environmental data makes it possible to correlate information about organism or function abundance to metadata that describe a variety of biologically intriguing characteristics of the samples, such as the biome the samples were collected from, the pH of the samples, etc.

Metabolic modeling can help elucidate the roles played by individual taxa in microbial communities by providing a detailed characterization of their functional repertoire. Once obtained this knowledge can be used for a number of purposes, such as the prediction of cultivation conditions for functionally important taxa. While the current state of modeling and our ability to annotate microorganisms has advanced greatly in recent years, the emerging models should be viewed as a first approximation rather than the final answer to these questions.

With KBase the functionality is in place to perform comparisons of multiple strategies for deriving metabolic models from microbial community data. Those strategies include the use of PCR primer amplified ribosomal genes as a reporter for the organisms informing PICRUSt predictions, the use of shotgun metagenomics to obtain functional information, the use of EMIRGE to extract complete ribosomal sequences from shotgun metagenomics sequences and the use of taxonomic information obtained from metagenomic sequences to inform PiCRUST predictions.

KBase is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

161. Analysis of Microbes

Paramvir Dehal*[1](pdehal@lbl.gov), Chris Henry[2], Jason Baumohl[1], Aaron Best[1], Jared Bischof[2], Ben Bowen[1], Tom Brettin[2], Tom Brown[2], Shane Canon[1], Stephen Chan[1], John- Marc Chandonia[1], Dylan Chivian[1], Ric Colasanti[2], Neal Conrad[2], Brian Davison[3], Matt DeJongh[6], Paramvir Dehal[1], Narayan Desai[2], Scott Devoid[2], Terry Disz[2], Meghan Drake[3], Janaka Edirisinghe[2], Gang Fang[7], José Pedro Lopes Faria[2], Mark Gerstein[7], Elizabeth M. Glass[2], Annette Greiner[1], Dan Gunter[1], James Gurtowski[5], Nomi Harris[1], Travis Harrison[2], Fei He[4], Matt Henderson[1], Chris Henry[2], Adina Howe[2], Marcin Joachimiak[1], Kevin Keegan[2], Keith Keller[1], Guruprasad Kora[3], Sunita Kumari[5], Miriam Land[3], Folker Meyer[2], Steve Moulton[3], Pavel Novichkov[1], Taeyun Oh[8], Gary Olsen[9], Bob Olson[2], Dan Olson[2], Ross Overbeek[2], Tobias Paczian[2], Bruce Parrello[2], Shiran Pasternak[5], Sarah Poon[1], Gavin Price[1], Srivdya Ramakrishnan[5], Priya Ranjan[3], Bill Riehl[1], Pamela Ronald[8], Michael Schatz[5], Lynn Schriml[10], Sam Seaver[2], Michael W. Sneddon[1] , Roman Sutormin[1], Mustafa Syed[3], James Thomason[5], Nathan Tintle[6], Will Trimble[2], Daifeng Wang[7], Doreen Ware[5], David Weston[3], Andreas Wilke[2], Fangfang Xia[2], Shinjae Yoo[4], Dantong Yu[4], **Robert Cottingham[3]**, **Sergei Maslov[4]**, **Rick Stevens[2]**, **Adam P. Arkin[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley, CA, [2]Argonne National Laboratory, Argonne, IL, [3]Oak Ridge National Laboratory, Oak Ridge, TN, [4]Brookhaven National Laboratory, Upton, NY, [5]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY [6]Hope College, Holland, MI, [7]Yale University, New Haven, CT, [8]University of California, Davis, CA, [9]University of Illinois at Champaign-Urbana, Champaign, IL, [10]University of Maryland, College Park, MD

http://kbase.us

Project Goals: The KBase project aims to provide the capabilities needed to address the grand challenge of systems biology: to predict and ultimately design biological function. KBase enables users to collaboratively integrate the array of heterogeneous datasets, analysis tools and workflows needed to achieve a predictive understanding of biological systems. It incorporates functional genomic and metagenomic data for thousands of organisms, and diverse tools for (meta)genomic assembly, annotation, network inference and modeling, allowing researchers to combine diverse lines of evidence to create increasingly accurate models of the physiology and community dynamics of microbes and plants. KBase will soon allow models to be compared to observations and dynamically revised. A new prototype Narrative interface lets users create a reproducible record of the data, computational steps and thought process leading from hypothesis to result in the form of interactive publications.

The Microbial Sciences component of the KBase project has three overall goals: 1) to enable the generation of predictive models for metabolism and gene regulation to facilitate the manipulation of microbial function; 2) to vastly increase the capability of the scientific community to communicate and utilize existing data; and 3) to enable the planning of effective experiments and to maximize our understanding of microbial system functions. To achieve these goals we have focused on unifying existing 'omics datasets and modeling toolsets within a single integrated framework that will enable users to move seamlessly from the genome assembly and annotation process through to a reconciled metabolic and regulatory model that is linked to all existing experimental data for a particular organism. The results are hypotheses for such things as gene-function matching and the use of comparative functional genomics to perform higher quality evidence-based annotations. KBase offers tools for reconciling the models

against experimental growth phenotype data, and using them to predict phenotypes in novel environments or under genetic perturbations.

To prioritize the development of the microbial science area and enable new science, we will focus on accomplishing prototype science workflows rather than general tasks. To date we have developed KBase capabilities and demonstrations workflows for: (1) genome annotation and metabolic reconstruction, (2) regulon reconstruction, (3) metabolic and regulatory model reconstruction, and (4) reconciliation with experimental phenotype and expression data.

KBase is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

162. KBase Outreach and Partnership

Elizabeth M. Glass*[1] (marland@mcs.anl.gov), Brian Davison[2], Jason Baumohl[3], Aaron Best[4], Jared Bischof[1], Ben Bowen[3], Tom Brettin[1], Tom Brown[1], Shane Canon[3], Stephen Chan[3], John-Marc Chandonia[3], Dylan Chivian[3], Ric Colasanti[1], Neal Conrad[1], Matt DeJongh[4], Paramvir Dehal[3], Narayan Desai[1], Scott Devoid[1], Terry Disz[1], Meghan Drake[2], Janaka Edirisinghe[1], Gang Fang[7], José Pedro Lopes Faria[1], Mark Gerstein[7], Annette Greiner[3], Dan Gunter[3], James Gurtowski[6], Nomi Harris[3], Travis Harrison[1], Fei He[5], Matt Henderson[3], Chris Henry[1], Adina Howe[1], Marcin Joachimiak[3], Kevin Keegan[1], Keith Keller[3], Guruprasad Kora[2], Sunita Kumari[6], Miriam Land[2], Folker Meyer[1], Steve Moulton[2], Pavel Novichkov[3], Taeyun Oh[8], Gary Olsen[9], Bob Olson[1], Dan Olson[1], Ross Overbeek[1], Tobias Paczian[1], Bruce Parrello[1], Shiran Pasternak[6], Sarah Poon[3], Gavin Price[3], Srivdya Ramakrishnan[6], Priya Ranjan[2], Bill Riehl[3], Pamela Ronald[8], Michael Schatz[6], Lynn Schriml[10], Sam Seaver[1], Michael W. Sneddon[3], Roman Sutormin[3], Mustafa Syed[2], James Thomason[6], Nathan Tintle[4], Will Trimble[1], Daifeng Wang[7], Doreen Ware[5,6], David Weston[2], Andreas Wilke[1], Fangfang Xia[1], Shinjae Yoo[5], Dantong Yu[5], **Robert Cottingham[2]**, **Sergei Maslov[5]**, **Rick Stevens[1]**, **Adam P. Arkin[3]**

[1]Argonne National Laboratory, Argonne, IL, [2]Oak Ridge National Laboratory, Oak Ridge, TN, [3]Lawrence Berkeley National Laboratory, Berkeley, CA, [4]Hope College, Holland, MI, [5]Brookhaven National Laboratory, Upton, NY, [6]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, [7]Yale University, New Haven, CT, [8]University of California, Davis, CA, [9]University of Illinois at Champaign-Urbana, Champaign, IL, [10]University of Maryland, College Park, MD

http://kbase.us

Project Goals: The KBase project aims to provide the capabilities needed to address the grand challenge of systems biology: to predict and ultimately design biological function. KBase enables users to collaboratively integrate the array of heterogeneous datasets, analysis tools and workflows needed to achieve a predictive understanding of biological systems. It incorporates functional genomic and metagenomic data for thousands of organisms, and diverse tools for (meta)genomic assembly, annotation, network inference and modeling, allowing researchers to combine diverse lines of evidence to create increasingly accurate models of the physiology and community dynamics of microbes and plants. KBase will soon allow models to be compared to observations and dynamically revised. A new prototype Narrative interface lets users create a reproducible record of the data, computational steps and thought process leading from hypothesis to result in the form of interactive publications.

Outreach goals derive from the main KBase objective of building a community-driven computational system for systems biology research. Our ultimate goal is to engage a diverse user community and maximize successful use of KBase to advance predictive biology. To accomplish this requires developing relationships and understanding researchers, their scientific objectives, and how KBase could benefit their research. Since KBase is a community-driven project, feedback from users or potential users is very important as we aim to drive KBase development to meet their needs. This includes improvements to KBase system capabilities, workflows, documentation and training. In part Outreach acts as the user advocate to the KBase development team, ensuring the development team is aware of the research user community's needs, priorities and perspective on KBase. Outreach is also responsible for effectively communicating KBase utilities and benefits to researchers and training them to be high performing KBase

users and contributors. Success for the KBase project ultimately depends on learning from the user community and determining how best to reach, inform, and respond to their needs and priorities so we can help them deliver successful, high-quality science.

Engagement is accomplished via a variety of mechanisms and venues. We are making personal contact with DOE-BER funded researchers (and over time expanding our scope to universities, other agencies and industry) to talk one-on-one to learn about their research and needs. Detailed information about our community enables us to provide tailored outreach**.** Outreach based on the type of KBase user, research area and scientific questions help us to more productively communicate with users and effectively engage them in KBase events.

Outreach produces webinars and holds tutorials and developer boot camps to provide education on the use of KBase and how to contribute. To accompany such efforts and to promote independent learning, we will develop education materials to assist users in navigating and using KBase effectively. Outreach attends relevant conferences and meetings to present KBase, provide training, and engage in scientific discussions with the community. We inform the community of new developments and outreach events through the website (kbase.us), blogs and social media.

In addition, the KBase team is building partnering relationships with other large projects. For important stakeholders--such as JGI, the BRCS, EMSL and iPlant--we have been co- designing science and software milestones, sharing infrastructure, defining routes for users to migrate between software and data systems, and developing cross-training programs. This allows KBase to maximize impact and relevance to the community and to prioritize its development goals.

KBase is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

163. The KBase Narrative User Interface

Stephen Chan[1], Jason Baumohl[1], Aaron Best[2], Jared Bischof[2], Ben Bowen[1], Tom Brettin[2], Tom Brown[2], Shane Canon[1], John-Marc Chandonia[1], Dylan Chivian[1], Ric Colasanti[2], Neal Conrad[2], Brian Davison[3], Matt DeJongh[6], Paramvir Dehal*[1] (psdehal@lbl.gov), Narayan Desai[2], Scott Devoid[2], Terry Disz[2], Meghan Drake[3], Janaka Edirisinghe[2], Gang Fang[7], José Pedro Lopes Faria[2], Mark Gerstein[7], Elizabeth M. Glass[2], Annette Greiner[1], Dan Gunter[1], James Gurtowski[5], Nomi Harris[1], Travis Harrison[2], Fei He[4], Matt Henderson[1], Chris Henry[2], Adina Howe[2], Marcin Joachimiak[1], Kevin Keegan[2], Keith Keller[1], Guruprasad Kora[3], Sunita Kumari[5], Miriam Land[3], Folker Meyer[2], Steve Moulton[3], Pavel Novichkov[1], Taeyun Oh[8], Gary Olsen[9], Bob Olson[2], Dan Olson[2], Ross Overbeek[2], Tobias Paczian[2], Bruce Parrello[2], Shiran Pasternak[5], Sarah Poon[1], Gavin Price[1], Srivdya Ramakrishnan[5], Priya Ranjan[3], Bill Riehl[1], Pamela Ronald[8], Michael Schatz[5], Lynn Schriml[10], Sam Seaver[2], Michael W. Sneddon[1], Roman Sutormin[1], Mustafa Syed[3], James Thomason[5], Nathan Tintle[6], Will Trimble[2], Daifeng Wang[7], Doreen Ware[5], David Weston[3], Andreas Wilke[2], Fangfang Xia[2], Shinjae Yoo[4], Dantong Yu[4], **Robert Cottingham[3]**, **Sergei Maslov[4]**, **Rick Stevens[2]**, **Adam P. Arkin[1]**

[1] Lawrence Berkeley National Laboratory, Berkeley, CA, [2] Argonne National Laboratory, Argonne, IL, [3] Oak Ridge National Laboratory, Oak Ridge, TN, [4] Brookhaven National Laboratory, Upton, NY, [5] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, [6] Hope College, Holland, MI, [7] Yale University, New Haven, CT, [8] University of California, Davis, CA, [9] University of Illinois at Champaign-Urbana, Champaign, IL, [10] University of Maryland, College Park, MD

http://kbase.us

Project Goals: The KBase project aims to provide the capabilities needed to address the grand challenge of systems biology: to predict and ultimately design biological function. KBase enables users to collaboratively integrate the array of heterogeneous datasets, analysis tools and workflows needed to achieve a predictive understanding of biological systems. It incorporates functional genomic and metagenomic data for thousands of organisms, and diverse tools for (meta)genomic assembly, annotation, network inference and modeling, allowing researchers to combine diverse lines of evidence to create increasingly accurate models of the physiology and community dynamics of microbes and plants. KBase will soon allow models to be compared to observations and dynamically revised. A new prototype Narrative interface lets users create a reproducible record of the data, computational steps and thought process leading from hypothesis to result in the form of interactive publications.

The Narrative interface enables users to easily customize, execute, and share a set of ordered KBase actions that target their specific systems biology hypotheses. The datasets, analyses, and thought processes underlying the execution of these actions are captured in the form of an interactive, dynamic publication called a Narrative. Within each Narrative, users can interleave text and commentary with workflows, so that hypotheses and conclusions can be captured as well as raw results and procedural notes. The Narrative, with the help of Workspaces, maintains provenance and metadata for all data, thereby providing a virtual reference list for all Narratives. A sophisticated social framework will soon allow members of a research team to share, execute, modify, and comment on Narratives at multiple levels of granularity enabled by the KBase infrastructure.

Intrinsically reproducible, Narratives will serve as a new type of publication by (1) explicitly capturing the parameters associated with various algorithms, (2) clearly showing how users applied the algorithms to selected datasets, and (3) transparently recording the process by which resulting output was used to derive biological insight and conclusions. Users who access a Narrative that another researcher has created and shared in KBase not only will be able to see a complete inventory of the data and algorithms underlying a conclusion, they will be able to repeat the computational experiment with the push of a button, even altering parameters to achieve different or improved results.

The Narrative interface is built on top of the open source iPython platform and deeply integrated with the KBase infrastructure. For those seeking to develop new KBase functionality, this allows for rapid integration of KBase-compliant services and KBase data, and leverages the existing community support. Detailed documentation and support is available on the kbase.us website explaining how to use the Narrative interface and the process for adding new features.

KBase is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

164. Collaborative Workspaces in the DOE Systems Biology Knowledgebase

Christopher S. Henry*[1] (chenry@mcs.anl.gov), Jason Baumohl[2], Aaron Best[1], Jared Bischof[1], Ben Bowen[2], Tom Brettin[1], Tom Brown[1], Shane Canon[2], Stephen Chan[2], John- Marc Chandonia[2], Dylan Chivian[2], Ric Colasanti[1], Neal Conrad[1], Brian Davison[3], Matt DeJongh[6], Paramvir Dehal[2], Narayan Desai[1], Scott Devoid[1], Terry Disz[1], Meghan Drake[3], Janaka Edirisinghe[1], Gang Fang[7], José Pedro Lopes Faria[1], Mark Gerstein[7], Elizabeth M. Glass[1], Annette Greiner[2], Dan Gunter[2], James Gurtowski[5], Nomi Harris[2], Travis Harrison[1], Fei He[4], Matt Henderson[2], Adina Howe[1], Marcin Joachimiak[2], Kevin Keegan[1], Keith Keller[2], Guruprasad Kora[3], Sunita Kumari[5], Miriam Land[3], Folker Meyer[1], Steve Moulton[3], Pavel Novichkov[2], Taeyun Oh[8], Gary Olsen[9], Bob Olson[1], Dan Olson[1], Ross Overbeek[1], Tobias Paczian[1], Bruce Parrello[1], Shiran Pasternak[5], Sarah Poon[2], Gavin Price[2], Srivdya Ramakrishnan[5], Priya Ranjan[3], Bill Riehl[2], Pamela Ronald[8], Michael Schatz[5], Lynn Schriml[10], Sam Seaver[1], Michael W. Sneddon[2], Roman Sutormin[2], Mustafa Syed[3], James Thomason[5], Nathan Tintle[6], Will Trimble[1], Daifeng Wang[7], Doreen Ware[5], David Weston[3], Andreas Wilke[1], Fangfang Xia[1], Shinjae Yoo[4], Dantong Yu[4], **Robert Cottingham[3]**, **Sergei Maslov[4]**, **Rick Stevens[1]**, **Adam P. Arkin[2]**

[1]Argonne National Laboratory, Argonne, IL, [2]Lawrence Berkeley National Laboratory, Berkeley, CA, [3]Oak Ridge National Laboratory, Oak Ridge, TN, [4]Brookhaven National Laboratory, Upton, NY, [5]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, [6]Hope College, Holland, MI, [7]Yale University, New Haven, CT, [8]University of California, Davis, CA, [9]University of Illinois at Champaign-Urbana, Champaign, IL, [10]University of Maryland, College Park, MD

http://kbase.us

Project Goals: The KBase project aims to provide the capabilities needed to address the grand challenge of systems biology: to predict and ultimately design biological function. KBase enables users to collaboratively integrate the array of heterogeneous datasets, analysis tools and workflows needed to achieve a predictive understanding of biological systems. It incorporates functional genomic and metagenomic data for thousands of organisms, and diverse tools for (meta)genomic assembly, annotation, network inference and modeling, allowing researchers to combine diverse lines of evidence to create increasingly accurate models of the physiology and community dynamics of microbes and plants. KBase will soon allow models to be compared to observations and dynamically revised. A new prototype Narrative interface lets users create a reproducible record of the data, computational steps and thought process leading from hypothesis to result in the form of interactive publications.

Collaboration, provenance, data consolidation, data standards, and data validation are all major objectives of the KBase project, and the Workspace is the engine being used to drive towards these goals. The Workspace is an online data store in KBase, where the wide variety of data entities (e.g., metabolic models, genome annotations, phenotype data) being produced and consumed by KBase analysis pipelines are stored.

User-generated data including genomes, contigs, phenotype data, and expression data may be loaded directly into the Workspace for analysis by KBase tools. The Workspace stores the many derived data products (e.g., genome annotations, metabolic models) produced by KBase analyses, with all objects generated from an analysis pipeline being interconnected all the way back to the raw data. For example, a

metabolic model is linked to the genome it was constructed from, the genome is linked to the contigs with the genome sequence, and the contigs are linked to the reads they were assembled from. The Workspace also facilitates collaboration by enabling project data, which could encompass thousands of objects, to be rapidly shared with a few select collaborators or—if desired— with the entire research community (for example, during the publication phase of a research project).

The Workspace supports full versioning of all data objects, enabling a researcher to rapidly access, compare, and restore prior versions of all objects stored in the Workspace. Provenance is also provided for data in the Workspace, providing researchers with detailed information on the analysis pipelines and parameters that produced each data object. We anticipate that these features will greatly facilitate the evaluation, validation, and replication of even complex systems biology studies.

All objects stored in the Workspace are typed, with each data type specified in detail (e.g., required fields, field types, field indexing). For example, a Genome object in the workspace must include a scientific name, a domain, and a list of features with functional annotations and locations on the chromosome. Data objects are validated against these specifications, enabling the Workspace to enforce data standards, as well as metadata requirements. The Workspace also provides a rich API for users to specify their own new data types, supporting the rapid integration of new tools, data, and analysis pipelines by the KBase user community. In this poster, we will highlight the functionality encompassed in the Workspace, we will explore the typed objects already available within the Workspace, and we will demonstrate how these typed objects connect to the many analysis tools and pipelines already available in KBase.

KBase is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

165. The KBase User Experience

Adam P. Arkin[1], Jason Baumohl[1], Aaron Best[2], Jared Bischof[2], Ben Bowen[1], Tom Brettin[2], Tom Brown[2], Shane Canon[1], Stephen Chan[1], John-Marc Chandonia[1], Dylan Chivian[1], Ric Colasanti[2], Neal Conrad[2],, Brian Davison[3], Matt DeJongh[6], Paramvir Dehal[1], Narayan Desai[2], Scott Devoid[2], Terry Disz[2], Meghan Drake[3], Janaka Edirisinghe[2], Gang Fang[7], José Pedro Lopes Faria[2], Mark Gerstein[7], Elizabeth M. Glass[2], Annette Greiner[1], Dan Gunter[1], James Gurtowski[5], Nomi Harris[1], Travis Harrison[2], Fei He[4], Matt Henderson[1], Chris Henry[2], Adina Howe[2], Marcin Joachimiak[1], Kevin Keegan[2], Keith Keller[1], Guruprasad Kora[3], Sunita Kumari[5], Miriam Land[3],, Folker Meyer[2], Steve Moulton[3], Pavel Novichkov[1], Taeyun Oh[8], Gary Olsen[9], Bob Olson[2], Dan Olson[2], Ross Overbeek[2], Tobias Paczian[2], Bruce Parrello[2], Shiran Pasternak[5], Sarah Poon[1], Gavin Price[1], Srivdya Ramakrishnan[5], Priya Ranjan[3], Bill Riehl[1], Pamela Ronald[8], Michael Schatz[5], Lynn Schriml[10], Sam Seaver[2], Michael W. Sneddon[1], Roman Sutormin[1], Mustafa Syed[3], James Thomason[5], Nathan Tintle[6], Will Trimble[2], Daifeng Wang[7], Doreen Ware[5], David Weston[3], Andreas Wilke[2], Fangfang Xia[2], Shinjae Yoo[4], Dantong Yu[4], **Robert Cottingham***[3] (cottinghamrw@ornl.gov)**, Sergei Maslov**[4] **, Rick Stevens**[2]

[1] Lawrence Berkeley National Laboratory, Berkeley, CA, [2] Argonne National Laboratory, Argonne, IL, [3] Oak Ridge National Laboratory, Oak Ridge, TN, [4] Brookhaven National Laboratory, Upton, NY, [5] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, [6]Hope College, Holland, MI, [7] Yale University, New Haven, CT, [8] University of California, Davis, CA, [9] University of Illinois at Champaign-Urbana, Champaign, IL, [10]University of Maryland, College Park, MD

http://kbase.us

Project Goals: The KBase project aims to provide the computational capabilities needed to address the grand challenge of systems biology: to predict and ultimately design biological function. KBase enables users to collaboratively integrate the array of heterogeneous datasets, analysis tools and workflows needed to achieve a predictive understanding of biological systems. It incorporates functional genomic and metagenomic data for thousands of organisms, and diverse tools including (meta)genomic assembly, annotation, network inference and modeling, thereby allowing researchers to combine diverse lines of evidence to create increasingly accurate models of the physiology and community dynamics of microbes and plants. KBase will soon allow models to be compared to observations and dynamically revised. A new prototype Narrative interface lets users create a reproducible record of the data, computational steps and thought process leading from hypothesis to result in the form of interactive publications.

KBase provides several ways for scientists to access the system's functions and data including programmatically via a number of different programming languages; via a web-based scripting interface (IRIS); and using our prototype graphical user interface, the Narrative Interface (see poster on IRIS and the Narrative Interface). This poster introduces how to register as a KBase user; find documentation and tutorials; search and analyze KBase data; upload your own data; and use the first generation Narrative Interface. We explain the emerging model of how users interact with the system and with each other: how their data can be uploaded and shared as desired; how users can start projects and add narratives to them; and how users can use the Narrative Interface together with workspace (see related posters) to perform complex computational biological, human-in-the-loop workflows on heterogeneous biological data and share their processes and thoughts with the community.

KBase is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

166. Design of KBase Infrastructure

Thomas Brettin*[1] (brettin@cels.anl.gov), Daniel Olson[1], Jason Baumohl[2], Aaron Best[3], Jared Bischof[1], Ben Bowen[2], Tom Brown[1], Shane Canon[1], Stephen Chan[2], John-Marc Chandonia[2], Dylan Chivian[2], Ric Colasanti[1], Neal Conrad[1], Brian Davison[4], Matt DeJongh[3], Paramvir Dehal[2], Narayan Desai[1], Scott Devoid[1], Terry Disz[1], Meghan Drake[4], Janaka Edirisinghe[1], Gang Fang[7], José Pedro Lopes Faria[1], Mark Gerstein[7], Elizabeth M. Glass[1], Annette Greiner[2], Dan Gunter[2], James Gurtowski[6], Nomi Harris[2], Travis Harrison[1], Fei He[5], Matt Henderson[2], Chris Henry[1], Adina Howe[1], Marcin Joachimiak[2], Kevin Keegan[1], Keith Keller[2], Guruprasad Kora[4], Sunita Kumari[6], Miriam Land[4], Folker Meyer[1], Steve Moulton[4], Pavel Novichkov[2], Taeyun Oh[8], Gary Olsen[9], Bob Olson[1], Dan Olson[1], Ross Overbeek[1], Tobias Paczian[1], Bruce Parrello[1], Shiran Pasternak[6], Sarah Poon[2], Gavin Price[2], Srivdya Ramakrishnan[6], Priya Ranjan[4], Bill Riehl[2], Pamela Ronald[8], Michael Schatz[6], Lynn Schriml[10], Sam Seaver[1], Michael W. Sneddon[2], Roman Sutormin[2], Mustafa Syed[4], James Thomason[6], Nathan Tintle[3], Will Trimble[1], Daifeng Wang[7], Doreen Ware[5,6], David Weston[4], Andreas Wilke[1], Fangfang Xia[1], Shinjae Yoo[5], Dantong Yu[5], **Robert Cottingham[4], Sergei Maslov[5], Rick Stevens[1], Adam P. Arkin[2]**

[1]Argonne National Laboratory, Argonne, IL, [2]Lawrence Berkeley National Laboratory, Berkeley, CA, [3]Hope College, Holland, MI, [4]Oak Ridge National Laboratory, Oak Ridge, TN, [5]Brookhaven National Laboratory, Upton, NY, [6]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, [7]Yale University, New Haven, CT, [8]University of California, Davis, CA, [9]University of Illinois at Champaign-Urbana, Champaign, IL, [10]University of Maryland, College Park, MD

http://kbase.us

Project Goals: The KBase project aims to provide the capabilities needed to address the grand challenge of systems biology: to predict and ultimately design biological function. KBase enables users to collaboratively integrate the array of heterogeneous datasets, analysis tools and workflows needed to achieve a predictive understanding of biological systems. It incorporates functional genomic and metagenomic data for thousands of organisms, and diverse tools for (meta)genomic assembly, annotation, network inference and modeling, allowing researchers to combine diverse lines of evidence to create increasingly accurate models of the physiology and community dynamics of microbes and plants. KBase will soon allow models to be compared to observations and dynamically revised. A new prototype Narrative interface lets users create a reproducible record of the data, computational steps and thought process leading from hypothesis to result in the form of interactive publications.

At the core of the KBase architecture is a set of rich data models and stores, scalable computing, and workflow management. Our KBase physical infrastructure is built on the successes of DOE investment in our national scientific cyber-infrastructure and therefore leverages enormous intellectual resources present in the DOE community. Building on ESNet allows us to construct a wide area network between the partner labs that enables a virtual hardware infrastructure. Our use of cloud-computing supports development of new tools and provides compute resources for production services. The acceptance of virtualization technology is growing, and the use of machine images produced by others is already visible in our core services. Additionally, machine images are now provided which contain multiple components of the KBase infrastructure and services. Cluster Computing has long been a critical part of biological data

analysis. In collaboration with computing centers created by the Office of Advanced Computing Research such as NERSC, our underlying cluster services can leverage these resources and scale to meet needs.

KBase aims to power the next wave of biological research in DOE and beyond. Enabling these capabilities requires a software and hardware infrastructure that is integrated, extensible, and scalable. The architecture is designed to meet these needs and support user functionality to visualize data, create models or design experiments based on KBase- generated suggestions.

KBase is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

167. Developing for KBase

Thomas Brettin*[1] (brettin@cels.anl.gov), **Rick Stevens**[1], Jason Baumohl[2], Aaron Best[3], Jared Bischof[1], Ben Bowen[2], Tom Brown[1], Shane Canon[1], Stephen Chan[2], John-Marc Chandonia[2], Dylan Chivian[2], Ric Colasanti[1], Neal Conrad[1], Brian Davison[4], Matt DeJongh[3], Paramvir Dehal[2], Narayan Desai[1], Scott Devoid[1], Terry Disz[1], Meghan Drake[4], Janaka Edirisinghe[1], Gang Fang[7], José Pedro Lopes Faria[1], Mark Gerstein[7], Elizabeth M. Glass[1], Annette Greiner[2], Dan Gunter[2], James Gurtowski[6], Nomi Harris[2], Travis Harrison[1], Fei He[5], Matt Henderson[2], Chris Henry[1], Adina Howe[1], Marcin Joachimiak[2], Kevin Keegan[1], Keith Keller[2], Guruprasad Kora[4], Sunita Kumari[6], Miriam Land[4], Folker Meyer[1], Steve Moulton[4], Pavel Novichkov[2], Taeyun Oh[8], Gary Olsen[9], Bob Olson[1], Dan Olson[1], Ross Overbeek[1], Tobias Paczian[1], Bruce Parrello[1], Shiran Pasternak[6], Sarah Poon[2], Gavin Price[2], Srivdya Ramakrishnan[6], Priya Ranjan[4], Bill Riehl[2], Pamela Ronald[8], Michael Schatz[6], Lynn Schriml[10], Sam Seaver[1], Michael W. Sneddon[2], Roman Sutormin[2], Mustafa Syed[4], James Thomason[6], Nathan Tintle[3], Will Trimble[1], Daifeng Wang[7], Doreen Ware[5,6], David Weston[4], Andreas Wilke[1], Fangfang Xia[1], Shinjae Yoo[5], Dantong Yu[5], **Robert Cottingham**[4], **Sergei Maslov**[5], **Adam P. Arkin**[2]

[1]Argonne National Laboratory, Argonne, IL, [2]Lawrence Berkeley National Laboratory, Berkeley, CA, [3]Hope College, Holland, MI, [4]Oak Ridge National Laboratory, Oak Ridge, TN, [5]Brookhaven National Laboratory, Upton, NY, [6]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, [7]Yale University, New Haven, CT, [8]University of California, Davis, CA, [9]University of Illinois at Champaign-Urbana, Champaign, IL, [10]University of Maryland, College Park, MD

http://kbase.us

Project Goals: The KBase project aims to provide the capabilities needed to address the grand challenge of systems biology: to predict and ultimately design biological function. KBase enables users to collaboratively integrate the array of heterogeneous datasets, analysis tools and workflows needed to achieve a predictive understanding of biological systems. It incorporates functional genomic and metagenomic data for thousands of organisms, and diverse tools for (meta)genomic assembly, annotation, network inference and modeling, allowing researchers to combine diverse lines of evidence to create increasingly accurate models of the physiology and community dynamics of microbes and plants. KBase will soon allow models to be compared to observations and dynamically revised. A new prototype Narrative interface lets users create a reproducible record of the data, computational steps and thought process leading from hypothesis to result in the form of interactive publications.

One of the key operating principles of KBase is to allow the scientific community to incorporate their own new algorithms into the system to make them available to others easily; to avail themselves of the KBase computational architecture; and to make use of the KBase data sources. The KBase team aims to make this process simple and to provide an easy route for dissemination of new tools and comparison to existing tools in a common framework.

To achieve these goals KBase system design is firmly rooted in our service-oriented architecture and in our commitment to open source development and distribution models. These aspects of the project have been with the system since its inception, and will continue to be a guiding force for future work.

The KBase system design is based on several sound best practice principles including consistent code use, code re-use, and the decoupling of modular system components. We have established standard software engineering processes for version control, software and data builds, testing, QA/QC, deployments and releases. These enable the deployment of a large number of services by a relatively small release engineering team.

To prepare for potential future services contributed by the community, we provide developer training materials on our website as well in hands-on developer training sessions called bootcamps. In the future, we plan to offer a wider range of bootcamps and webinars to target different types of developers and different scientific focus areas. In the meantime, prospective developers and computational biologists can find information about KBase service design at http://kbase.us/developer-zone/ .

KBase is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

168. Command Line Environment in the DOE Systems Biology Knowledgebase

Bob Olson[1], James Thomason[2], Thomas Brettin*[1] (brettin@cels.anl.gov), Jason Baumohl[4], Aaron Best[3], Jared Bischof[1], Ben Bowen[4], Tom Brown[1], Shane Canon[4], Stephen Chan[4], John-Marc Chandonia[4], Dylan Chivian[4], Ric Colasanti[1], Neal Conrad[1], Brian Davison[5], Matt DeJongh[3], Paramvir Dehal[4], Narayan Desai[1], Scott Devoid[1], Terry Disz[1], Meghan Drake[5], Janaka Edirisinghe[1], Gang Fang[7], José Pedro Lopes Faria[1], Mark Gerstein[7], Elizabeth M. Glass[1], Annette Greiner[4], Dan Gunter[4], James Gurtowski[2], Nomi Harris[4], Travis Harrison[1], Fei He[6], Matt Henderson[4], Chris Henry[1], Adina Howe[1], Marcin Joachimiak[4], Kevin Keegan[1], Keith Keller[2], Guruprasad Kora[5], Sunita Kumari[2], Miriam Land[5], Folker Meyer[1], Steve Moulton[5], Pavel Novichkov[4], Taeyun Oh[8], Gary Olsen[9], , Dan Olson[1], Ross Overbeek[1], Tobias Paczian[1], Bruce Parrello[1], Shiran Pasternak[2], Sarah Poon[4], Gavin Price[4], Srivdya Ramakrishnan[2], Priya Ranjan[5], Bill Riehl[4], Pamela Ronald[8], Michael Schatz[2], Lynn Schriml[10], Sam Seaver[1], Michael W. Sneddon[4], Roman Sutormin[4], Mustafa Syed[5], James Thomason[2], Nathan Tintle[3], Will Trimble[1], Daifeng Wang[7], Doreen Ware[2], David Weston[5], Andreas Wilke[1], Fangfang Xia[1], Shinjae Yoo[6], Dantong Yu[6], **Robert Cottingham[5]**, **Sergei Maslov[6]**, **Rick Stevens[1]**, **Adam P. Arkin[4]**

[1]Argonne National Laboratory, Argonne, IL, [2]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, [3]Hope College, Holland, MI, [4]Lawrence Berkeley National Laboratory, Berkeley, CA, [5]Oak Ridge National Laboratory, Oak Ridge, TN, [6]Brookhaven National Laboratory, Upton, NY, [7]Yale University, New Haven, CT, [8]University of California, Davis, CA, [9]University of Illinois at Champaign-Urbana, Champaign, IL, [10]University of Maryland, College Park, MD

http://kbase.us

Project Goals: The KBase project aims to provide the capabilities needed to address the grand challenge of systems biology: to predict and ultimately design biological function. KBase enables users to collaboratively integrate the array of heterogeneous datasets, analysis tools and workflows needed to achieve a predictive understanding of biological systems. It incorporates functional genomic and metagenomic data for thousands of organisms, and diverse tools for (meta)genomic assembly, annotation, network inference and modeling, allowing researchers to combine diverse lines of evidence to create increasingly accurate models of the physiology and community dynamics of microbes and plants. KBase will soon allow models to be compared to observations and dynamically revised. A new prototype Narrative interface lets users create a reproducible record of the data, computational steps and thought process leading from hypothesis to result in the form of interactive publications.

KBase provides a variety of user experiences targeted to meet the needs of a diverse user- base in the systems biology community. A rich programmatic API is provided in wide variety of languages for maximum performance and flexibility for computationally advanced users. The graphical Narrative interface provides an intuitive visual experience to facilitate hands-on interactive systems biology. Between these is the command line interface, providing a powerful but simple command-driven interface capable of high- throughput interaction with KBase tools and data. The online IRIS web application (http://iris.kbase.us) permits users to run commands in KBase from any computer with web access without installing any local software. It offers a wide variety of online tutorials, scripting, and data visualization features. The installable KBase client for Mac and Ubuntu Linux enables users to run KBase commands from a shell on their own computers, supporting seamless integration with locally installed data and tools.

The installable client also includes API libraries in numerous programming languages (e.g., Java, JavaScript, Perl, Python, JSON RPC), enabling users to program directly against the KBase API from their own computers and workstations. In this way, the existing computational infrastructure in any academic lab can seamlessly integrate KBase tools and data into local data analysis pipelines.

The KBase command line environment now offers a powerful suite of over 800 distinct commands that encompass genome assembly, annotation, metabolic modeling, expression data analysis, phenotype analysis, and GWAS. Since the initial release in February 2013, the KBase command line environments have collectively been applied by over 300 unique users to: (i) annotate or import over 34K genomes, (ii) construct over 40K metabolic models, and (iii) run over 110K FBA simulations. In this poster, we will highlight all command-line clients and tools available in KBase, with example workflows for genome assembly, annotation, metabolic modeling, and phenotype reconciliation.

KBase is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

169. RDP: Data and Tools for Studying Structure and Function of Microbial Communities

Benli Chai[1*] (chaibenl@msu.edu), Jordan A. Fish[1], Qiong Wang[1], Yanni Sun[3], C. Titus Brown[2,3], James M. Tiedje[1,2] and **James R. Cole**[1]

[1]Center for Microbial Ecology, [2]Microbiology and Molecular Genetics, [3]Computer Science and Engineering, Michigan State University, East Lansing, MI 48824

http://rdp.cme.msu.edu
http://fungene.cme.msu.edu

Project Goals: The Ribosomal Database Project (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, greenhouse gas production, and environmental bioremediation.

In the current release (October 2013), RDP offers 2,809,406 aligned and annotated quality- controlled public bacterial, archaeal, and fungal rRNA sequences. Over the past year, the RDP website was visited, on average, by **10,000 users** (unique IP) in **22,000 analysis sessions** each month. RDP recently released new alignments of bacterial and archaeal 16S rRNA gene sequence alignments and a fungal 28S gene sequence alignment using the latest Infernal 1.1 aligner with specially-tuned covariance models (CMs). As part of RDP's efforts to support the fungal research community following the release of the Fungal 28S RDP Classifier (Liu et al., 2011), most RDP tools, including the RDP Hierarchy Browser, Sequence Match, Probe Match, and RDPipeline, have been updated to work with the new fungal 28S sequences.

The **new RDPipeline** (Cole et al., 2013) expands upon our existing high-throughput tool offerings and is designed to accommodate the latest benchtop high-throughput sequencing technologies. RDPipeline integrates with researchers' existing *my*RDP accounts for streamlined analysis job submission and monitoring. The new RDPipeline includes both improved performance in optimizing back-end job load distribution and increased capacity for larger datasets. It also provides additional user-friendly features such as a "my jobs" page for each user to track the job status, download results, and retrieve process parameters for past analysis tasks submitted to RDPipeline. Other enhancements include optimized paired-end read assembly (Assembler). Tested on Illumina MiSeq paired-end data, this tool outperformed its peers in selectively filtering out error-containing sequence reads, and also better handles different types of paired-end overlaps. A new data validation mechanism implemented in RDPipeline provides feedback if incorrect data input is submitted before an analysis job starts running--a feature especially valuable for inexperienced users.

FunGene, RDP's Functional Gene Pipeline and Repository (Fish et al., 2013), offers databases of many common ecofunctional genes and proteins, as well as integrated tools that allow researchers to browse these collections and choose subsets for further analysis, build phylogenetic trees, test primers and probes for coverage, and download aligned sequences. Additional FunGene tools are specialized to process coding gene amplicon data. For example, **RDP FrameBot** (Wang et al., 2013) produces frameshift-corrected protein and DNA sequences from raw reads while finding the most closely related protein reference sequence. These tools can help provide better insight into microbial communities by directly

studying key genes involved in important ecological processes. Over the past year, RDP FunGene **usage increased 1.8-fold** to 1997 researchers in 2823 analysis sessions per month.

Porting RDP tools to KBase will provide an opportunity for RDP to reach the broader research community. To learn how to develop RDP tools into KBase services (modules). RDP staff hosted a two-day **KBase Bootcamp at MSU** in May 2013. Two instructors from Argonne National Laboratory, four RDP staff members and one additional student from MSU, three graduate students and one post-doc from the University of Oklahoma, and one graduate student from Georgia Institute of Technology all participated in the KBase Bootcamp. Following the bootcamp all nine participants completed a survey. Overall, they all agreed or agreed strongly with the statement "I found the KBase bootcamp useful". Five of the participants said they would apply for KBase developer accounts, while the other four indicated they would be developing tools that will use KBase services.

In addition to web-based services, RDP now distributes many of its process/analysis tools as stand-alone, open-source versions through https://github.com/rdpstaff. Tutorials are provided to guide researchers through the otherwise complex data processing steps in well-defined, task- oriented workflows with detailed instructions. RDP's mission includes user support; email rdpstaff@msu.edu or call +1(517)432-4997.

References:

Cole J.R., Wang Q., Fish J.A., Chai B., McGarrell D.M., Sun Y., Brown C.T., Porras-Alfaro A., Kuske C.R., and Tiedje J.M. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucl. Acids Res.* 42 (Database issue): D633-D642 (2014).

Fish J.A., Chai B., Wang Q., Sun Y., Brown C.T., Tiedje J.M., and Cole J.R. FunGene: the Functional Gene Pipeline and Repository. *Front. Microbiol.* 4:291 (2013).

Liu K.L., Porras-Alfaro A., Kuske C.R., Eichorst S.A., and Xie G. Accurate, rapid taxonomic classification of fungal large-subunit rRNA genes. *Appl. Environ. Microbiol.* 78:1523- 1533 (2012).

Wang Q., Quensen III J.F., Fish J.A., Lee T.-K., Sun Y., Tiedje J.M., and Cole J.R. Ecological patterns of nifH genes in four terrestrial climatic zones explored with targeted metagenomics using FrameBot, a new informatics tool. *mBio* 4:e00592-13 (2013).

Ribosomal Database Project (RDP) is supported by the Office of Science (Biological and Environmental Research), U.S. Department of Energy grant DE-FG02-99ER62848.

170. Semantic Index of Phenotypic and Genotypic Data

Charles T. Parker* (chuck.t.parker@namesforlife.com),[1] Nenad Krdzavac,[2] Kevin Petersen,[2] Amber Roberts,[2] Grace Rodriguez,[1] and **George M. Garrity**[1,2]

[1]NamesforLife, LLC; East Lansing, Michigan and [2]Michigan State University; East Lansing, Michigan

http://ontology.namesforlife.com

Project Goals: The goal of this project is to develop a semantic data resource that can serve as a basis for predictive modeling of microbial phenotype. The core technical objectives are twofold: (1) to build a database of normalized phenotypic descriptions (observational data) using the primary taxonomic literature of bacterial and archaeal type strains, and (2) to construct an ontology with reasoning capabilities to make accurate phenotypic and environmental inferences based on that data. This project is tightly coupled with ongoing DOE projects (the Genomic Encyclopedia of Bacteria and Archaea, the Microbial Earth Project, the Community Sequencing Project) and with two key publications, *Standards in Genomic Sciences* (SIGS) and the *International Journal of Systematic and Evolutionary Microbiology* (IJSEM).

The DOE Systems Biology Knowledgebase (KBase) was envisioned to provide a framework for modeling dynamic cellular processes of microorganisms, plants and metacommunities. The KBase will enable rapid iteration of experiments that draw on a wide variety of data and allow researchers to infer how cells and communities respond to natural or induced perturbations, and ultimately to predict outcomes.

Predictive models rely on high quality input data, but not all data are of similar quality nor are they amenable to computational analysis without extensive cleaning, interpretation and normalization. Key among those needed to make the KBase fully operational are phenotypic data, which are more complex than sequence data, occur in a wide variety of forms, often use complex and non-uniform descriptors and are scattered about specialized databases and scientific and technical literature. Incorporating phenotypic information into the KBase requires expertise in harvesting, modeling and interpreting these data.

The Semantic Index of Phenotypic and Genotypic Data will address this problem by providing a resource of reference phenotypic data for all validly published type strains of *Bacteria* and *Archaea*, based on concepts and observational data drawn from the primary taxonomic literature. In the Phase I project we developed software to construct and analyze a corpus of this literature and to extract putative feature domain vocabularies comprising approximately 40,000 candidate phenotypic terms used in 5,750 (now expanded to 11,018 of 17,793 total) new and emended descriptions of the 11,492 distinct type strains of *Bacteria* and *Archaea*. In Phase II, these vocabularies are serving as the basis for developing a phenotypic ontology, a repository of phenotypic data and normalized phenotypic descriptions for each species. Many of the phenotypes applied to microbes describe a combination of quantitative environmental conditions and qualitative growth and metabolic capabilities. Such terms are challenging to implement in query systems due to their context-based interpretations and conceptual overlap across multiple feature domains. In our past year of research, we have discovered novel design patterns for ontology development [1] that address these problems and remove barriers to machine reasoning over these complex terms.

In our current work, we are applying these novel modeling techniques to encode axioms for automatically resolving ambiguity attributed to the semantic equivalence and imprecision of phenotypic terms arising in literature [2]. These axioms will enable reasoners to make correct inferences over the ontology and phenotypic data. We are also developing a query and retrieval service linked to the ontology that will provide researchers with consistent, accurate interpretations that are usable for predictive modeling and in other research and commercial applications.

Several additional software components were developed to overcome technical barriers that arose during this project [3]. Originally implemented as command-line utilities for vocabulary extraction, annotation and document analysis, we are now developing these into a commercial semantic desktop application for document/corpus analysis and for bootstrapping terminology/ontology development.

Publications

1. Parker, CT, Garrity, GM and Krdzavac, NB. Systems and Methods for Inferring Properties of Objects in the Absence of Direct Observations. U.S. Provisional Patent Application No. 61/880,244. Filed September 20, 2013. Washington, DC: *U.S. Patent and Trademark Office*.
2. Krdzavac, NB, Parker, CT, Garrity, GM: An Observation Ontology. *Bioinformatics*. 2014, *Under review*.
3. Garrity GM: The NamesforLife Semantic Index of Phenotypic and Genotypic Data: Phase I Final Technical Report [Internet]. 2012, doi:10.1601/report.sc0006191p1

Funding for this project was provided through the DOE SBIR/STTR program (DE-SC0006191). Public funding for development of the NamesforLife infrastructure was received from the DOE SBIR/STTR program (DE-FG02-07ER86321), the Michigan Small Business Technology Development Corporation, the Michigan Strategic Fund, and the Michigan Universities Commercialization Initiative.

171. Biological and Environmental Research Information System: A Multifaceted Approach to DOE Systems Research Communication

Jennifer L. Bownas, Kris S. Christen, Holly L. Haun, Brett Hopwood, Marilyn E. Langston, Sheryl A. Martin, Marissa D. Mills, Judy M. Wyrick, and **Betty K. Mansfield*** (mansfieldbk@ornl.gov)

Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://genomicscience.energy.gov
http://doesbr.org/

Project Goals: Develop and distribute programmatic materials to help build the multidisciplinary community needed to advance systems research for DOE energy and environmental missions. The Biological and Environmental Research Information System (BERIS) group works with program managers and the scientific community to help develop and communicate key scientific and technical concepts for discourse within the scientific community and the public sector.

Concerted communication is key to progress in cutting-edge science and public accountability. BERIS goals focus on three objectives: (1) facilitate science planning, research, and communication; (2) inform a broader audience about Department of Energy (DOE) research projects, progress, and significance to science and society; and (3) respond to outreach and information exchange needs of related DOE projects. Over the past 24 years, BERIS has focused on presenting all facets of genomics research for DOE's Office of Science. The materials produced by our group have helped ensure that scientists can participate in and reap the bounty of the genome revolution, have facilitated the training of new generations of students in genomics and systems biology, and have aided the public in making informed decisions regarding genetics issues.

In 2009, our scope was extended to include more programs within the Office of Biological and Environmental Research (BER), which conducts frontier research in climate, subsurface biogeochemistry, and genome science within the Office of Science. BER research explores scientific complexity at temporal and spatial scales requiring contributions from teams of interdisciplinary scientists, thereby necessitating an unprecedented integrative approach both to the science and to research communication strategies. Because each scientific discipline has different perspectives and languages, effective communication to help foster information flow across disciplines and translation of scientific discovery into appropriate DOE mission areas is critical to BER's success. Our work with DOE staff and the research community has resulted in the production and dissemination of information in various formats: technical reports, roadmaps, websites, brochures, databases, technical compilations, presentations, exhibits for scientific meetings, topical text, graphics, and posters. We maintain the searchable BER Research Highlights database (*public.ornl.gov/hgmis/bernews/*). We also assist with outreach efforts of DOE grantees—especially the Bioenergy Research Centers, Joint Genome Institute, Environmental Molecular Sciences Laboratory, KBase, and Atmospheric Radiation Measurement Climate Research Facility—to help increase their reach and impact.

Examples of Recent Work

DOE Biological and Environmental Research Advisory Committee

- *BER Virtual Laboratory: Innovative Framework for Biological and Environmental Grand Challenges* (February 2013)

DOE BER Biological Systems Science Division—completed and ongoing

- *Applications of New DOE National User Facilities in Biology Workshop Report* (February 2012)
- *DOE Genomic Science Awardee Meeting X, February 26–29, 2012*
- *Biosystems Design* workshop report (April 2012)
- *DOE Joint Genome Institute Strategic Planning for Genomic Sciences* workshop report (September 2012)
- *DOE User Facilities: Advanced Technologies for Biology, Structural Biology* brochure (updated 2012)
- *Revealing the Role of Microbial Communities in Carbon Cycling* brochure (October 2013)
- *Plant Feedstock Genomics for Bioenergy Joint USDA-DOE Awards* (January 2014)
- *U.S. Department of Energy's Bioenergy Research Centers: An Overview of the Science* (update in progress)
- Genomic Science website ongoing (*genomicscience.energy.gov*)
    - Key sections: Annual research summaries, systems biology computing, advanced biofuels, DOE Bioenergy Research Centers, carbon cycling and climate, DOE-USDA Plant Feedstock Genomics for Bioenergy, biosystems design, and user facilities enabling science

DOE BER Climate and Environmental Sciences Division—completed and ongoing

- *Research Priorities for Tropical Ecosystems under Climate Change* workshop report (October 2012)
- *Terrestrial Ecosystem Science* brochure (July 2012)
- *Spruce and Peatland Responses Under Climatic and Environmental Change* brochure (July 2012)
- *Next-Generation Ecosystem Experiment: Arctic Landscapes* brochure (July 2012)
- *Climate and Environmental Sciences Division Overview* brochure (Summer 2012)
- *Environmental Molecular Sciences Laboratory* brochure (Summer 2012)
- Subsurface Biogeochemical Research website ongoing (doesbr.org)
    - Annual summary (Summer 2013)
    - Research bibliography (January 2014)

Related Information

The DOE BER Human Genome Project Information website was updated for archival use in 2013. When this BERIS-created and maintained site was active during the HGP era, it was the central site for research community and public access to the HGP. The site will continue to enable HGP information flow for use for educational and research purposes enabling a "peek" into science and program planning history for "biology's moon shot." http://web.ornl.gov/sci/techresources/Human_Genome/index.shtml

Awards

Awards of excellence for *Applications of New DOE National User Facilities in Biology Workshop Report* and Merit for the Subsurface Biochemical Research website in the 2012-2013 Society for Technical Communication (STC) Summit Competition. Cumulatively, BERIS has received more than 75 STC awards, including 8 international.

The Biological and Environmental Research Information System (DOE ERKP 153) at Oak Ridge National Laboratory is supported by the U.S. Department of Energy Office of Biological and Environmental Research in the DOE Office of Science.

172. Algorithms Enabling Metaproteomics

Seungjin Na[1], Grant Fujimoto[2], Sangtae Kim[2], Brian LaMarche[2], Nuno Bandeira[1], Stephen Callister[2], **Samuel H Payne**[2]* (samuel.payne@pnnl.gov)

1 University of California, San Diego; 2 Pacific Northwest National Laboratory

http://omics.pnl.gov

Project Goals: This project is focused on improving algorithms 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.

Unlike laboratory experiments that focus on one single organism, biofuel production is often explored in experiments with either natural or synthetic communities composed of numerous organisms. To understand the active state of these communities it is advantageous to assay the proteome as opposed to the genome. Unfortunately current algorithms for peptide/spectrum matching are reliant on protein databases. The practical implication of this dependence for proteomics of communities (metaproteomics) is that the algorithms often fail to identify sufficient number of peptides and proteins.

The major goal of this project is to improve peptide and protein identification in community proteomics datasets. Thus, our algorithms must be designed to operate with millions of protein sequences (e.g. for an experiment with very deep metagenomic sequencing) or without sequences at all (for a project without matched metagenomics). To this end, we have designed several new algorithms: SpectralNetworks with AlignGF and tag-filtering, Informed Quantitation of Top-down proteomics, and alignment via spectrum/spectrum matches.

SpectralNetworks has been effectively used to discover protein post-translational modification and mutations. Although it proved to be useful in analyzing small spectral datasets, it poses challenges in its reliability and scalability for large spectral datasets. The key problem in SpectralNetworks is distinguishing between correct and incorrect spectral pairs. AlignGF (Alignment generating function) can be used to calculate theoretical distributions of scores for all possible alignments between two spectra, and provide a rigorous solution to the problem of computing statistical significance of spectral alignments. A binomial distribution is assumed to regulate the probability of aligned (or matching) peaks among total peaks of a spectrum. The implementation of AlignGF has dramatically improved the sensitivity of SpectralNetworks in large datasets, achieving ~75% sensitivity at ~90% precision for spectral pairs of mutated peptides. A second improvement to the SpectralNetworks algorithm is the inclusion of tag-based filtering. The previous algorithm needed to compute all pairwise spectral alignments for the construction of spectral networks, requiring significant computational resource.

Properly aligned spectra, from peptides with similar sequences, share a substring of the sequence. Therefore, we use spectrum tag generation to filter which spectral pairs can be aligned. The tag-based filtering currently accelerates SpectralNetwork generation by 200x with little loss in sensitivity. As mass spectrometers continue to improve, the analysis of intact proteins is becoming more common. Top-down proteomics is advantageous as it can fully characterize a protein with all its post-translational modifications which are critical for function. We are developing algorithms for top-down, in anticipation of their use in community studies of biofuels, particularly for simplified or synthetic communities. Our

current algorithmic progress is with a quantitation tool called IQ, which accurately measures the abundance of each individual proteoform in a sample.

As a final algorithm for metaproteomics, we are developing methods to align mass spectrometry datasets without the requirement of peptide identifications. LC-MS alignment is crucial when researchers are looking for common features across several LC-MS/MS runs. Instrumental variation, especially in the chromatographic dimension can hinder the ability to find the same species. When genomic information is incomplete or missing (e.g. metaproteomic applications) existing alignment techniques that require peptide identifications for alignment may fail. Using MS/MS spectra to define these anchor points can provide a highly confident alignment, even when there are few anchor points as might exist in rapidly changing communities.

173. Metabolic RouteSearch in Pathway Databases and Expansion of MetaCyc

Mario Latendresse,[1] Ron Caspi,[1] **Peter D. Karp[1]**

[1]Bioinformatics Research Group, SRI International, Menlo Park, CA.

Project Goals:

- **Extend Pathway Tools software to support pathway design**
- **Extend the MetaCyc pathway database to support metabolic engineering and pathway analysis of bioenergy-related organisms**

The RouteSearch Software for Metabolic Engineering

A key computational problem in metabolic engineering is finding efficient metabolic routes from a source to a target compound in genome---scale reaction networks, potentially considering the addition of new reactions. Efficiency can be based on many factors, such as route lengths, atoms conserved, and the number of new reactions added to the route. Fast algorithms are needed to systematically search these large genome-scale reaction networks.

We present the algorithm used in the new RouteSearch tool within the Pathway Tools software. This algorithm is based on a general Branch-and-Bound search and involves constructing a network of atom mappings to facilitate efficient searching. As far as we know, it is the first published algorithm that finds guaranteed optimal routes where atom conservation is part of the optimality criteria. RouteSearch includes a graphical user interface that speeds user understanding of its search results. We evaluated the algorithm on five example metabolic-engineering problems from the literature; for one problem the published solution was equivalent to the optimal route found by RouteSearch; for the remaining four problems, RouteSearch found the published solution as one of its best-scored solutions. These problems were each solved in less than five seconds of computational time.

In addition, RouteSearch can be used to search for reaction paths between metabolites in an organism's metabolic network without the addition of new engineered reactions.

RouteSearch is accessible at BioCyc.org by using the menu command Metabolism ---> Route Search, and by downloading Pathway Tools. Download from: http://biocyc.org/download.shtml

Expansion of the MetaCyc Pathway Database

The goal of the curation part of this project was to enhance the bioenergy-related content of the MetaCyc database in several respects. All in all, we have created on the order of 50 new bioenergy pathways.

Lignocellulose compounds. We significantly enhanced this category of chemical compounds in MetaCyc, with a focus on natural cellulosic and hemicellulosic polymers [e.g. a (1→3)-β-D-xylan]. We also curated many of the compounds that are routinely used by researchers in the field for characterization of these enzymes (for example, Avicel, p-nitrophenyl-derivatives).

Lignocellulose-degrading enzymes and pathways. We created many pathways that describe the degradation of lignocellulosic compounds, along with the enzymes that are involved. A few of the polymers included in these pathways are cellulose, rhamnogalacturonan, (1,3)-β-D-xylan, (1,4)-β-D-xylan, L-arabinan, xyloglucan, L-arabinan, and glucuronoarabinoxylan.

New display for Glycan degradation pathways. The traditional pathway layout, which shows successive ordered reactions, is often not suitable for describing glycan degradation pathways, where a number of enzymes act at the same time at different locations of the polymer. To help illustrate this type of degradation pathways we implemented a new display tool by integrating the GylcanBuilder software into Pathway Tools. Using this tool, polymer structures are shown using symbols for the glycan monomers, and the location of sites cleaved by different enzymes is shown using color-coded arrows pointing to the bonds within the polymer structure that are cleaved.

Biofuel production. Under this category we created pathways for the biosynthesis of bioenergy-relevant chemical products, such as alcohols, isoprenoids, alkanes, algal oil compounds, etc.

Hydrogen-production. We created eight different hydrogen production pathways, describing the process in different organisms, encompassing all known types of hydrogenase enzymes.

Engineered pathways. Since many of the important bioenergy-related pathways are engineered pathways that combine enzymes from multiple organisms within a single host, we added to MetaCyc the ability to describe such pathways, and curated a number of relevant engineered pathways.

References

[1] The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nuc. Acids Res. Jan 2014 42:D459-71.

Funding Statement

This work was funded by DOE grant DE-SC0004878.

174. Development of a core *Escherichia coli* inetic metabolic model using the E formalism

Ali Khodayari[1*](auk241@psu.edu), Ali R. Zomorrodi[1], James C. Liao[2], and **Costas D. Aranas**[1]

[1]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA; [2]Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA.

http://maranas.che.psu.edu/
http://www.seas.ucla.edu/~liaoj/

Project Goals: The goal of this project is to explore the use of Ensemble odeling E concepts for constructing inetic models of metabolism consistent with available measurements metabolomics fluxomic and capable of guiding genetic interventions for overproduction. Here we describe the development of an E based inetic model of core *E. coli* metabolism.

Recent availability of fluxomic and metabolomics data has paved the way for large-scale kinetic modeling of metabolism. Challenges still remain on deriving meaningful kinetic descriptions and parameterizations that faithfully replicate metabolic responses to genetic and/or environmental perturbations. In this study we introduce a kinetic model of *E. coli* core metabolism that satisfies the fluxomic data for wild-type and seven mutant strains [1] by making use of the recently introduced Ensemble Modeling (EM) concepts. This model encompasses 138 reactions, 93 metabolites and 60 substrate-level regulatory interactions and accounts for glycolysis/gluconeogenesis, pentose phosphate pathway, TCA cycle, major pyruvate metabolism, anaplerotic reactions and a number of reactions in other parts of the metabolism (Figure 1). Parameterization is performed using a formal optimization approach that minimizes the uncertainty-scaled discrepancies between model predictions and flux measurements. The predicted fluxes by the model are within the uncertainty range of experimental flux data for 78% of the reactions (with measured fluxes) for both the reference (wild-type) and seven mutant strains. The remaining flux predictions fall within three standard deviations of measured values. Converting the EM-based parameters into a Michaelis-Menten equivalent formalism revealed that 80% of K_m and k_{cat} parameters are within one order of magnitude of the literature available values. The predicted metabolite concentrations by the model are also within uncertainty ranges of metabolomic data for 68% of the metabolites. A leave-one-out cross-validation test to evaluate the flux prediction performance of the model showed that metabolic fluxes for the mutants located in the proximity of mutations used for training the model are predicted more accurately. The constructed model and parameterization procedure provides the means for the construction of even larger-scale models as well as models with more narrowly distributed parameter values as new metabolomics/fluxomic data sets are becoming available for *E. coli* and other well studied organisms.

Reference

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.

The work was supported by the grant from Department of Energy, USA (grant# DE-ER65254).

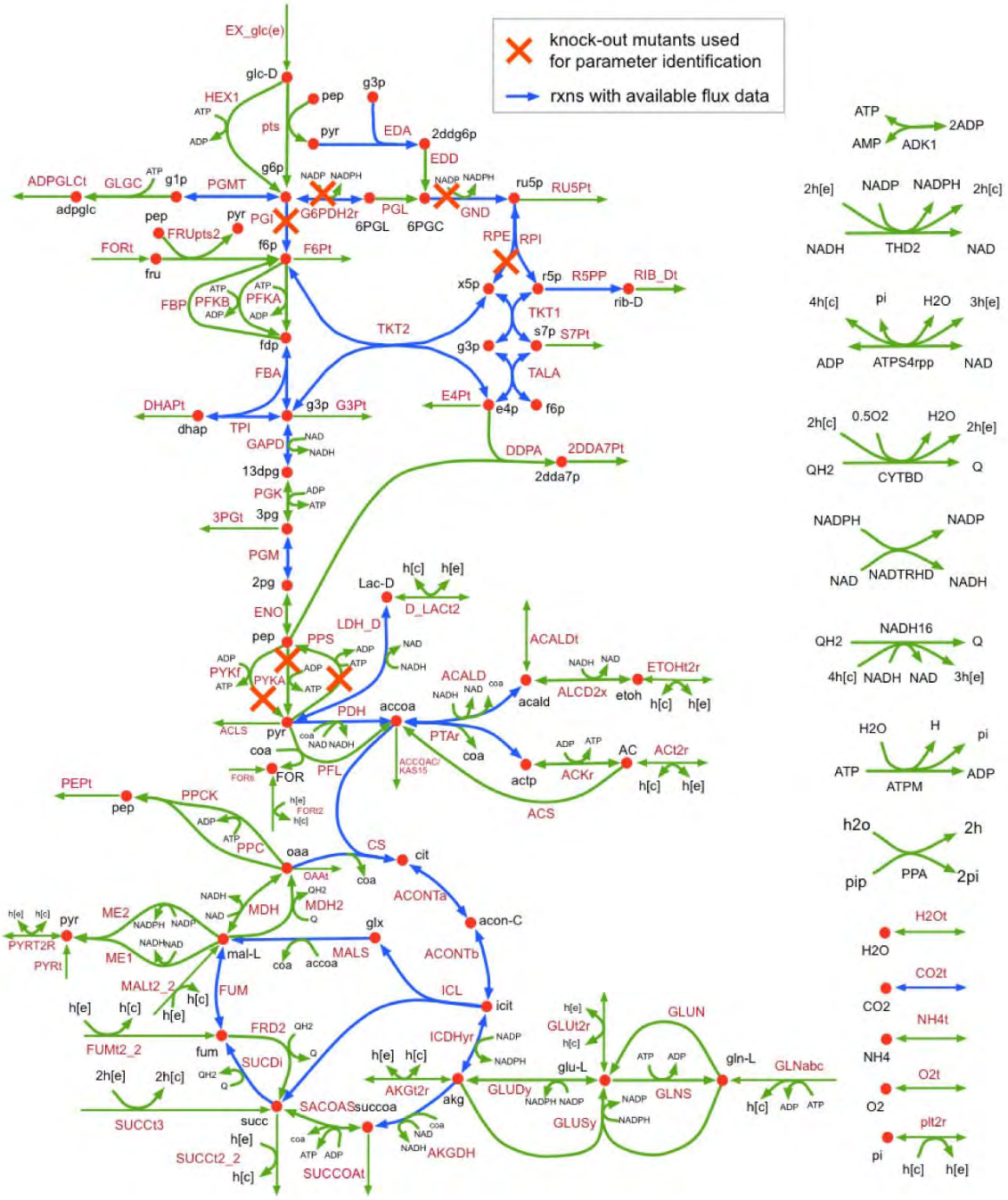


Figure 1. The constructed kinetic model of *E. coli* core metabolism**.** The red crosses represent the knockout mutants with available flux data, used for parameter estimation and cross-validation analysis. Reactions in blue represent those with available flux measurements for the wild-type and mutant strains.

175. Integrating atom mapping information within MetRxn: Application to metabolic flux elucidation through MFA

Akhil Kumar[1]* (azk172@psu.edu), Saratram Gopalakrishnan[2] and **Costas D. Aranas[2]**

[1]The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA;
[2]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA

http://maranas.che.psu.edu/

Project Goals: This project aims to organi e and disseminate standardi ed metabolite and reaction information to improve metabolic modeling by accurately describing reaction stoichiometry, directionality, atom mapping from reactants to products, and gene to protein to reaction relations. Developed standardi ation algorithms automatically curate information to remove incompatibilities in content representation, fix stoichiometric errors such as elemental or charge imbalances and resolve incomplete atomistic details.

MetRxn is a standardized non-redundant searchable collection of published metabolic models and databases from a wide variety of organisms. The standardization procedure follows a workflow that starts by matching metabolite entries using lexicographic and phonetic techniques, and structure comparison using atomistic details. The reactions are first charge and mass balanced and subsequently atom/bond mapping resolution algorithms are applied. For each reaction, metabolite stoichiometry, atom transition and metabolite compartment information is stored. The reaction and metabolite information is downloadable in SBML 3.0 and in a tabular format. The current MetRxn update includes recently published metabolic data for a total of 112 metabolic models and 8 metabolic databases. The number of distinct reactions that have been mapped is greater than 20,000 and MetRxn contains tools that allow users to download atom mapping data for each reaction.

As part of our ongoing effort we have enhanced the MetRxn knowledgebase with additional information such as reaction transition information and reaction standard free energies. In accordance with our data integration goal, we have integrated the ncbi taxonomy database, uniprot gene id's and ncbi gene id's within MetRxn. We developed a customized algorithm for quickly generating unique molecular graphs and detecting symmetries for all metabolites in the database. This is used to create atom transition information between reactants and products for all reactions contained in MetRxn. This information is leveraged for the construction of genome- scale size mapping models to support metabolic flux elucidation using C13 labeled substrates through MFA. Algorithmic details and the impact of migrating to genome-scale models on flux elucidation fidelity are discussed. The current release, MetRxn 2.0 (http://www.metrxn.che.psu.edu/) makes available information on metabolites, reactions, enzymes and reaction atom transitions required by metabolic flux elucidation tools such as Metran, OpenFlux, 13CFlux2 and FiatFlux.

The work was supported by the genomic science grant from Department of Energy, USA (grant# DEFG02-05ER25684).

176. Opt orce: train design using inetic models and synthetic biology tools

Anupam Chowdhury[1*] (axc1028@psu.edu), Chiam Y. Ng[1], Howard M. Salis[1], and **Costas D. Aranas[1]**

[1]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA.

http://maranas.che.psu.edu

Project Goals: The goals of this effort are to i improve the fidelity of computational strain design procedures by incorporating substrate level regulation in the form of inetic expressions in stoichiometric models using the Opt orce procedure, and ii implement the identified intervention strategies by creating a library of R of different Translation nitiation Rates using the R library calculator to construct synthetic operons that maximi es the yield of the target product.

Existing computational strain-design approaches relying solely on stoichiometry and on- off regulation ignore the effects of metabolite concentrations and substrate-level enzyme regulation while identifying metabolic interventions. In this work, we developed the k- OptForce procedure which integrates the available kinetic descriptions of metabolic reactions with stoichiometric models to sharpen the prediction of intervention strategies for improving the bio-production of a chemical of interest. The interventions suggested by k-OptForce are comprised of both direct enzymatic parameter changes (for reactions with available kinetics) and indirect reaction flux changes (for reactions with only stoichiometric information). Application of k-OptForce to the overproduction of L-serine in *E. coli* and triacetic acid lactone (TAL) in *S. cerevisiae* revealed that the identified interventions tend to cause less dramatic rearrangements of the flux distribution so as not to violate concentration bounds. In some cases, additional modifications are needed to overcome the substrate-level regulations imposed by the representative kinetic model. The mechanism of action of these modifications is often subtle by alleviating substrate inhibition or draining away cofactors from competing pathways. In other cases, kinetic expressions shape flux distributions so as to favor the overproduction of the desired product requiring fewer direct interventions. This work paves the way for the integrated analysis of kinetic and stoichiometric models and enables elucidating system-wide metabolic interventions while capturing regulatory and kinetic effects.

The prioritized set of interventions is subsequently used as a guideline to construct the mutant strain using synthetic biology approaches. Flux through a metabolic reaction is rationally controlled by altering the Translation Initiation Rates (TIR) of the Ribosome Binding Sites (RBSs) using predictive RBS library calculator. Each of this degenerate RBS sequence is able to span a large range of TIRs, allowing us to generate libraries of cells with varying levels of that particular protein. After completing all the computer- aided design, we combine commercial DNA synthesis, Gibson DNA assembly, and multiplex automated genome engineering (MAGE) to construct synthetic bacterial operons. These recombineering techniques are used to insert genetic systems to the *E. coli* chromosome and rationally control the enzyme expression levels via directed genome mutagenesis. We are currently applying this strategy to manipulate the level of proteins of *E. coli* native genes to pinpoint the flux changes consistent with those metabolic interventions predicted by k-OptForce.

The work was supported by the genomic science grant from Department of Energy, USA (grant# DEFG02-05ER25684).

177. Metabolic modeling of multi tissue and multi organism systems

Margaret Simons[1]*(mns157@psu.edu), Rajib Saha[1]*(rus184@psu.edu), Mohammad Mazharul Islam[1] (mzi108@psu.edu), Ali R. Zomorrodi[2], Bertrand Hirel[3], and **Costas D. Aranas**

[1]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA; [2]Bioinformatics Program & Biomedical Engineering Department, Boston University, Boston, MA; [3]Institut Jean-Pierre Bourgin, Institut National de la Recherche Agronomique, Centre de Versailles-Grignon, UR 511, Route de St Cyr, F-78026 Versailles Cedex, France

http://maranas.che.psu.edu/

Project Goals: The project aims to improve the understanding of metabolic interactions both within organisms containing multiple tissue types i.e., plants and microbial communities containing multiple species. Overall, the specific goals of this project are two fold: i to reconstruct a multi tissue mai e model for determining bottlenec s in nitrogen metabolism, suggesting genetic manipulations that improve nitrogen use efficiency, and enhancing the understanding of nitrogen flow through the plant, and ii to develop and implement a constraint based and multilevel optimi ation approach for analy ing the physiological responses and interactions within microbial communities. This dynamic modeling framewor will ultimately allow us to study the metabolic trade offs within natural and bioengineered microbial communities by capturing temporal changes and incorporating substrate upta e inetics.

Genome-scale metabolic models in combination with flux balance analysis can be used to explore the metabolic repertoires and restrictions of complex organisms or microbial communities. By reconstructing multi-tissue and multi-organism models, we can determine the interactions between different cell/tissue-types or organisms, resolve bottlenecks in limiting pathways, and study the metabolic trade-offs between species- level and community-level fitness functions.

Towards the first goal, a second-generation genome-scale model of *Zea Mays* has been constructed that captures C_4 carbon fixation by modeling the interactions between the bundle sheath and mesophyll cells in the leaf tissue. By integrating our earlier model, *i*RS1563, with information from the Kyoto Encyclopedia of Genes and Genomes (KEGG), MaizeCyc, and MetaCrop databases, we have constructed a cell-type specific metabolic model. The model combines gene-protein relationships (GPRs) with elemental and charge-balanced reactions. Experimental evidence pertaining to the biomass composition, compartmentalization, and flux constraints was incorporated into the model. Transcriptomic and proteomic data is used to introduce regulatory constraints in the model to improve the simulation of nitrogen rich/poor conditions as well as the impact of a glutamine synthase (GS) deletion. Using flux balance analysis combined with condition-specific biomass equations, we have simulated nitrogen rich and nitrogen poor supply conditions. Model suggestions achieve 80% accuracy in predicting the direction of change in metabolite pool sizes under the excess nitrogen versus limiting nitrogen conditions for 71 metabolites with metabolome data. Similarly, we attain approximately 76% and 75% accuracy in predicting the impact of GS1.3 and GS1.4 deletions on the 71 metabolites pool sizes, respectively. Ultimately, the goal is to reconstruct a multi-tissue model of not just leaf but all five major tissue-types in maize (root, stalk, leaf, tassel and seed), analyze the flow of nitrogen from the plant root to the other tissues, suggest genetic interventions to improve nitrogen use, and study the effect of nitrogen on sugar storage in the seed.

Moving from a single organism, we aim to develop efficient computational tools for the metabolic modeling and analysis of multi-species microbial systems, where more than one microorganism is involved and can interact through the unidirectional or bidirectional exchange of biochemical cues. Toward this end, we previously developed a procedure called OptCom for the steady-state flux balance analysis of microbial communities using genome-scale metabolic models. Microbial communities are known to exhibit dynamic shifts in their metabolism and inter-species interactions in response to changes/perturbations in environmental conditions to support co-growth, survival, and stability. The temporal variations in inter-species metabolic interactions can significantly affect the community structure and functions. In order to capture the temporal dynamics of microbial communities, we have developed a modeling framework called d-OptCom, by extending the OptCom procedure, which allows for incorporating a kinetic description of the uptake of shared metabolites while integrating species- and community-level fitness functions. The applicability of d-OptCom was demonstrated by modeling the dynamic co-growth of a number of auxotrophic mutant pairs of *E. coli* and by computationally assessing the dynamics within a uranium-reducing community comprised of *Geobacter sulfurreducens*, *Rhodoferax ferireducens* and *Shewanella oneidensis*. d-OptCom was also employed to examine the impact of electron donor addition on the relative abundance of uranium reducing species. These studies elucidate the importance of simultaneously accounting for both species- and community-level fitness functions when modeling microbial communities and demonstrate that the incorporation of uptake kinetics substantially restricts the feasible space of inter-species flux trafficking. Overall, this study paves promising frontiers for the dynamic multi- objective analysis of complex microbial ecosystems.

The work was supported by the genomic science grant from Department of Energy, USA (grant# DEFG02-05ER25684).

178. ¦sing MetRxn for metabolic model reconstruction, flux elucidation and redesign

Costas D. Aranas

[1]Department of Chemical Engineering, Pennsylvania State University, University Park, PA

http://maranas.che.psu.edu/

Project Goals: This project aims to organi e and disseminate standardi ed metabolite and reaction information to improve metabolic modeling by accurately describing reaction stoichiometry, directionality, atom mapping from reactants to products, and gene to protein to reaction relations. This resource is deployed for microbial, multi tissue and multi organism metabolic model reconstruction, metabolic flux elucidation using labeled substrates A and computational strain design.

MetRxn is a standardized non-redundant searchable collection of published metabolic models and databases from a wide variety of organisms. The current MetRxn 2.0 update includes recently published metabolic data for a total of 112 metabolic models and 8 metabolic databases. The number of distinct reactions that have been mapped is greater than 20,000 and MetRxn contains tools that allow users to download atom mapping data for each reaction. In this talk, we will elaborate on new features of MetRxn 2.0 (http://www.metrxn.che.psu.edu/) including atom mapping information across all reactions and enhanced integration with other databases. We will describe how this resource can impact genome-scale metabolic model reconstruction by providing curated reaction and metabolite content. Progress towards the development of a multi- tissue metabolic model for maize, rapid generation of cyanobacterial models and metabolic modeling of microbial communities will be briefly highlighted. Reaction atom transition information in MetRxn can rapidly be leveraged to create genome-scale atom mapping models. Efforts towards resolving metabolic fluxes at a genome-scale and computational challenges with current flux elucidation tools will be described and the impact on flux elucidation fidelity will be quantified.

Existing computational strain design approaches relying solely on stoichiometry and rudimentary constraint-based regulation overlook the effect of metabolite concentrations and substrate-level enzyme regulation while identifying metabolic interventions. This may lead to suggested interventions that cannot be implemented. To remedy this, we developed the k-OptForce procedure which integrates all available kinetic descriptions of metabolic reactions with stoichiometric models to sharpen the prediction of intervention strategies for improving the bio- production of a chemical of interest. In addition, we have used the Ensemble Modeling (EM) procedure for constructing kinetic models of core *E. coli* metabolism consistent with available measurements (metabolomic & fluxomic).

The work was supported by the genomic science grant from Department of Energy, USA (grant# DEFG02-05ER25684).

179. A Regulated Model for *Clostridium acetobutylicum* Based on Response to Butanol and Butyrate Stress

Satyakam Dash[1]* (sid5219@engr.psu.edu), Thomas J. Mueller[1]* (tjm5293@psu.edu), Keerthi P. Venkataramanan[2], **Eleftherios T. Papoutsais**[2], and **Costas D. Aranas**[1]

[1]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA; [2]Department of Chemical Engineering and Delaware Biotechnology Institute, University of Delaware, Newark, DE.

http://maranas.che.psu.edu

Project Goals: This project aims to understand and model the stress response of *Clostridium acetobutylicum* ATCC 2 to two important toxic metabolites: butanol and butyrate, using a regulated genome scale model.

Clostridia are anaerobic Gram-positive Firmicutes containing broad and flexible systems for substrate utilization, which have been used successfully to produce a range of industrial compounds. In particular, *Clostridium acetobutylicum* has been used to produce butanol on an industrial-scale through acetone–butanol–ethanol (ABE) fermentation. A genome-scale model is a powerful tool for understanding the metabolic capacity of an organism. The inclusion of regulatory information to a genome-scale model provides additional reliability for phenotype predictions. This work describes the construction of a genome-scale metabolic model of *C. acetobutylicum* ATCC 824, *i*CAC802, and its integration with experimental gene transcription data. *i*CAC802 spans 802 genes and includes 1137 metabolites and 1470 reactions along with gene-protein-reaction associations.

Both ^{13}C-MFA and gene deletion data in the ABE fermentation pathway were used to test the predicted allowable flux ranges by the model. Transcription data measured in response to two stressors, butanol and butyrate, were used to impose 1071 regulatory constraints in the form of flux bound constraints for the *i*CAC802 model using the E-Flux method. These bounds affected the flux of tens of reactions in core metabolism. This regulated metabolic model was tested through comparisons with experimental fermentation data under the same stressed conditions. The regulated model showed down-regulation of glucose uptake as observed experimentally under stress conditions. The model including the regulatory restrictions under stress exhibited an approximately 50% reduction in biomass yield which is in broad agreement with experimental data. The experimental fermentation data for acetate and butyrate also lie within the flux ranges predicted by the model.

Our ultimate goal in developing a regulated model is to achieve precise, condition-specific metabolic predictions to aid redesign in pursue of a desired overproduction phenotype.

The work was supported by the genomic science grant from Department of Energy, USA (grant # DE-SC0007092).

180. Parametrizing a Genome-Scale Model of Metabolism and Expression in E. coli with Multi-omic Data

Ali Ebrahim[1]*, Edward O'Brien[1], Joshua Lerman[1], Donghyuk Kim[1], Karsten Zengler[1], Adam Feist[1], **Bernhard Palsson[1]**

The quantitative relationship between the rates of transcription, translation, and catalytic activity for any given enzyme are important constraints on models coupling metabolism with gene expression (ME). We have assembled a dataset of proteomic and transcriptomic data for *E. coli* growing in batch on four different carbon sources: glucose, fumarate, pyruvate, and acetate. We have found that the protein/mRNA ratio remains constant for most genes across the different environments. Using a recently reconstructed ME model for *E. coli* , we have developed an algorithm which uses expression from experimental data and metabolic flux computed in the model to iteratively estimate parameters constraining enzymatic activities, sampling from different starting values. The converging model parameters proved to be remarkably consistent across all four conditions, and correlate with reported enzyme efficiencies. For some reactions for which fluxomic data is available, parameters calculated using measured instead of modeled fluxes were integrated into the parameter set. Using this parameter set increases the accuracy of the ME model when predicting differential expression between different conditions.

181. Dynamic Model Building Based on the Ensemble Modeling Approach

Jimmy G. Lafontaine Rivera[a*](lafonj@gmail.com), Yun Lee[a], Ali Khodayari[b], Thomas Wasylenko[c], **Costas D Maranas**[b], **Greg Stephanopoulos**[c], **James C. Liao**[a]

[a] Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, 5531 Boelter Hall, Los Angeles, CA 90095-1592, USA
[b] Department of Chemical Engineering, The Pennsylvania State University, University Park, PA, USA
[c] Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

Project Goals: The goal of this project is to develop a novel modeling approach to describe the dynamic behavior of metabolic networks (in particular, flux changes upon enzyme tuning) by integrating multiple data platforms, including flux, metabolites, transcriptome, and enzyme tuning data. Although the utility of such models is undeniable, their development has been impaired by inadequate modeling approaches, the sheer size of the problem, and difficulties in accessing the intracellular environment. As a result, little progress has been made in realizing such dynamic models despite the continuously increasing number of intracellular measurements that are becoming available by high throughput methods. The resulting models from the proposed research will account for pathway enzyme kinetics and aim to predict the effects of genetic manipulations designed to bring about changes in metabolic flux and overproduction of metabolites, such as tuning various enzyme levels or the Michaelis-Menten constants (K_m) of key enzymes. In this context, such models will be instrumental for constructing microbial strains to produce various biofuels such as ethanol, 1-butanol, and isobutanol from renewable resources. We will use production of these biofuels in *Escherichia coli* as a model system, both because of its central role as a test bed in systems biology, the wealth of kinetic and regulatory information available and its successful usage for the production of biofuels. While the *E. coli* focus will facilitate model development, the approach developed will be general and applicable to other microorganisms and eventually plants. The project is based on the Ensemble Modeling (EM) approach, robust flux and metabolite measurements, and an efficient optimization scheme developed in the PIs laboratories.

Presently, there are no satisfactory dynamic models of cellular function. This unique deficiency persists despite the extraordinary advances that have taken place during the past decade in the areas of high throughput measurement of cell-wide intracellular biomolecules and molecular level simulations of various systems. Present models of microbial metabolism suffer from serious drawbacks that limit their applicability as a robust and versatile tool for re-engineering metabolic networks. Such limitations include: (a) reliance on coarsely lumped kinetic and regulatory information, (b) sparse and/or unreliable kinetic parameters derived under mostly in vitro conditions that poorly approximate the cellular environment, (c) small scale models describing only fractions of the cellular metabolism, (d) difficulty in scaling up to levels required by our current understanding of cellular function and, also, potentially allowed by available genomic and cell-wide measurements. This difficulty stems from the inherent problems in accurately estimating the large number of parameters required for kinetic cell-wide models. Consequently, current cell-wide models are mainly stoichiometric in nature and capitalize on genomic sequence information to define cell-wide bioreaction networks whose rates are determined such as to optimize a cell objective (such as maximizing growth rate) subject to constraints derived from metabolite balances (the FBA approach).

Ensemble Modeling (EM) was recently introduced to address the many shortcomings of conventional theory-based models. The basic approach of EM is to construct an ensemble of dynamic models that span the entire kinetic parameter space allowable by thermodynamics. All such models attain the same steady state in terms of flux distribution and metabolite concentrations. However, these steady states are altered upon introduction of specific genetic perturbations. Ultimately, we expect to develop models that will encompass all reactions involved in current metabolic reconstructions of the bacterium *Escherichia coli*, a sum total of approximately 150-200 reactions that carry significant flux. (Other reconstructed networks carry zero or negligible fluxes under most conditions.) Besides being able to predict the outcomes of manipulation of any enzyme within the model, it will have the flexibility to be adapted to incorporate the production of novel compounds. This flexibility will allow experimentalist to quickly pinpoint targets for genetic manipulation regardless of the novel pathway they have introduced.

In this work, we use parameter continuation methods for analyzing the robustness of metabolic models. Parameter continuation allows us to quickly observe the metabolic effects of perturbations, such as enzyme activity changes, as function of perturbation magnitude without the need to run computationally expensive time domain integrations. This gives us a more complete view of a perturbation's effect on metabolism and allows us to examine a model's robustness. Parameter continuation will play a big role in the development of cell-wide models, because robustness is a key element of living organisms and should be a property of its models and numerical integration of the time-domain system becomes computationally infeasible as models increase in scope.

182. PhyloFacts FAT-CAT Ortholog Identification and Functional Annotation

Kimmen Sjölander[1]* (kimmen@berkeley.edu) and Cyrus Afrasiabi[1] [1]University of California, Berkeley
http://phylogenomics.berkeley.edu/phylofacts/

Project goals: The PhyloFacts project aims to develop novel methods and web servers to integrate multiple data types and to infer and curate (meta)genomic functional annotations. We address these aims by using evolutionary reconstruction to integrate and organize heterogeneous data from homologous genes across thousands of species, to simultaneously derive functional and taxonomic annotations for environmental sample sequences, and for functional annotation of whole genomes.

The PhyloFacts FAT-CAT web server is designed to improve the accuracy of genome functional annotation. Standard pipelines use a BLAST-based annotation-transfer protocol to assign predicted functions to genes. This approach is now known to have high error rates due to hidden paralogy and promiscuous domains, and errors can be propagated through sequence databases by annotation transfer protocols. Although the actual fraction of gene annotation errors is unknown (and errors can be difficult to detect), numerous studies suggest that as many as 25% of genes have errors in their functional annotations. If we add to this number the ~30% of genes labeled simply as "hypothetical", it becomes clear that significant work remains to improve on this status quo. We address these challenges in the PhyloFacts resource through the use of phylogenomic analysis, protein structure information and integration of heterogeneous experimental and annotation data.

The fundamental assumption underlying phylogenomic analysis is that accurate prediction of protein function depends on the phylogenetic identification of orthologs. Orthology is a phylogenetic term: two proteins are each other's orthologs if they are related by speciation from a common ancestor. We combine phylogenomic ortholog identification with protein structure analysis, for instance, building trees for Pfam domains as well as for multi-domain architectures, an approach we call *structural phylogenomics.* Protein function is multi-faceted, and informed by experimental data of different types; we therefore retrieve data from numerous resources, including GO annotations for molecular function, biological process, and cellular location, pathway role, protein-protein interaction, Pfam multi-domain architecture, Enzyme Commission numbers, and other types of experimental and annotation data. We overlay these data on protein family trees, allowing the function(s) of individual proteins to be inferred based on annotations retrieved for their homologs in the tree, weighting sequences nearby in the tree more than those that are distant, and weighting sequence annotations with experimental support more than those that have been derived using noisy annotation transfer protocols.

Because building phylogenetic trees is computationally expensive, we have developed a system to enable phylogenomic classification for novel sequences (e.g., newly sequenced genomes) using hidden Markov models placed at all nodes (or vertices) in PhyloFacts protein family trees. Sequences can be assigned to positions in trees based on the top-scoring HMM in the tree. From this phylogenetic placement, we identify orthologs and derive a functional annotation.

PhyloFacts' broad taxonomic and functional coverage, with >7.3 M proteins from across the Tree of Life, enables FAT-CAT to predict orthologs and assign function for most sequence inputs. Benchmarking experiments comparing FAT-CAT against the major orthology web servers – eggNOG, KEGG, OrthoMCL, InParanoid, PhylomeDB and OrthoDB – demonstrate FAT-CAT's high precision and robustness to both promiscuous domains and recent duplication events. On this dataset, FAT-CAT was the only webserver with perfect precision, with OMA and PhylomeDB making a very small number of errors.

By contrast, other orthology web servers mix paralogs with predicted orthologs and include proteins with only partial homology to query sequences.

The FAT-CAT web server is available at http://phylogenomics.berkeley.edu/phylofacts/fatcat/. Details on the method and validation experiments are presented in Afrasiabi *et al*, "The PhyloFacts FAT-CAT Webserver: Ortholog Identification and Function Prediction using Fast Approximate Tree Classification," *Nucleic Acids Research* 2013. Supplementary data are available online at http://phylogenomics.berkeley.edu/phylofacts/fatcat/supplementary/.

The PhyloFacts FAT-CAT webserver is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

183. Determining the control circuitry of redox metabolism at the genome-scale

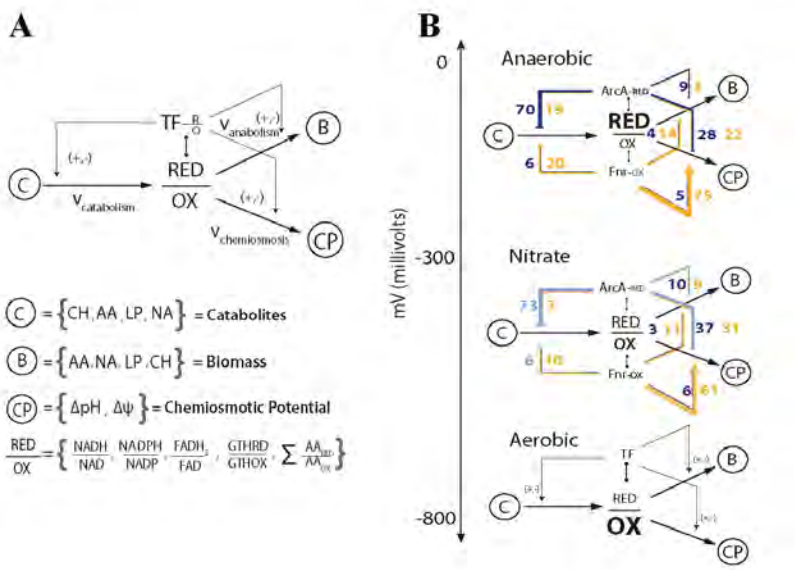
Authors and Affiliations:
Stephen A. Federowicz[1*], Donghyuk Kim[1], Ali Ebrahim[1], Joshua Lerman[3], Harish Nagarajan[4], **Byung-Kwan Cho[2]**, **Karsten Zengler[1]**, **Bernhard O. Palsson[1]**

[1]Department of Bioengineering, University of California – San Diego, La Jolla, CA, U.S.A., [2]Department of Biological Sciences, KAIST, Daejoen, Korea., [3]Amryis Inc. El Cerrito, CA., [4] Genomatica Inc. San Diego, CA 92122. *Presenter

Presenter contact: sfederow@ucsd.edu
URL: http://systemsbiology.ucsd.edu

Project Goals:
This project aims to: (1) create a fully curated, bottom up reconstruction of the transcriptional regulatory network in *Escherichia Coli*, (2) determine fundamental constraints on the regulatory response via network and sequence level features, (3) develop a non-Boolean constraints based modeling approach for regulation, (4) integrate the transcriptional regulatory network with metabolic and macromolecular synthesis models, and (5) provide a platform for genome scale metabolic engineering and synthetic design.

Project Description:
Determining how facultative anaerobic organisms sense and direct cellular responses to oxygen availability has been the subject of intense study. However, even in the model organism Escherichia coli, existing mechanisms only explain a small fraction of the hundreds of genes that are regulated. Here we propose a model that accounts for the full breadth of regulated genes by detailing how two global transcription factors (TFs), ArcA and Fnr of Escherichia coli, sense redox ratios and act on a genome-wide basis to coordinately balance anabolic, catabolic, and energy generation pathways. We first addressed gaps in the regulatory network by carrying out ChIP-chip and gene expression experiments to identify 463 regulatory events. We then interfaced this reconstructed regulatory network with a highly curated genome-scale metabolic model and show that ArcA and Fnr control > 80% of total metabolic flux. Finally, we provide evidence for a proposed feedforward with feedback trim model by calculating a 0.71($p < 1e-6$) correlation between changes in metabolic flux and changes in regulatory activity across fermentative and nitrate respiratory conditions. We also are able to relate the proposed model to a wealth of previously generated data.


A
TF
RED
OX
B
CP
C
$V_{catabolism}$
$V_{anabolism}$
$V_{chemiosmosis}$
C = {CH, AA, LP, NA} = Catabolites
B = {AA, NA, LP, CH} = Biomass
CP = {ΔpH, Δψ} = Chemiosmotic Potential
RED/OX = {NADH/NAD, NADPH/NADP, FADH2/FAD, GTHRD/GTHOX, Σ AA_RED/AA_OX}
B
0
-300
-800
mV (millivolts)
Anaerobic
ArcA-RED
RED
OX
Fnr-OX
Nitrate
Aerobic
TF
RED
OX

184. Reconstruction of Protein Translocation in *Escherichia coli* Allows For Bottom-Up Predictions of Compartmentalization Properties

Joanne Liu[1]*, Edward O'Brien[1], Joshua Lerman[1], Adam Feist[1], **Bernhard Palsson**[1], and **Karsten Zengler**[1]

[1]Department of Bioengineering, University of California San Diego, 9500 Gilman Drive, Mail Code 0412, La Jolla, CA, 92093, USA.

http://sbrg.ucsd.edu

Project goal: Reconstruct protein translocation in a genome-scale metabolic and gene- expression model of *Escherichia coli* to serve as an accessible tool that represents a consolidation of research and permits studying the effects compartmentalization has on bacterial metabolism

Compartmentalization is essential for life, and bacteria have evolved pathways to translocate proteins from one compartment to another. Although the individual pathways have been studied in detail, the cumulative effects the pathways have on the whole cell, and vice versa, has yet to be studied. A recent and novel genome-scale model of metabolism and gene-expression of *Escherichia coli* allows us to now study the effects of compartmentalization on a systems-level. To enable such analysis, the model was significantly expanded to include a comprehensive reconstruction of protein translocation pathways. Other improvements to the model include the incorporation of five distinct protein compartments (cytoplasm, periplasm, the inner and the outer membrane, and the extracellular space), published enzymatic rates of the translocases, and a membrane constraint based on cell morphology. The metabolic and phenotypic demands of the improved model allow for the de novo prediction of enzyme abundances. Comparison against experimental data reveals that we are capable of estimating accurate protein expression levels (Pearson correlation of 0.998 for translocase pathways). Furthermore, the model can be used to examine optimal energy generation and growth states and how they differ from an observed measured state in vivo. For example, one observation was that the model predicted more resources can be diverted to oxidative phosphorylation instead of unnecessary ABC transporters not needed in a given condition. This model reveals that the consolidation of current scientific knowledge enables calculation of protein abundances based on enzymatic activities and compartmental demands.

185. The Core Regulons Orchestrating the Response of *Clostridium acetobutylicum* to Butanol and Butyrate Stress

Qinghua Wang[1,2,3,#,*] (wangq@dbi.udel.edu), Keerthi P Venkataramanan[1,4,#], Hongzhan Huang[1,2,3], **Eleftherios T Papoutsakis**[1,4], and **Cathy H Wu**[1,2,3]

[1]Delaware Biotechnology Institute, University of Delaware, Newark; [2]Center for Bioinformatics and Computational Biology, University of Delaware, Newark; [3]Department of Computer and Information Sciences, University of Delaware, Newark; and [4]Department of Chemical and Biomolecular Engineering, University of Delaware, Newark

http://clostress.org/

Project Goal: The objective of this project is to develop integrated, predictive models of the metabolic and regulatory networks of the metabolite stress response in solventogenic clostridia using enabling systems-biology approaches. Clostridia are Gram$^+$, obligate anaerobic, endospore forming bacteria of major importance to fermentative biofuel production. We focus on understanding and modeling the stress response of *Clostridium acetobutylicum* to two important toxic metabolites: butanol and butyrate. This is a problem of major importance not only in clostridial biotechnologies, but in all microbial systems of interest to DOE for bioenergy production. Coupling a large set of transcriptome data we generated with genome-scale regulon analyses, we have successfully constructed a stress response network model. The bioinformatics and systems biology efforts are applicable to a broad set of microorganisms of interest to the production of biofuels and chemicals from renewable sources.

Organisms of the genus *Clostridium* are Gram$^+$ endospore formers of great importance in human normo- and pathophysiology, and biofuel and biorefinery applications. Exposure of *Clostridium* organisms to chemicals, in particular toxic metabolites, is ubiquitous in both natural (such as in human gut) and engineered environments, engaging both the general stress response as well as specialized programs. Yet, despite its fundamental and applied significance, it remains largely unexplored at the systems level.

In this study, we generated a total of 96 individual sets of high-resolution microarray data examining the transcriptional changes in *C. acetobutylicum*, a model *Clostridium* organism, in response to three levels of chemical stress from the native metabolites, butanol and butyrate. We identified 164 significantly differentially expressed transcription regulators and detailed the cellular programs associated with general and stressor-specific responses, many previously unexplored. Pattern-based, comparative genomic analyses enabled, for the first time, to construct the stress-responsive regulons in *C. acetobutylicum* under butanol and butyrate stress. Notably, the regulons and binding motifs of the stress-related transcription factors (HrcA, CtsR, LexA, Rex and PerR) were defined together with those controlling stress-responsive amino acid and purine metabolism (ArgR, HisR, CymR and PurR).

Using a large set of high-throughput temporal gene expression data in combination with genome-scale regulon analyses, we have successfully built a stress response network model integrating important players for the general and specialized metabolite stress response in *C. acetobutylicum*[1]. Since the majority of the transcription factors and their target genes are highly conserved in other organisms of the *Clostridium* genus, this network would be largely applicable to other *Clostridium* organisms. The network informs the molecular basis of *Clostridium* responses to toxic metabolites in natural ecosystems

and the microbiome, and will facilitate the construction of genome-scale models with added regulatory-network dimensions to guide the development of tolerant strains.

References

1. Wang, Q., Venkataramanan, K. P., Huang, H., Papoutsakis, E. T., and Wu, C. H. (2013) Transcription factors and genetic circuits orchestrating the complex, multilayered response of *Clostridium acetobutylicum* to butanol and butyrate stress, BMC Syst Biol 7, 120.

The work was supported by the genomic science grant from Department of Energy, USA (grant # DE-SC0007092).

[#] Authors contributed equally

186. Heterogeneity in Protein Expression Induces Metabolic Variability in a Modeled *Escherichia coli* Population

Piyush Labhsetwar[1], John A. Cole[2], Elijah Roberts[3], Nathan D. Price[4] and **Zaida A. Luthey-Schulten**[2,3]* (zan@illinois.edu)

[1]School of Molecular and Cell Biology, [2]Department of Physics, [3]School of Chemical Sciences, Univserity of Illinois at Urbana-Champaign, [4]Institute for Systems Biology, Seattle, WA

Project Goals: Methodology used to develop whole cell models of methanogens will be applied to the model organism *E. coli* for which considerable single cell proteomics and kinetic data have just become available. The integration of the effects of stochastic gene expression with genome-scale metabolic/regulatory models is a novel contribution to systems biology. Besides helping to elucidate the underlying mechanisms by which cells respond to intrinsic noise with differential usage of metabolic pathways, the method could be put to practical use in synthetic biology and bioengineering applications. The goal of this project is to validate the computational methodology for analyzing the effect cellular noise have on population heterogeneity within monoclonal populations and demonstrate its applicability to *E. coli*.

Stochastic gene expression can lead to phenotypic differences among cells even in isogenic populations growing under macroscopically identical conditions [1]. Here flux-balance analysis is applied in investigating the effects of single cell proteomics data on the metabolic behavior of an *in silico E. coli* population. The latest metabolic reconstruction is integrated with transcriptional regulatory data in order to model realistic cells growing in a glucose minimal medium under aerobic conditions. Our modelled cells exhibit a broad distribution of growth rates (Figure 1). Well-defined subpopulations that differ in terms of pathway usage can be identified using principal component analysis. The population differentiates into slow-growing acetate- secreting cells and fast-growing CO_2- secreting cells, with a large population growing at intermediate rates shift from glycolysis to Entner-Doudoroff (ED) pathway usage (Figure 2). Constraints on pathway usage imposed by integrating transcriptional data have a large impact on NADH oxidizing pathway usage within the cell. Finally, stochasticity in the expression of only a few genes may be necessary to capture most of the metabolic variability seen in the entire population (Figure 3).

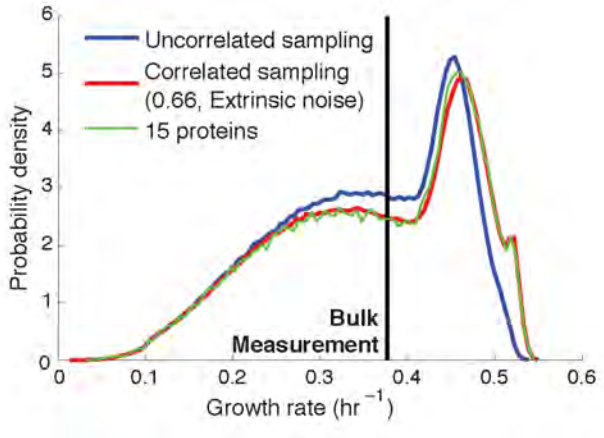


Figure 1 Distributions of specific growth rates predicted by uncorrelated protein sampling (blue), by imposing correlations of correlation coefficient 0.66 among proteins in the extrinsic noise regime (red), and by sampling only the 15 proteins whose copy numbers are most likely to constrain the growth of modeled cells (green). The black line represents the experimentally determined bulk specific growth rate.

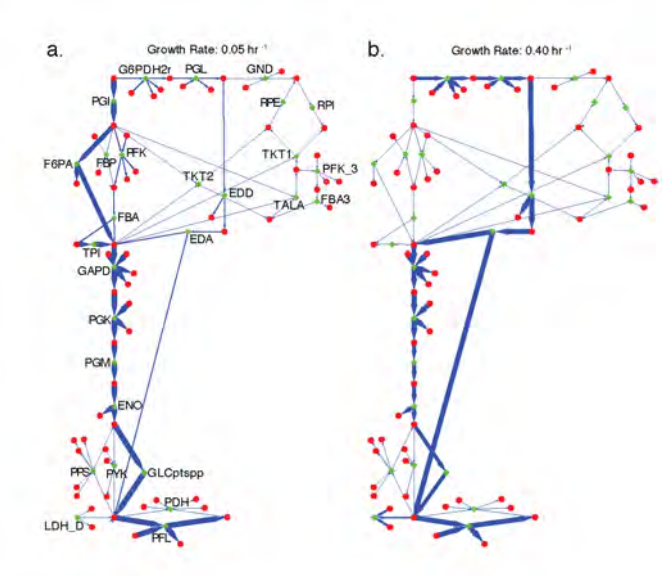


Figure 2 Examples of differences in usage between glycolysis and the ED pathway by representative cells in our metabolize the same amount of glucose as glycolysis, but at the cost of substrate-level ATP generation. Slowgrowing cells tend to use glycolysis (a), whereas intermediate to fast-growing cells tend to use the ED pathway (b).

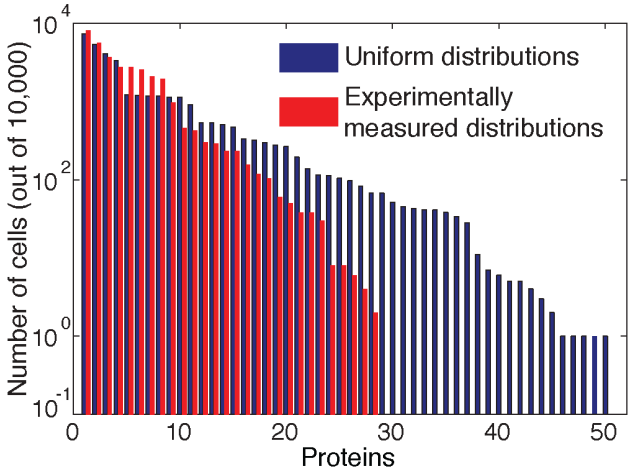


Figure 3 Bar graph indicating the number of cells whose growth is directly limited by a given protein. Only 28 proteins sampled from the experimentally measured protein distributions (shown in red) limit the growth rate of at least one cell in a population of 10,000. For reference, over 50 proteins would be expected to limit the growth rate of at least one cell, had all enzyme counts been sampled from a uniform distribution from 1 to 1,000 (shown in blue).

References

1) P. Labhsetwar, J.A. Cole, E. Roberts, N.D. Price, Z. Luthey-Schulten. *Proc. Nat. Acad. Sci.*, 110(34):14006-11, 2013.

This work was supported by the Department of Energy grant "Computational Modeling of Fluctuations in Energy and Metabolic Pathways of Methanogenic Archaea" number DOE DE- FG02-10ER6510.

187. A Computational Model of Methane Producing Archaea

Joseph R. Peterson[1]* (jrptrsn3@illinois.edu), Piyush Labhsetwar[2], Jeremy R. Ellermeier[2], Petra R.A. Kohler[2], Ankkur Jain[3], Taekjip Ha[3], William W. Metcalf[2] and **Zaida Luthey-Schulten**[1,3]

[1]Shool of Chemical Sciences, [2]School of Molecular and Cell Biology, [3]Department of Physics University of Illinois at Urbana-Champaign

http://www.scs.illinois.edu/schulten/research/index.html

Project Goals: Methanogens utilize a wide variety of substrates to produce methane, which has implications for bioenergy production and global carbon cycling. This project's main goal is the development of a whole-cell model for making predictions about the Methanogenesis pathways in members of the *Methanosarcina* genus. This entails the integration of a stochastic kinetic model of the methanogenic pathways with a transcriptionally regulated metabolic model with which the environmental conditions that induce population heterogeneity may be identified. We will model the response of *M. acetivorans* (marine) and *M. barkeri* (freshwater) to environmental fluctuations. The modeling leverages the expertise of the C. Woese and W. Metcalf labs on the genetics and evolution of the methanogens, as well as single molecule and biochemical experiments. The integration of kinetics with genome-scale metabolic/regulatory models constitutes a novel methodology for systems biology studies of cellular phenotypes.

Progress towards a more complete model of the methanogenic archaeum *Methanosarcina acetivorans* is reported [1]. We characterized size distribution of the cells using differential interference contrast (DIC) microscopy, finding them to be ellipsoidal with mean length and width of 2.9 µm and 2.3 µm respectively when grown on methanol, and on average 2.3 µm long and 1.7 µm nwide when grown on acetate (Figure 1). We used the single molecule pull down (SiMPull) technique to measure average copy number of the Mcr complex and ribosomes (Table 1). In creating a model, RNA expression data (RNA-seq) measured for cell cultures grown on methanol and trimethylamine (TMA) can be used to estimate relative protein production per mole of ATP consumed. A kinetic model for the methanogenesis pathways based on biochemical studies that have been further validated by recent metabolic reconstructions for several related methanogens, is presented (Figure 2a). The kinetic model is capable of capturing experimentally observed mean methane production and mean substrate consumption rates for cell cultures growing on methanol and TMA/methanol mixtures (Figure 2b,c) In this model, twenty-six reactions in the methanogenesis pathways are coupled to a cell mass production reaction that updates enzyme concentrations. The archaeum's growth was most sensitive to the number of methyl-coenzyme- M reductase (Mcr) and methyl-

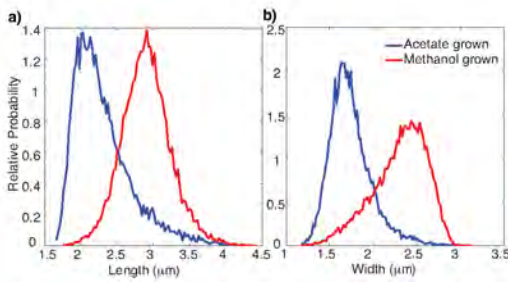


Figure 1 Distributions of (a) length and (b) width of *M. acetivorans* cells grown in methanol and acetate. Methanol grown cells are on average 9 fL and about 2.2x larger than those grown in acetate.

Biological Replicate	Protein Copy/Cell
Mcr	
1	320±713
2	273±124
Rpl18	
1	10038±3340
2	18135±6040

Table 1 Measured copy number per cell (grown in methanol) of two highly expressed proteins

tetrahydromethanopterin: coenzyme-M methyltransferase (Mtr) proteins. A draft model of transcriptional regulation based on known interactions is proposed which we intend to integrate with the kinetic model to allow dynamic regulation (Figure 3).

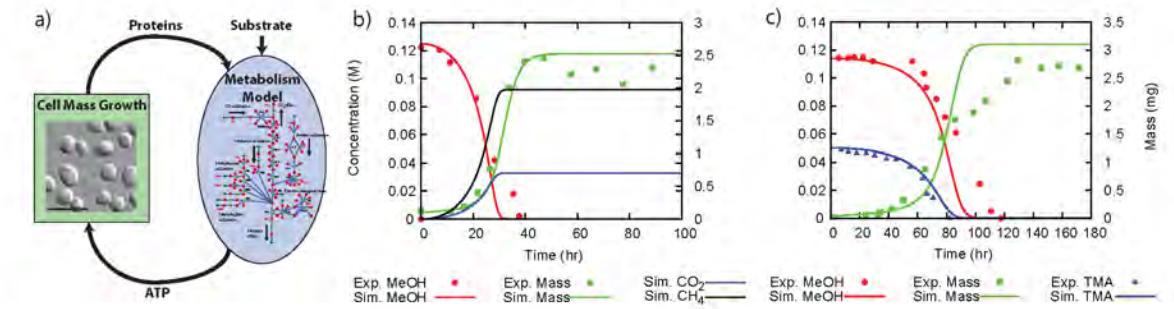


Figure 2 a) Kinetic model for the methanogenesis pathway is coupled to a cell culture mass growth reaction to simulate colony growth. Relative protein production ratios are based on RNA expression data. b) Simulated colony growth on 125 mM methanol for 100 hours showing good agreement to experiments [2]. c) Simulated colony growth on 125 mM methanol and 50 mM TMA for 180 hours also showing good agreement to experiments [2].

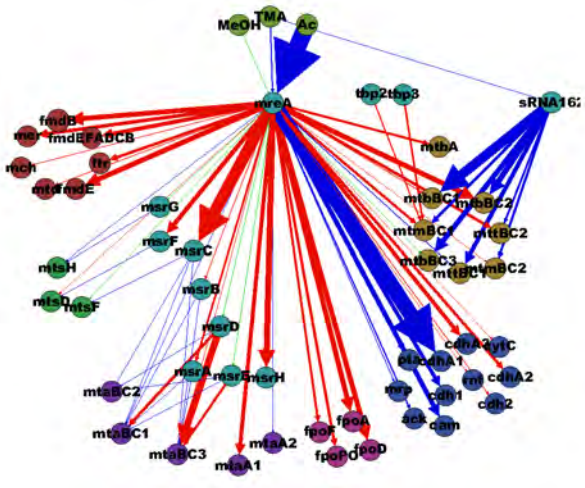


Figure 3A regulation model based on transcription factors (cyan) that promote (blue arrows) or repress (red arrows) expression of different methanogenesis proteins. mreA, turns off many of the methylotrophic proteins and turns on manyMethylamine e into the cells as well as proteins that fix nitrogen, in the presence of MMA, DMA or TMA

References

1. J.R. Peterson, P. Labhsetwar, J.R. Ellermeier, P.R.A. Kohler, A. Jain, T. Ha, W.W. Metcalf, Luthey-Schulten. *Archaea*, 2014, *In Press*.
2. Bose, M. Pritchett, M. Rother, and W. Metcalf. *Journal of Bacteriology*, vol. 188, no. 20, pp. 7274–7283, 2006.

This work was supported by the Department of Energy grant "Computational Modeling of Fluctuations in Energy and Metabolic Pathways of Methanogenic Archaea" number DOE DE- FG02-10ER6510.

188. A notion before projecting biological networks across species: have the genes co- occurred through evolution?

Gang Fang[1,2*] (gang.fang@yale.edu), Doreen Ware,[4,5] Sergei Maslov,[6] and **Mark Gerstein**[1,2,3]

[1]Department of Molecular Biophysics and Biochemistry, [2]Program in Computational Biology and Bioinformatics, [3]Department of Computer Science, Yale University, New Haven, CT; [4]Cold Spring Harbor Laboratory, NY; [5]United States Department of Agriculture-Agricultural Research Service; [6]Brookhaven National Laboratory, NY

Project Goals: This university-KBase project is to develop a set of confidence values for cross-species biological network and modules projection.

KBase and other resources have stored many pre-computed biological networks and gene clusters as functional modules. This changes the way of biological database search from one of querying individual gene's function to examining relationships in a group of genes in networks. Biological network query is particularly useful for studying genes from less investigated species, and often the first step of such studies is to look for clues in database by checking how their orthologs are connected in various functional modules in other species. The reliability of cross-species projection of network information, namely the connections (edges) between genes (nodes) and node attributes, is dependent on many factors. Among which, one factor is fundamental: co-occurrence of genes on phylogenetic tree. If the genes were randomly distributed in phylogeny, projection of their biological connections from one species to another would be less reliable. Therefore, we suggest a set of confidence values for projected modules. First, we propose an evolutionary profile $\{x_0, x_1...x_n\}$ describing the conservation of module among n species, where x_i is percentage of member genes present in species i. With this profile, correlation between this module and phylogenetic tree of the n species is calculated. This is the K statistic [1]. Significant K suggests this module is not randomly distributed on phylogenetic tree. In addition, MPD and MNTD Z-scores are calculated [2]. The two values describe how the module preserves through evolution. Significant MPD Z-score means the module supports phylogenetic tree level clustering, i.e. significant association between phylogeny co-occurrence and pre-computed biological relationships of the member genes, so that projection of this module is reliable. Significant MNTD Z-score also supports projected biological relationships between genes. It is different from MPD Z-score in that it suggests the group of genes subject to co-horizontal transfer or very species-specific. Several study cases of these confidence scores are reported [3] and presented in our poster. A further note is that our strategy is different from the workflow where co-occurrence clusters of input genes is first calculated followed by examining overlaps between co-occurred gene clusters and pre-computed modules. That kind of workflow may introduce extra issues due to inconsistent clustering thresholds etc. Instead, our suggested confidence values leverage the pre-computed functional modules and cluster input genes into groups and then evaluate these modules based on phylogeny of member genes. The projected modules could be further sorted according to their confidence values and provide guidance for subsequent experimental work.

References:

1. Blomberg, S. P., T. Garland, Jr., and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57:717- 745.
2. Webb, C.O., Ackerly, D.D., and Kembel, S.W. 2008. Phylocom: software for the analysis of phylogenetic community structure and trait evolution. Version 4.0.1.

3. Fang, G., Passalacqua, K.D, Hocking, J., Montero Llopis, P., Gerstein, M., Bergman, N. H., and Jacobs-Wagner, C. 2013. Transcriptomic and phylogenetic analysis of a bacterial cell cycle reveals strong associations between gene co- expression and evolution. BMC Genomics 14:450

189. Fermentation Performance of Field-Grown Natural Variant and Transgenic Feedstocks

Kelsey L. Yee,[1,3*] (yeekl@ornl.gov) Miguel Rodriguez Jr.,[1,3] Olivia A Thompson,[1,3] Wellington Muchero,[1,3] Jin-Gui Chen,[1,3] Lee E. Gunter,[1,3] Sara S. Jawdy,[1,3] Charleson R. Poovaiah,[3,5] Hui Shen,[3,6] Mitra Mazarei,[3,5] Holly L. Baxter,[3,5] Nancy L. Engle,[1,3] Angela Ziebell,[2,3] Robert W. Sykes,[2,3] Erica Gjersing,[2,3] Chunxiang Fu,[3,4] Zeng-Yu Wang,[3,4] Mark F. Davis,[2,3] Timothy J. Tschaplinski,[1,3] C. Neal Stewart Jr., [3,5] Richard A. Dixon,[3,6] Gerald A. Tuskan,[1,3] Jonathan R. Mielenz,[1,3] Brian H. Davison[1,3] and **Paul Gilna[3] (BESC PI)**

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]National Bioenergy Center, National Renewable Energy Laboratory, Golden, Colorado; [3]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [4]The Samuel Roberts Noble Foundation, Ardmore, Oklahoma; [5]University of Tennessee, Department of Plant Sciences, Knoxville; [6]University of North Texas, Department of Biological Sciences, Denton

http://bioenergycenter.org

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). BESC researchers provide enabling technologies in 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.

Populus trichocarpa and switchgrass are dedicated bioenergy feedstocks and their inherent recalcitrance to bioconversion is one of the major economic hurdles for the production of biofuels. Implications of improvements in these feedstocks through use of natural variant and transgenic selection will be discussed with the goal of moving towards an improved phenotype which maintains normal growth characteristics and has increased fermentation yield. Microbial bioconversion assays of the feedstocks (unpretreated and pretreated) were performed with traditional enzymatic hydrolysis and yeast-based fermentation platform: simultaneous saccharification and fermentation (SSF) or separate hydrolysis and fermentation (SHF). In addition, the top ethanol yielding biomass lines from yeast-based fermentation were subjected to a consolidated bioprocessing platform using *Clostridium thermocellum*, a thermophilic bacterial species that produces its own hydrolytic enzymes.

We examined 21 *Populus* sp. variants from two environmentally different common garden field sites. These 21 variants have a mutation that affects the secondary carbon pathway metabolism and leads to reduced lignin content, as low as approximately 16-17% in some variants. In addition, we evaluated field-grown transgenic switchgrass lines with reduced lignin content due to genetic modifications in the lignin biosynthetic pathway. The results demonstrated superior conversion yields for the transgenic switchgrass and the *Populus* natural variants compared to their controls for both yeast-based and *C. thermocellum* fermentations. We concluded that transgenic switchgrass and *Populus* natural variants that have reduced recalcitrance are a valuable resource for producing economical biofuels. We also determined that when characterizing new biomass sources, *in vitro* assays, such as sugar release, should be supplemented with *in vivo* fermentation tests, which we have shown to detect inhibitory compounds.

Finally, we have continued our effort in fermenting the improved feedstock lines with genetically engineered microbes. The engineered and evolved *C. thermocellum* strain (M5170) was found to respond to the apparent reduced recalcitrance of the COMT transgenic switchgrass with no substrate inhibition, producing more ethanol on the transgenic feedstock than the wild-type substrate. Since ethanol was the main fermentation metabolite produced by an engineered and evolved *C. thermocellum* strain, its fermentation ethanol yield on a transgenic switchgrass substrate (g/g) is the highest produced thus far.

References:

Studer MH, DeMartini JD, Davis MF, Sykes RW, Davison B, Keller M, Tuskan GA, Wyman CE: Lignin Content in natural *Populus* variants affects sugar release. *Proc. Natl. Acad. Sci. USA* 2012, 15:108, 6300-6305.

Fu C, Mielenz JR, Xiao X, Ge Y, Hamilton CY, Rodriguez M Jr, Chen F, Foston M, Ragauskas AJ, Bouton J, Dixon RA, Wang Z-Y. Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proc. Natl. Acad. Sci. USA* 2011, (9)108:3803- 3808.

Tschaplinski TJ, Standaert RF, Engle NL, Martin MZ, Sangha AK, Parks JM, Smith JC, Samuel R, Jiang N, Pu Y, Ragauskas AJ, Hamilton CY, Fu C, Wang Z-Y, Davison BH, Dixon RA, Mielenz JR: Down-regulation of the caffeic acid O-methyltransferase gene in switchgrass reveals a novel monolignol analog. *Biotechnology for Biofuels* 2012, 5:71.

Yee KL, Rodriguez M Jr., Tschaplinski TJ, Engle NL, Martin MZ, Fu C, Wang Z-Y, Hamilton- Brehm SD, Mielenz JR. Evaluation of the bioconversion of genetically modified switchgrass using simultaneous saccharification and fermentation and a consolidated bioprocessing approach. *Biotechnology for Biofuels* 2012, 5:81

Shen H, He X, Poovaiah CR, Wuddineh WA, Ma J, Mann DGJ, Wang H, Jackson L, Stewart CN Jr., Chen F, Dixon RA: Functional characterization of the switchgrass (*Panicum virgatum*)

R2R3-MYB transcription factorPvMYB4 for improvement of lignocellulosic feedstocks. *New Phytologist* 2012, 193:121-136.

Shen H, Poovaiah CR, Ziebell A, Tschaplinski TJ, Pattathil S, Gjersing E, Engle NL, Katahira R, Pu Y, Sykes R, Chen F, Ragauskas AJ, Mielenz JR, Hahn MG, Davis M, Stewart CN Jr., Dixon RA: Enhanced characteristics of genetically modified switchgrass (*Panicum virgatum* L.) for high biofuel production**.** *Biotechnology for Biofuels* 2013, 6:71.

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.

190. Direct Conversion of Plant Biomass to Ethanol by Engineered *Caldicellulosiruptor bescii*

Daehwan Chung,[1,3†] Minseok Cha,[1,3†] Adam M Guss,[2,3] and Janet Westpheling[1,3]* (janwest@uga.edu) and **Paul Gilna[3] (BESC PI)**

[1]Department of Genetics, University of Georgia, Athens; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]The BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

†These authors contributed equally to this work

http://bioenergycenter.org

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). BESC research in biomass deconstruction and conversion targets CBP by studying model organisms and thermophilic anaerobes to understand novel strategies and enzyme complexes for biomass deconstruction.

Ethanol is the most widely used renewable transportation biofuel in the US - 13.3 billion gallons in 2012. In spite of considerable effort to produce fuels from lignocellulosic biomass, chemical pretreatment and the addition of saccharolytic enzymes prior to microbial bioconversion remain economic barriers to industrial deployment[2]. We began with a bacterium that efficiently uses unpretreated biomass and engineered it to produce ethanol. Here we report the direct conversion of switchgrass, a non-food, renewable feedstock, to ethanol by an engineered thermophilic, anaerobic, bacterium, and *Caldicellulosiruptor bescii*. This was accomplished by heterologous expression of a *Clostridium thermocellum* bifunctional acetaldehyde/alcohol dehydrogenase.

While wild-type *C. bescii* lacks the ability to make ethanol, 70% of the fermentation products in the engineered strain was ethanol from switchgrass without pretreatment. Direct conversion of biomass to ethanol represents a new paradigm for consolidated bioprocessing (CBP) offering the potential for carbon neutral, cost effective, sustainable fuel production.

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.

191. Development of a High-throughput Genetic Transformation System for Switchgrass (Panicum virgatum L.)

Zeng-Yu Wang[1,2*] (zywang@noble.org), Chunxiang Fu,[1] Hiroshi Hisano,[1,2] Frank Hardin,[1,2] Xirong Xiao,[1,2] Amy Mason[1,2] and **Paul Gilna[2] (BESC PI)**

[1]Forage Improvement Division, The Samuel Roberts Noble Foundation, Ardmore, Oklahoma;
[2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

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 multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC biomass formation and modification research involves working directly with two potential bioenergy crops (switchgrass and *Populus*) to develop varieties that are easier to break down into fermentable sugars. We are using both testing and generation of large numbers of natural and modified plant samples as well as developing genomics tools for detailed studies into poorly understood cell wall biosynthesis pathways.

Switchgrass (*Panicum virgatum* L.) is an outcrossing C4 perennial bunch grass native throughout North America. Genetic improvement of switchgrass through biotechnological approaches is expected to play a crucial role in modifying quality or quantity of biomass suitable for biofuel production. Development of genetic tools is essential for effective improvement of existing switchgrass cultivars. Switchgrass, like many other grasses, is generally considered difficult to genetically manipulate at the cellular level. The low transformation efficiency has been recognized as a bottleneck in genetic manipulation and functional test of transgenes in grasses.

The establishment of a well defined, rapid and highly efficient genetic transformation system is an important prerequisite for genetic engineering of this species. We have successfully solved this bottle-neck problem in switchgrass and established a high throughput system for the production of large numbers of transgenic plants.

By identification of highly tissue culture responsive genotypes and by optimization of transformation parameters, we have developed a highly efficient genetic transformation system for the widely used switchgrass cultivar Alamo. Embryogenic calli were induced from seeds or immature inflorescences. Overexpression vectors and RNAi vectors were constructed and transferred into *Agrobacterium tumefaciens* strain EHA105 or AGL1. The hygromycin phosphotransferase (*hph*) gene was used as a selectable marker. After co-cultivation with *Agrobacterium*, the infected calli were transferred onto selection medium containing the antibiotic hygromycin. Resistant calli obtained after about six weeks of selection were transferred to regeneration medium. Regenerated green shoots were transferred to rooting medium, and the rooted plantlets were later transferred to the greenhouse. The timeline from callus infection to rooted plants was about 4 months. Regenerated plants were screened by PCR analysis. Stable integration of the transgenes into the plant genome was confirmed by Southern blot analysis. The transformation efficiency (number of independent transgenic plants/number of calli used for infection) reached more than 90%. Since switchgrass is an outcrossing species, individual seeds within a cultivar may represent different genotypes. To solve this problem, we use inflorescences from a defined genotype to induce calli. Transgenic plants obtained from these calli have the same genetic background, allowing a precise comparison of the effects of transgenes. Importantly, our transformation system is

consistent and highly reproducible. We have produced transgenics from more than 70 gene constructs, with at least 30 independent transgenic events generated for each construct. This high throughput transformation system offers a solid basis for functional analysis of large numbers of genes in switchgrass.

To better characterize transgenic switchgrass plants, we proposed a standardized procedure for switchgrass sample collection by describing various developmental stages of switchgrass, defining the R1 stage as the stage at which tillers should be collected, and providing a description of how and what material should be analyzed. Such a standardized procedure will help to maintain consistency in switchgrass evaluation methods, enable comparisons of data obtained from different approaches and studies, and facilitate efforts towards improving switchgrass as a bioenergy crop.

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.

192. *In vitro* Synthesis of Xylan Catalyzed by Purified Plant Xylosyl Transferases

Breeanna Urbanowicz[1,2*] (breeanna@uga.edu), Maria Peña,[1,2] Heather Moniz,[1] Kelley Moremen,[1] William York,[1,2] and **Paul Gilna[2] (BESC PI)**

[1]The Complex Carbohydrate Research Center, University of Georgia, Athens; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org

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 multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC biomass formation and modification research involves working directly with two potential bioenergy crops (switchgrass and *Populus*) to develop varieties that are easier to break down into fermentable sugars. We are using both testing and generation of large numbers of natural and modified plant samples as well as developing genomics tools for detailed studies into poorly understood cell wall biosynthesis pathways.

Secondary cell walls are composed mainly of a cellulose/xylan network impregnated with lignin, resulting in a reinforced structure that is recalcitrant to deconstruction by microbial enzymes. Xylans are the dominant hemicellulosic polysaccharide found in the plant kingdom, second only to cellulose in abundance, and are present in load-bearing secondary cell walls of dicots and in both primary and secondary cell walls of grasses and cereals. The glucuronoxylan (GX) present in hardwoods including *Populus* and in mature stems of *Arabidopsis* is a homodisperse polymer that has a backbone composed of 1,4-linked β-D-xylosyl (Xyl) residues that are often substituted at *O*-2 with glucuronic acid (GlcA) or 4-*O*-methyl glucuronic acid (MeGlcA) and also contain a distinct acidic reducing-end sequence.

Beginning with the identification of cellulose synthase (CesA) genes from cotton almost two decades ago, the genes that encode enzymes responsible for catalyzing β-1,4 linked backbone formation have been identified for all principal hemicellulosic polysaccharides, with the notable exception of xylan. All of these identified backbone synthases have been cellulose synthase-like (Csl) genes, which are integral membrane proteins from glycosyltransferase family 2, structurally related to CesA. Despite predictions that the β-1,4-xylan synthase will be a member of the Csl family, most of the candidate enzymes are members of other families and none of them have been functionally characterized due to difficulties in heterologous expression of plant GTs. We investigated the function of several GT candidates for the xylan synthase in *Arabidopsis* with our MALDI-TOF MS based assay using heterologously expressed and purified enzymes.
These experiments provide direct biochemical evidence establishing the identity of the *Arabidopsis* gene encoding the glycosyltransferase that catalyzes the transfer of xylosyl units from UDP-xylose to xylan oligosaccharide acceptors in the absence of other proteins *in vitro*, establishing this enzyme as the xylan synthase.

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.

193. Discovering Naturally Occurring Allelic Variants Associated with Biomass Recalcitrance in *Populus trichocarpa*

Muchero Wellington,[5] Kelsey Yee,[5] Priya Ranjan,[5] Arthur J. Ragauskas,[3,5] Stephen P. DiFazio,[2,5] Luke Evans,[2,5] Gancho Slavov,[2,5] Eli Rodgers-Melnick,[2,5] Samarthya Bhagia,[1,5] Charles E. Wyman,[1,5] Daniel Rokhsar,[4,5] David M. Goodstein,[4,5] Joel Martin,[4] Wendy Schackwitz,[4] Gerald A. Tuskan[5*] (tuskanga@ornl.gov) and **Paul Gilna[5] (BESC PI)**

[1]Department of Chemical and Environmental Engineering, University of California, Riverside; [2]Department of Biology, West Virginia University, Morgantown; [3]Institute of Paper Science & Technology, School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta; [4]DOE Joint Genome Institute, Walnut Creek, California; [5]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org

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 multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC biomass formation and modification research involves working directly with two potential bioenergy crops (switchgrass and *Populus*) to develop varieties that are easier to break down into fermentable sugars. We are using both testing and generation of large numbers of natural and modified plant samples as well as developing genomics tools for detailed studies into poorly understood cell wall biosynthesis pathways.

Adaptive variation in natural populations is a rich source of genetic variation that can be mined for accelerated domestication of biofuel crops. 1034 unique native genotypes from 22 river systems/ 5 subpopulations have been established in 4 common gardens and have been fully resequenced using Illumina HiSeq platform. Approximately, 40 million SNP with more than 50% of loci having a minor allele frequency < 1%; $\pi = 0.007$and a match rate to reference genotype of 97.7%. Measuring a wide variety of traits in clonally-replicated common gardens over multiple years and environmental conditions we were able associate vegetative phenology, height and diameter growth, cell wall chemistry and sugar release phenotypes to specific SNP markers. For the adaptive phenotypes these genes include a UDP-Galactose transporters and glycosyl hydrolases. Elite genotypes carrying these beneficial alleles have been identified and display a 905 improvement in growth relative to the wild-type clones. For cell wall phenotypes these genes include an EPSP synthase, a copper-related ATOX-1, a calcium transporting ATPase, an amino acid transporter, a MYB transcription factor and a protein kinase. Allele specific variants at each of these loci result in sugar release analysis that show up to 1.4X lower lignin, 2.7X increase in sugar release, and up to 2.4X higher ethanol yield compared to wildtype genotypes. The top performing genotypes consistently show high sugar yields in low severity hydrothermal pretreatment conditions and at low enzyme loadings. Several favorable alleles produce similar favorable phenotypes in transgenic *Arabidopsis*.

The ultimate goal of these efforts is to identify candidate genes that can be commercially deployed in superior biomass feedstock lines that express reduced recalcitrance, increased sugar yield and long term economic and environmental sustainability.

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.

194. First Year Field Results of *PvMYB4*-overexpressing Transgenic Switchgrass and Other Multiple Lines

Charleson Poovaiah,[1,2] Kelsey Yee,[2] Miguel Rodriguez,[2] Hui Shen,[2,3] Mitra Mazarei,[1,2*] (mmazarei@utk.edu), Holly Baxter,[1,2] Angela Ziebell,[2,4] Erica Gjersing,[2,4] Robert Sykes,[2,4] Crissa Doeppke,[2,4] Geoff Turner,[2,4] Steve Decker,[2,4] Melvin Tucker,[2,4] Mark Davis,[2,4] Jonathan Mielenz,[2] Richard Dixon,[2,3] C. Neal Stewart Jr.,[1,2] and **Paul Gilna[2] (BESC PI)**

[1]University of Tennessee, Knoxville; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]University of North Texas, Denton; [4]National Renewable Energy Laboratory, Golden, Colorado

http://bioenergycenter.org

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 multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC biomass formation and modification research involves working directly with two potential bioenergy crops (switchgrass and *Populus*) to develop varieties that are easier to break down into fermentable sugars. We are using both testing and generation of large numbers of natural and modified plant samples as well as developing genomics tools for detailed studies into poorly understood cell wall biosynthesis pathways.

High biomass production and wide adaptation has made switchgrass (*Panicum virgatum* L.) a leading dedicated lignocellulosic feedstock candidate in the United States. One of the major limitations is the recalcitrance of complex carbohydrates to hydrolysis for conversion of lignocellulosic biomass into biofuel. Lignin is a primary contributor to recalcitrance as it creates a physical and chemical barrier to enzymatic access of cell wall polysaccharides. Therefore, genetic manipulation of the lignin biosynthetic pathway in an effort to reduce lignin content is a promising approach for overcoming this inherent cell wall recalcitrance.

Transgenic switchgrass plants with reduced lignin content were produced by over-expressing MYB4 transcription factor, an R2R3 type MYB repressor of the lignin biosynthetic pathway. Resulting MYB transgenic plants had reduced height, increased tillering, and reduced lignin under greenhouse conditions. Sugar release efficiency of some lines increased by 300%, which yielded up to 2.7 times more ethanol compared to control plants.

These analyses were performed on greenhouse-grown plants using live, green plant tissue and produced results that could be of great interest to the biofuel industry. However, results from using senesced and brown plant tissue harvested at the end of the growing season from the field experiments and productivity and performance of field-grown transgenics could potentially differ from greenhouse-grown transgenics because of exposure of plants to environmental stresses not present in a greenhouse setting. Therefore, it is essential to determine whether the improved traits observed in greenhouse experiments will be maintained in the field in order to further assess the relevance of these transgenics to the biofuel industry.

Field experiments of MYB4 transgenic switchgrass have been underway at the UT Plant Sciences Unit in Knoxville, Tennessee for two field seasons. In this project, plants are being assessed for 1) agronomic performance including biomass metrics (tiller height, plant width, tiller number) and biomass yield (dry weight), 2) cell wall chemical characterization including lignin content and lignin composition prior to and after senescence, 3) biorefinery performance including pretreatment response, sugar release efficiency, and ethanol yield, and 4) disease susceptibility of transgenic plants compared to the controls including the rust disease.

Several lines of MYB4-overexpressing plants did not survive the first year of field experiments with even more mortality over the first field winter. Out of 4 transgenic lines survived in the field, 2 lines had increased biomass yield and two lines had decreased biomass yield compared to the controls. Biomass produced appeared to be inversely associated with the expression level of the MYB4 gene. Cell wall recalcitrance analyses indicated lower lignin content in 2 transgenic lines compared to the controls and no significant difference in the S/G ratio between transgenics and the controls. The 2 transgenic lines with lower lignin content also had increase in sugar release compared to the controls. Three transgenic lines had an increase of 36%, 29% and 16% in ethanol yield compared to the controls. The most prevalent diseases observed were rust and leaf spot caused by a *Bipolaris* species on both transgenic and the controls. Additional data are forthcoming from the second field season.

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.

195. Chemical Factors that Control Lignin Polymerization

Amandeep K. Sangha[1,2*] (sanghaak@ornl.gov), Brian H. Davison,[2] Robert F. Standaert,[2,3] Mark F. Davis,[2,4] Jeremy C. Smith,[1,2,3] Jerry M. Parks,[1,2,3] and **Paul Gilna[2] (BESC PI)**

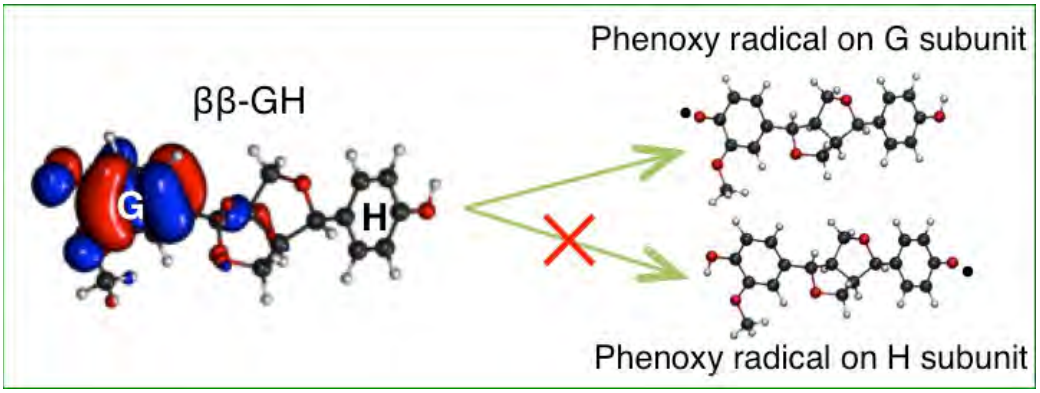
[1]UT/ORNL Center for Molecular Biophysics; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]Biology and Soft Matter Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [4]Bioenergy Science Center, National Renewable Energy Laboratory, Golden, Colorado

http://bioenergycenter.org

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). BESC researchers provide enabling technologies in 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.

Lignin structure, degree of polymerization and composition play an important role in biomass recalcitrance. Lignin formation involves the biosynthesis of monolignols, their transport to the cell wall lignification zone, peroxidase-mediated oxidation of lignin precursors, and finally radical coupling reactions to elongate growing lignin chains. We have performed quantum chemical calculations to investigate whether there might be chemical control of lignin polymerization.

We used density functional theory calculations to investigate the reactivity of various mono-, di- and trilignols comprising *p*-hydroxyphenyl (H) and guaiacyl (G) subunits linked through common β-O4, β-β and β-5 inter-unit bonds. We found that the composition of the highest occupied molecular orbital (HOMO) of neutral lignol oligomers plays a critical role in their oxidation: only lignol precursors with strong HOMO electron density on the phenolic oxygen can undergo facile oxidation to form radicals, which then participate in lignin chain elongation.



These results provide an explanation of why terminal H subunits and β-5 linkages involving either H or G subunits can stop chain elongation. The results also provide a coherent framework for understanding the propensity for growth or termination of different terminal subunits and inter-unit linkages in various lignin polymers.

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.

196. Surface Characterization of Populus during C. *bescii* Growth: Understanding the Relationship between Biomass Degradation and C. *bescii* Penetration in Wood Sticks

Seokwon Jung,[1,3] Daehwan Chung,[2,3] Allison Tolbert,[1,3] Janet Westpheling,[2,3] Arthur J. Ragauskas[1,3*] (arthur.ragauskas@chemistry.gatech.edu) and **Paul Gilna[3] (BESC PI)**

[1]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta; [2]Department of Genetics, University of Georgia, Athens; [3]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org

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). BESC researchers provide enabling technologies in 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.

Caldicellulosiruptor bescii is a cellulolytic/hemicellulolytic anaerobic bacterium that is considered for biomass conversion because it is a heat-tolerant bacterium which efficiently degrades non-pretreated biomass. It is reported that *C. bescii* is able to grow up to 90ºC on untreated plant biomass and to degrade crystalline cellulose/xylan on untreated plant biomass such as switchgrass (SG) and poplar. Surface characterization of biomass during *C. bescii* growth provides insights into the biomass recalcitrance. To understand the chemical and physical changes on the surface of biomass with *C. bescii* growth, we employed ToF-SIMS. As a model substrate, cross-sections of a juvenile *Populus* stem (ca. 80 μm thick) were used. The sectioned poplar stem was incubated with/without *C. bescii*. Fresh juvenile poplar stems (ca. 1.3 cm length) were debarked and incubated with *C. bescii* for 5 different times in order to understand the *C. bescii* penetration mechanism. The poplar stems after *C. bescii* growth (i.e., positive control) were cross-sectioned to 50 μm thick slices and collected from top to center. These sectioned samples can be used to observe the level of cell wall degradation depending on the vertical depth by SEM and ToF-SIMS. Here, we found that *C. bescii* growth conditions partially affect cell wall damage, but the damage should not change chemical composition on the surface of cell wall. We also found that *C. bescii* can penetrate into the short wood stick through the lumen area and result in severe cell wall damage at vertical center area.

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.

197. Comparative Solubilization of Minimally Pretreated Lignocellulose as Impacted by Choice of Feedstock and Biocatalyst

Julie Paye[1,2*] (Julie.Paye@Dartmouth.edu), Anna Guseva,[1,2] Siva Pattathil,[2,3] Yunqiao Pu,[2,4] Xianzhi Meng,[2,4] Erica Gjersing,[2.5] Arthur Ragauskas,[2,4] Michael Hahn,[2,3] Mark Davis,[2,5] Brian Davison,[2] Lee Lynd[1,2] and **Paul Gilna[2] (BESC PI)**

[1]Dartmouth College, Hanover, New Hampshire; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]Complex Carbohydrate Research Center, University of Georgia, Athens; [4]Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta; [5]National Renewable Energy Laboratory, Golden, Colorado

http://bioenergycenter.org

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). BESC research in biomass deconstruction and conversion targets CBP by studying model organisms and thermophilic anaerobes to understand novel strategies and enzyme complexes for biomass deconstruction.

Following reports of solubilization of minimally-pretreated lignocellulose by *C. bescii* (Yang AEM 2009, Kataeva 2013), we undertook controlled experiments to determine how widespread such ability is and how it is impacted by the choice of feedstock. We find that many anaerobic bacteria are able to achieve substantial solubilization of minimally pretreated lignocellulose, with *Clostridium thermocellum* giving among the highest yields - but not uniquely high yields - relative to other cultures tested. Solubilization yields were about two times higher for *C. thermocellum* cultures than for commercial fungal for minimally-pretreated switchgrass harvested at both mid-season and late-season. Performance of fungal enzymes was not significantly improved by addition of yeast, higher enzyme loading, increased hydrolysis temperature, or lower substrate loadings. For both *C. thermocellum* and fungal cellulase, conversion is twice as high for switchgrass harvested at mid-season as compared to late season. For poplar, *C. thermocellum* cultures achieve higher solubilization yields than fungal cellulases, although conversion for both is lower than that seen in grasses. Solubilization yields declined only slightly with increasing particle size for all feedstock-biocatalyst combinations except fermentation of minimally-pretreated wood by *C. thermocellum*, for which a sharp decline was observed. Fermentation residues were analyzed to develop a better understanding of the underlying cause for the substantial differences observed.

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.

198. The D494G Point Mutation in the Bifunctional Alcohol and Aldehyde Dehydrogenase of *Clostridium thermocellum* Leads to Improved Ethanol Production

Dan Olson[1,2*] (dan268@gmail.com), Tianyong Zheng,[1,2] Adam Guss,[2] Daniel Amador-Noguez,[3] Lee Lynd[1,2] and **Paul Gilna[2] (BESC PI)**

[1]Dartmouth College, Thayer School of Engineering, Hanover, New Hampshire; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]University of Wisconsin, Madison

http://bioenergycenter.org

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). BESC research in biomass deconstruction and conversion targets CBP by studying model organisms and thermophilic anaerobes to understand novel strategies and enzyme complexes for biomass deconstruction.

Previously we have engineered *Clostridium thermocellum* for increased ethanol production by eliminating lactate and acetate production. The resulting strain did not grow well and carbon flux was diverted to amino acid production instead of ethanol. The strain was evolved by serial transfer and ethanol production increased. To understand the reason for this change, the strain was resequenced and compared to the unevolved strain. A SNP in the bifunctional alcohol/aldehyde dehydrogenase (adhE) was found that creates a D494G substitution in the amino acid sequence. The mutation appears to alter the cofactor specificity for the alcohol dehydrogenase reaction and may explain the reason for the additional ethanol production.

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.

199. Toward Thermophilic Isobutanol Production in *Clostridium thermocellum*

Paul P. Lin[1,2*] (linpohen@gmail.com), Jennifer Takasumi,[1,2] James C. Liao[1,2] and **Paul Gilna[2] (BESC PI)**

[1]Department of Chemical and Biomolecular Engineering, University of California-Los Angeles;
[2]BioEnergy Science Center, Oak Ridge, National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org

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). BESC research in biomass deconstruction and conversion targets CBP by studying model organisms and thermophilic anaerobes to understand novel strategies and enzyme complexes for biomass deconstruction.

Isobutanol is an emerging biofuel that has a comparable energy density to gasoline and is compatible with existing transportation infrastructure. The thermophilic microorganism, *Clostridium thermocellum*, is a potential host for cost-effective biofuel production using cellulose directly as the raw material. We identified all biosynthetic pathway enzymes from pyruvate to isobutanol that are stable at least 50°C. This combination of enzymes produced more than 3 g/L of isobutanol at 50°C when overexpressed in the thermophile, *Geobacillus thermoglucosidasius*. In addition, the *Lactococcus lactis* ketoisovalerate decarboxylase has been functionally expressed in *C. thermocellum* using a plasmid construct; however plasmid stability may be an issue. We also cloned, purified and assayed 14 putative alcohol dehydrogenases from the facultative anaerobe *G. thermoglucosidasius*. Five of which demonstrated the ability to reduce isobutyraldehyde to isobutanol. In particular, Geoth_3237 and Geoth_3823 were found to have high activity using the cofactors NADH and NADPH, respectively. In addition, we have also identified small-scale fermentation conditions to produce isobutanol in wild-type *C. thermocellum*. High concentrations of either cellobiose or cellulose without citrate promote isobutanol production. This result indicates that all of the necessary isobutanol production enzymes are available in the *C. thermocellum* genome, but enzyme identification and improvement will be necessary to improve isobutanol production to desired levels.

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.

200. Alkaline Hydrogen Peroxide Pretreatment Differentially Affects Cell Wall Cross-Linking and Recalcitrance in Diverse Bioenergy Feedstocks

Sivakumar Pattathil,[1,2] Michael G. Hahn[1,2*](hahn@ccrc.uga.edu), Muyang Li,[3,4] David B. Hodge[3,4] and **Paul Gilna[1] (BESC PI)**

[1]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]University of Georgia, Athens; [3]Great Lakes Bioenergy Research Center, East Lansing, Michigan; [4]Michigan State University, East Lansing

http://bioenergycenter.org http://glbrc.org

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). BESC researchers provide enabling technologies in 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.

Conducting in-depth glycome analyses of plant biomass modified through pretreatment processes is a novel approach for studying/identifying cell wall components contributing to recalcitrance. Three taxonomically diverse bioenergy feedstocks [hybrid poplar (a woody dicot), goldenrod (a herbaceous dicot) and corn stover (a graminaceous monocot)] were subjected to alkaline hydrogen peroxide (AHP) pretreatment and subsequent enzymatic conversion studies in order to assess how they respond to mild alkaline oxidative pretreatment and to identify differing features of the cell wall matrix that contribute to their recalcitrance. After AHP pretreatment, these biomass types exhibited varied enzymatic conversion efficiencies with corn stover showing the highest sugar yield followed by golden rod and poplar. Glycome profiling was employed to screen various AHP-pretreated biomass samples along with untreated controls to determine changes in the composition and extractability of non-cellulosic cell wall glycans (Figure 1). The results obtained in this study demonstrate that distinct patterns of cell wall structural changes (and hence altered cell wall glycan extractability) occur among these biomass types when subjected to AHP pretreatment, causing varied enzymatic conversion efficiencies. Using glycome profiling, we found that hybrid poplar was relatively unaffected by AHP pretreatment in terms of composition, enzymatic digestibility, and the extractability of cell wall glycans, perhaps due to its higher lignin content and hence greater extent of base- and/or peroxide-stable cell wall crosslinking. In general, AHP-pretreated golden rod and corn stover biomasses showed an enhanced abundance of hemicellulose epitopes in mild cell wall extracts such as oxalate and carbonate, hinting at easier removal of a sub-class of hemicelluloses from pretreated biomass.

AHP pretreatment of goldenrod resulted in a decrease in all classes of alkali (1M KOH)- extractable glycans, notably xylans, xyloglucans, and pectic polysaccharides, indicating their solubilization during pretreatment. This was accompanied by an improvement in the subsequent digestibility of the remaining cell wall residue. AHP pretreatment of corn stover resulted in mild increases in the extractability of all classes of cell wall glycans, notably xylans, xyloglucans, pectic polysaccharides, and β-glucans, indicating overall weaker associations between cell wall polymers. In grasses, alkali-labile ester cross-links between cell wall matrix macromolecules and higher alkaline solubility of grass lignins are proposed to be important properties that are exploited by AHP pretreatment to solubilize lignin and "loosen" the

cell wall matrix, which in turn results in superior enzymatic conversion relative to goldenrod and poplar following AHP pretreatment.

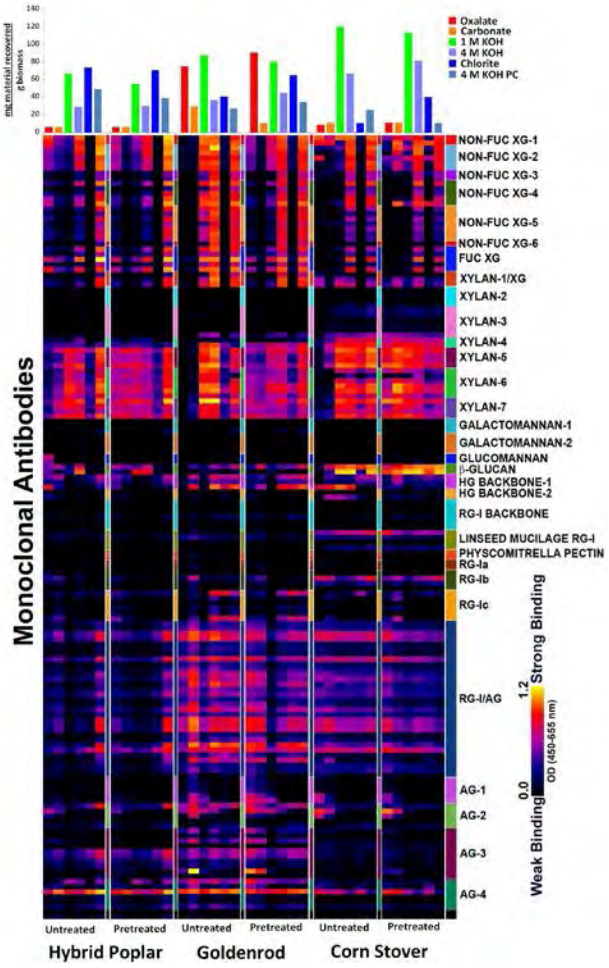


Figure 1. Glycome profiling of hybrid poplar, goldenrod and corn stover biomass samples before and after AHP pretreatment (12.5% H_2O_2 loading). Sequential cell wall extracts were made from untreated and pretreated biomass samples using increasingly harsh reagents. The extracts were ELISA screened using 155 mAbs directed against most major plant cell wall glycans. The resulting binding response data are represented as heatmaps with yellow-red- black scale indicating the strength of the ELISA signal (yellow, red and dark-blue colors depict strong, medium, and no binding, respectively). The mAbs are grouped based on the cell wall glycans they recognize as depicted in the panel at right hand side of the figure. The actual amounts of materials extracted out at each extraction condition are depicted as bar graphs at the top of heatmaps with color codes for reagents used for extraction.

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.

201. Elimination of Hydrogenase Post-translational Modification in *Clostridium thermocellum* Blocks H_2 Production and Increases Ethanol Yield

Ranjita Biswas,[1,2] Charlotte M. Wilson,[1,2] Tianyong Zhang,[2,4] Dawn Klingeman,[1,2] Richard J. Giannone,[2,3] Robert L. Hettich,[2,3] Daniel G. Olson,[2,4] Lee R. Lynd,[2,4] Steven D. Brown,[1,2] Adam M. Guss[1,2*] (gussam@ornl.gov) and **Paul Gilna[2] (BESC PI)**

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [4]Thayer School of Engineering at Dartmouth College, Hanover, New Hampshire

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESCS'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 for converting plant biomass into biofuels via consolidated bioprocessing. BESC research in biomass deconstruction and conversion targets CBP by studying model organisms and thermophilic anaerobes to understand novel strategies and enzyme complexes for biomass deconstruction.

Clostridium thermocellum is a leading candidate organism for implementing a consolidated bioprocessing (CBP) strategy for biofuel production due to its native ability to rapidly consume cellulose and its existing ethanol production pathway. *C. thermocellum* converts cellulose and cellobiose to lactate, formate, acetate, H_2, ethanol, amino acids, and other products. Elimination of the pathways leading to products such as H_2 could redirect carbon flux towards ethanol production. Rather than delete each hydrogenase individually, we targeted a hydrogenase maturase gene (*hydG*), which is involved in converting the three [FeFe] hydrogenase apoenzymes into holoenzymes by assembling the active site. This functionally inactivated all three [FeFe] hydrogenases simultaneously, as they were unable to make active enzymes. In the Δ*hydG* mutant, the [NiFe] hydrogenase-encoding *ech* was also deleted to obtain a mutant that functionally lacks all hydrogenase. The ethanol yield increased nearly 2-fold in Δ*hydG*Δ*ech* compared to wild type, and H_2 production was below the detection limit. Interestingly, Δ*hydG* and Δ*hydG*Δ*ech* exhibited improved growth in the presence of acetate in the medium.

Transcriptomic and proteomic analysis reveal that genes related to sulfate transport and metabolism were up-regulated in the presence of added acetate in Δ*hydG*, resulting in altered sulfur metabolism. Further genomic analysis of this strain revealed a mutation in the bi- functional alcohol/aldehyde dehydrogenase *adhE* gene, resulting in a strain with both NADH- and NADPH-dependent alcohol dehydrogenase activities, whereas the wild-type strain can only utilize NADH. This is the exact same *adhE* mutation found in ethanol tolerant *C. thermocellum* strain E50C, but Δ*hydG*Δ*ech* is not more tolerant to ethanol than the wild-type. Targeting protein post-translational modification represents a promising new approach to target multiple enzymes simultaneously for metabolic engineering.

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.

202. Coexistence but Independent Biosynthesis of Catechyl and Guaiacyl/Syringyl Lignin Polymers in Plant Seeds

Fang Chen[1,2*] (fang.chen@unt.edu), Yuki Tobimatsu,[3] Jin Nakashima,[4] Luis L. Escamilla- Treviño,[1,2] Lisa Jackson,[4] Richard A. Dixon,[1,2] John Ralph[3,5] and **Paul Gilna[2] (BESC PI)**

[1]University of North Texas, Denton; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]University of Wisconsin-Madison, Madison; [4]The Samuel Roberts Noble Foundation, Ardmore, Oklahoma, and [5]Great Lakes Bioenergy Research Center, Wisconsin Energy Institute, Madison, Wisconsin

http://bioenergycenter.org http://glbrc.org

Project Goals: The BioEnergy Research Centers BESC and GLBRC are 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 for converting plant biomass into biofuels via consolidated bioprocessing. Addressing the roadblock of biomass recalcitrance will require a greater understanding of plant cell walls from synthesis to deconstruction. This understanding will generate models, theories and finally processes in order to understand and overcome biomass recalcitrance. This grand challenge calls for an integrated research approach as illustrated by the project described below. BESC research is multi-disciplinary by design and multi-institutional in composition. GLBRC has similar aims and approaches that are both different and complementary, such as the joint BESC- GLBRC study described here.

Lignins are phenylpropanoid polymers, derived from monolignols, commonly found in terrestrial plant secondary cell walls. During lignin biosynthesis, the monolignol precursors are functionalized by aromatic hydroxylation and *O*-methylation to generate, successively, *p*- hydroxyphenyl (H), catechyl (C), guaiacyl (G), 5-hydroxyguaiacyl (5H), and syringyl (S) units. Natural lignins in dicotyledonous angiosperms consist of approximately equal amounts of G and S units, with less than 2% of H units. Lignins from gymnosperms are composed of G-units only with minor amounts of H units. C and 5H units are not considered as natural lignin subunits. 5H units are found in lignins in transgenic plants in which the second methylation step is blocked by mutation or gene silencing, but C units do not accumulate in the lignin of angiosperms in which the first or both methylation steps are similarly blocked. We recently reported evidence of an unanticipated catechyl lignin homopolymer (C-lignin) derived solely from caffeyl alcohol in the seed coats of several monocot and dicot plants. Here we report the identification of several dicot plants (*Euphorbiaceae* and *Cleomaceae*) that produce C-lignin together with traditional G/S lignins in their seed coats. Solution-state NMR analyses, along with an *in vitro* lignin polymerization study, determined that there is no copolymerization detectable, indicating that the synthesis and polymerization of caffeyl alcohol and conventional monolignols *in vivo* is spatially and/or temporally separated. C-lignin is structurally similar to the polymer synthesized *in vitro* by peroxidase-catalyzed polymerization of caffeyl alcohol. Circular dichroism spectroscopy did not detect any optical activity in the seed polymer. These data support the contention that the C- lignin polymer is produced *in vivo* via combinatorial oxidative radical coupling that is under simple chemical control, a mechanism analogous to that theorized for classical lignin biosynthesis. Of particular interest, the deposition of G and C lignins in *Cleome hassleriana* seed coats is developmentally regulated during seed maturation; C-lignin appears successively after G lignin within the same testa layers, concurrently with loss of the functionality of *O*- methyltransferases that are key enzymes for the conversion of C to G lignin

precursors. This study exemplifies the flexible assembly of different types of lignin polymers in plants, dictated by substantial, but still poorly understood, control of monomer supply by the cells in nature. A major challenge in current lignin bioengineering is achieving tight regulation of precursor synthesis to enable the flexible design of cell wall lignins with controlled structures. In this direction, future studies will focus on understanding the molecular genetic mechanisms that regulate the spatiotemporally specific production of C-lignin and the apparently abrupt cessation of *O*-methylation activity. The engineering of C-lignin as a novel, renewable, high-value polymer is also being explored.

The BioEnergy Science Center (BESC) and the Great Lakes Bioenergy Research Center (GLBRC) are both U.S. Department of Energy Bioenergy Research Centers supported by the Office of Biological and Environmental Research in the DOE Office of Science.

203. Revealing Nature's Cellulase Diversity: The Digestion Mechanism of the Hyperactive Cellulase *Caldicellulosiruptor bescii CelA*

Roman Brunecky,[1,3] Markus Alahuhta,[1,3] Qi Xu,[1,3] Bryon S. Donohoe,[1,3] Michael F. Crowley,[1,3] Irina A. Kataeva,[2,3] Sung-Jae Yang,[2,3] Michael G. Resch,[2,3] Michael W. W. Adams,[2,3] Vladimir V. Lunin,[2,3] Michael E. Himmel,[1,3] Yannick J. Bomble[1,3*] (Yannick.Bomble@nrel.gov) and **Paul Gilna[2] (BESC PI)**

[1]National Renewable Energy Laboratory, Golden, Colorado; [2]University of Georgia, Athens; [3]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESCS'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 for converting plant biomass into biofuels via consolidated bioprocessing. BESC research in biomass deconstruction and conversion targets CBP by studying model organisms and thermophilic anaerobes to understand novel strategies and enzyme complexes for biomass deconstruction.

The thermal tolerant CelA from *Caldicellulosiruptor bescii* is the most active cellulose degrading enzyme we have tested to date. In the saccharification of a common cellulose standard, Avicel, CelA outperforms mixtures of commercially relevant exo- and endoglucanases. We have solved the crystal structures of the two catalytic domains of CelA and modeled the solution structures of the unbound enzyme. From transmission electron microscopy studies of cellulose following incubation with CelA, we discovered morphological features that suggest CelA is capable of not only the common surface ablation/fibrillation strategies driven by processivity, but also of forming extensive cavities of roughly the size of the enzyme. It is proposed that CelA, and possibly other multi-functional glycoside hydrolases, act in this completely novel manner and thus, when combined with more common cellulases, will result in significant improvements in mixed cocktails. These results suggest that Nature's repertoire of cellulose digestion paradigms remain only partially understood.

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.

204. 'Caldi World': Unraveling the Mystery and Mechanisms of Plant Biomass Deconstruction by the Bacterial *Genus Caldicellulosiruptor*

Sara E. Blumer-Schuette[1, 3*] (seblumer@ncsu.edu), Mirko Basen,[2,3] MinSeok Cha,[2,3] Puja Chandrayan,[2,3] Jonathan M. Conway,[1,3] Daehwan Chung,[2,3] Joe Groom,[2,3] Laura L. Lee,[1,3] Amanda M. Rhaesa,[2,3] Israel Scott,[2,3] Jenna Young,[2,3] Jeffrey V. Zurawski,[1,3] Janet Westpheling,[2,3] Michael W. W. Adams,[2,3] Robert M. Kelly[1,3] and **Paul Gilna[3] (BESC PI)**

[1]North Carolina State University, Raleigh; [2]University of Georgia, Athens; [3] BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org

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). BESC research in biomass deconstruction and conversion targets CBP by studying model organisms and thermophilic anaerobes to understand novel strategies and enzyme complexes for biomass deconstruction.

Exploration of plant biomass deconstruction mechanisms beyond the fungal and cellulosomal enzyme paradigms has accelerated in recent years due in part to renewed interest in second- generation biofuels. An alternative strategy, consisting of modular, multi-functional carbohydrate active enzymes, is used by the extremely thermophilic bacterial genus *Caldicellulosiruptor* (T_{opt} 70~80°C) to harvest energy from a wide variety of C_5 and C_6 sugars found in plant biomass.

Recently, interest in this genus led to an exploration of biodiversity amongst globally isolated species and determined that the genus comprises a spectrum of weakly to strongly cellulolytic microorganisms. Common to all species is the ability to degrade amorphous cellulose, xylan and pectin as inferred from enzyme inventory and confirmed by growth physiology. Accessory enzymes, not common to all species, enable unique mechanisms by which individual species approach plant biomass deconstruction. Previous comparative proteomic screening implicated many cell associated proteins and enzymes as being integral for plant biomass deconstruction.

Many enzymes are found attached to the cell's S-layer, in addition to flagella and type IV pili being implicated in attachment to biomass. Two unique proteins directly downstream of the type IV pilus operon have been cloned and characterized as novel cellulose binding proteins. Taken together, cell-surface associated and modular, multifunctional enzymes participate in a process by which the cell strips away layers of lignin and polysaccharides, leaving behind biomass that resembles the starting material ('onion peeling effect'). This mechanism was discovered in *C. bescii* and confirmed in other strongly cellulolytic *Caldicellulosiruptor* species. Plant biomass originating from monocots will support growth of *C. bescii*, *C. kronotskyensis* and *C. saccharolyticus* as the only carbon source, and will also support growth on high loading levels (up to 20% w/v). A comparison of fermentation products from *Caldicellulosiruptor* species when grown on crystalline cellulose (Avicel) or untreated switchgrass demonstrates flexibility in their metabolism to handle only C_6 or a combination of C_5 and C_6 sugars. In addition to a better understanding of the enzymatic capacity and cellular metabolism of the genus *Caldicellulosiruptor*, a genetics system was developed for *C. bescii*. A unique restriction- modification

system was identified (CbeI – M. CbeI) and also determined as the barrier for DNA transformation. Transformation of *C. bescii* was enabled by a Caldi-*E. Coli* shuttle vector treated with M. CbeI and has facilitated the deletion of a lactate hydrogenase and also initial steps for metabolic engineering of Caldi strains. Overall, the genus *Caldicellulosiruptor*, by virtue of their diverse physiology and enzymatic capacity make an attractive model system for high temperature plant biomass deconstruction, and with the advances in genetic systems for the genus, can now be considered as promising platforms for biofuels production.

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.

205. Matrix Cell Wall Polysaccharide Synthesis Gene 1 Knockdown in Switchgrass – a BESC TOP Line

Ajaya K. Biswal[1,6*] (biswalajaya@ccrc.uga.edu), Michael K. Udvardi,[3,6] Yi-Ching Lee,[3,6] Ji-Yi Zhang,[3,6] Sushree Sangita Mohanty,[1,6] David Ryno,[1,6] Adam Bray,[4,6] Zachary King,[4,6] Hema Ramanna,[3,6] Robert Sykes,[5,6] Kelsey Yee,[6] Angela Ziebell,[5,6] Steve Decker,[5,6] Wayne Parrot,[4,6] Jonathan Mielenz,[6] Richard S. Nelson,[3] Mark Davis,[5,6] Debra Mohnen[1,2] and **Paul Gilna[6] (BESC PI)**

[1]Complex Carbohydrate Research Center, BioEnergy Science Center (BESC), University of Georgia, Athens; [2]Department of Biochemistry & Molecular Biology, University of Georgia, Athens; [3]Samuel Roberts Noble Foundation, Ardmore, Oklahoma; [4]Department of Crop and Soil Sciences, University of Georgia, Athens; [5]National Renewable Energy Laboratory, BioEnergy Science Center (BESC), Golden, Colorado; [6]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org

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 multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC biomass formation and modification research involves working directly with two potential bioenergy crops (switchgrass and *Populus*) to develop varieties that are easier to break down into fermentable sugars. We are using both testing and generation of large numbers of natural and modified plant samples as well as developing genomics tools for detailed studies into poorly understood cell wall biosynthesis pathways.

Lignocellulosic biomass such as woody *Populus* and perennial switchgrass is composed of the complex structures of lignin, cellulose, matrix polysaccharides (pectin and hemicellulose) and cell wall proteins. Due to its complex structure, lignocellulosic feedstock is highly recalcitrant to bioconversion into ethanol by microbes. Here we modified Matrix Cell wall Polysaccharide Synthesis gene 1 (*mcps1*) expression in switchgrass, *Populus*, rice and foxtail millet using an RNAi and Virus-Induced Gene Silencing (VIGS) approach to determine the effects of reduced expression of the encoded protein on recalcitrance and ethanol yield. The *MCPS1* gene was selected as a target due to its high expression in tissues enriched in secondary walls, since secondary wall-containing tissues are the main source of biomass in these feedstocks. The knockdown expression of MCPS1 in switchgrass (*Pvmcps1-KD*) resulted in increased tiller numbers and increased biomass yield. Down regulation of the switchgrass *MCPS1* gene yielded a significant 15% increase in glucose and total sugar/gm biomass and ethanol yields up to 36% (mg/g cellulose) and 51% (mg/g dry biomass), respectively, greater than controls. Similarly, 36% more glucose and 31% more total sugar release was obtained in foxtail millet *Simcps1-KD* VIGS knockdown lines. Down-regulation of MCPS1 homologs in *Populus* (*Pvmcps1*-KD) resulted in 6-44% greater plant height and 7-27% increased stem diameter compared to WT as well as 3-7% greater glucose release. Rice transgenic knockdown (*Osmcps1-KD*) lines had greater plant height, tiller numbers and dry biomass. The generation of biomass with reduced recalcitrance and increased biomass can lower biofuel production costs and increase biomass yield/acre, thereby positively affecting the biofuel industry. Efforts are underway to determine the mechanism by which reduced expression of MCPS1 leads to increased sugar release and enhanced growth.

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.

206. Fundamentals of Aqueous Pretreatment Chemistry and Cell Wall Cellular Structures of Low Recalcitrance *Populus* Lines for Enhanced Performance

Samarthya Bhagia,[1,2,4*] (sbhagia@engr.ucr.edu), Xianzhi Meng,[3,4] Kelsey Yee,[4] Muchero Wellington,[4] Garima Bali,[3,4] Yunqiao Pu,[3,4] Rajeev Kumar,[2,4] Gerald A. Tuskan,[4] Arthur J. Ragauskas,[3,4] Charles E. Wyman[1,2,4] and **Paul Gilna[4] (BESC PI)**

[1]Department of Chemical and Environmental Engineering, University of California, Riverside; [2]Center for Environmental Research and Technology (CE-CERT), University of California, Riverside; [3]Institute of Paper Science & Technology, School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta; [4]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org

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). BESC researchers provide enabling technologies in 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.

Lowering plant recalcitrance is a key to reducing pretreatment severity and enzyme doses that hinder large scale commercialization of cellulosic ethanol. Natural *Populus* lines with gene mutation in their lignin biosynthesis pathway were found to have reduced recalcitrance. High throughput pretreatment and co-hydrolysis (HTPH) was employed to screen 18 *Populus* lines along with 4 controls and BESC *Populus* standard to select those with the highest glucan and xylan release. At the same pretreatment severity, a temperature of 180°C greatly increased glucan and xylan release in several plant lines compared to 140°C and 160°C. This result likely indicates that some bonds have higher activation energies that require higher temperatures to overcome. We found several lines that gave 100% higher xylan yields in low severity hydrothermal batch pretreatment and 300% higher glucan yields in enzymatic hydrolysis at lower enzyme loadings relative to the BESC *Populus* standard. Correlations between glucan and xylan ranks for all lines indicated that higher glucan release led to higher xylan release and vice versa. ^{13}C-^{1}H NMR spectra of a low recalcitrant line showed a higher syringyl to guaiacyl lignin ratio and a higher p-hydroxybenzoate to guaiacyl ratio than for the *Populus* standard. This line was also found to contain a greater abundance of β-O-4 linkages and have a lower molecular weight than the standard. These variations in lignin are likely associated with enhanced sugar release.

Xylan yields were twice as high from selected lines than the *Populus* standard for hydrothermal batch pretreatment at low severities, while differences in xylan yields were less apparent for dilute acid batch pretreatment at low severities. These results suggest that changes in lignin architecture can lead to improved sugar release and several factors for better design of transgenic plants.

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.

207. Degradation of High Loads of Crystalline Cellulose and of Unpretreated Switchgrass by Caldicellulosiruptor bescii

Amanda M. Rhaesa,[1,3] Mirko Basen,[1,3] Irina Kataeva,[1,3] Cameron J. Prybol,[1,3] Israel M. Scott,[1,3] Farris L. Poole,[1,3] Robert M. Kelly,[2,3] Michael W. W. Adams[1,3*] (adams@bmb.uga.edu), and **Paul Gilna[3] (BESC PI)**

[1]University of Georgia, Athens; [2]North Carolina State University, Raleigh; [3]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org

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). BESC research in biomass deconstruction and conversion targets CBP by studying model organisms and thermophilic anaerobes to understand novel strategies and enzyme complexes for biomass deconstruction.

Deconstruction of plant biomass traditionally involves thermochemical pretreatment in order to overcome its recalcitrant nature. The anaerobic bacterium *Caldicellulosiruptor bescii* degrades switchgrass biomass without prior chemical treatment. Use of high substrate loads decreases the costs of biofuel generation, and will in fact be necessary to generate fuel titers required for a viable industrial process. Here we show that *C. bescii* grows at 78°C without inhibition on crystalline cellulose and on unpretreated switchgrass of concentrations up to 200 g L^{-1}. In contrast, it grew poorly on acid-pretreated switchgrass on concentrations as low as 20 g L^{-1}. Indeed, the presence of acid-pretreated switchgrass (50 g L^{-1}) inhibited the growth of the organism on the unpretreated biomass (10 g L^{-1}). Growth of *C. bescii* and degradation of crystalline cellulose, but not that of switchgrass, were limited by changes in pH, nitrogen and vitamin (folate) availability. Under optimal conditions *C. bescii* solubilized ~60% of the cellulose and ~30% of the unpretreated switchgrass using 50 g L^{-1} substrates. Further fermentation of cellulose, but not switchgrass, was inhibited by the organic acids generated as fermentation products, while prolonged degradation of high loads of switchgrass led to accumulation of an uncharacterized inhibitor of *C. bescii* growth. However, we present evidence here that *C. bescii* is much more salt tolerant than previously assumed. Soluble sugars, acetate, lactate, carbon dioxide and biomass, quantitatively accounted for the cellulose and plant biomass carbon utilized even using 50 g L^{-1} substrates.

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.

208. Simulation and Structure of Cel7A during Binding and Hydrolysis of Cellulose

Sai Venkatesh Pingali,[1] Junhong He,[1] Hugh M. O'Neill,[1] Volker S. Urban,[1] Loukas Petridis,[1] William T. Heller,[1] Marcus Foston,[2] Arthur Ragauskas,[2] Barbara R. Evans,[1] Jeremy C. Smith,[1] Paul Langan[1*] (langanp@ornl.gov), and **Brian H. Davison[1] (PI)**

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta, Georgia

Project Goals: Lignocellulosic biomass comprises the vast majority of biomass on Earth and has the potential to play a major role in generation of renewable biofuels if cost- effective conversion can be achieved. Largely composed of plant cell walls, it is a complex biological composite material that is recalcitrant to the structural deconstruction and enzymatic hydrolysis into sugars that is necessary for fermentation to bioethanol. The Scientific Focus Area in Biofuels is developing "Dynamic Visualization of Lignocellulose Degradation by Integration of Neutron Scattering Imaging and Computer Simulation" for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis. This is providing fundamental information about the structure and deconstruction of plant cell walls that is needed to drive improvements in the conversion of renewable lignocellulosic biomass to biofuels.

Investigation into the mode of action of cellulases on the surface of cellulose is of importance for understanding and optimizing the production of cellulo-oligosaccharides for biofuel production. A significant technical challenge has been the inability to probe the structure of the cellulose- cellulase system at a molecular level while the enzyme is digesting cellulose. Neutron scattering has the potential to overcome this challenge because of its unique ability to obtain scattering contrast between cellulose and enzyme by substituting hydrogen atoms of cellulose by deuterium. We have probed the structure of *Trichoderma reesei* Cel7A, a processive exocellulase enzyme, while it binds to and digests partially deuterated cellulose. The results provide insights into the pH dependent structural properties of cellulases when bound to the cellulose substrate.

Complementary molecular dynamics simulations validate the scattering data which show that *T. reesei* Cel7A assumes a compact conformation during cellulose hydrolysis (pH 4 to 5) and a more extended conformation when it has lower catalytic activity (pH 7). This observation lends support to a 'caterpillar mechanism' for Cel7A's action on cellulose which predicts that the energy required for processive action of Cel7A on cellulose is provided by stretching and compression of the linker region allowing the enzyme to move along a cellulose chain. This study provides important experimental insights into the mode-of-action of cellulases on cellulose surfaces.

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.

209. Investigation of Pretreatment-specific Changes in Biomass Structure

Loukas Petridis,[1] Sai Venkatesh Pingali,[1] Daisuke Sawada,[1] Volker Urban,[1] Hugh M. O'Neill,[1] Garima Bali,[2] Arthur J. Ragauskas,[2] Barbara R. Evans,[1] Jeremy C. Smith,[1] Paul A. Langan[1] and **Brian H. Davison[1*] (PI)**

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta, Georgia

Project Goals: Lignocellulosic biomass comprises the vast majority of biomass on Earth and has the potential to play a major role in generation of renewable biofuels if cost- effective conversion can be achieved. Largely composed of plant cell walls, lignocellulosic biomass is a complex biological composite material that shows significant recalcitrance towards the structural deconstruction and enzymatic hydrolysis into sugars that is necessary for fermentation to bioethanol. This Scientific Focus Area in Biofuels seeks to develop and demonstrate the "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 will provide 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.

Here, we report on three integrated studies neutron scattering, fiber diffraction and high- performance computational simulation to understand the physicochemical processes taking place across multiple length scales during thermochemical pretreatment of lignocellulosic biomass.

Morphological changes to the different components of lignocellulosic biomass were observed during steam pretreatment using a reaction cell to enable time-resolved neutron scattering.

Cellulose morphology changed mainly in the heating phase, whereas changes in lignin morphology occurred mainly in the holding and cooling phases. To optimize access of cellulose fibers, ammonia and diamine 1,2 diaminoethane (EDA) solvents were found to efficiently penetrate cellulose fibers, making it useful in several industrial processes to convert cellulose to biofuels. The neutron structure of EDA-cellulose complex revealed the location of hydrogen atoms and the mechanism behind the disruption of the hydrogen-bonding pattern by EDA. Comparison to molecular dynamic simulation showed the hydrogen-bonding arrangement to be highly dynamic with bonds continually being formed and broken. High-performance molecular dynamics simulations of lignin-cellulose complex in aqueous media found lignin to associate strongly with itself and cellulose. However, due to stronger hydration, non-crystalline regions show a lower tendency to associate with lignin than do crystalline regions. The information obtained could be useful in optimizing the different stages of pretreatment to allow control of lignin aggregate sizes leading to reduced lignin-cellulose association; enhanced enzyme accessibility and hence higher efficiency of biomass conversion.

Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725.

210. Production and Characterization of Deuterated Switchgrass and Annual Grasses for Neutron Studies

Barbara R. Evans[1*] (evansbr@ornl.gov), Garima Bali,[2] Riddhi Shah,[1] Arthur J. Ragauskas,[2] Hugh M. O'Neill,[1] Sai Venkatesh Pingali,[1] Daisuke Sawada,[1] Volker Urban,[1] Paul A. Langan[1] and **Brian H. Davison[1] (PI)**

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta, Georgia

Project Goals: Lignocellulosic biomass comprises the vast majority of biomass on Earth and has the potential to play a major role in generation of renewable biofuels if cost- effective conversion can be achieved. Largely composed of plant cell walls, lignocellulosic biomass is a complex biological composite material that shows significant recalcitrance towards the structural deconstruction and enzymatic hydrolysis into sugars that is necessary for fermentation to bioethanol. This Scientific Focus Area in Biofuels seeks to develop and demonstrate the "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 will provide 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.

We have developed and demonstrated a suite of methods for production and characterization of highly deuterium-enriched herbaceous biomass with morphology and chemical composition similar to protiated controls. Three grass species, annual ryegrass (*Lolium multiflorum*), winter grain rye (*Secale cereale*), and switchgrass (*Panicum virgatum*) have successfully been grown in 50% and higher concentrations of D_2O for production of lignocellulosic biomass containing 35 – 45% deuterium substitution as determined by NMR and FTIR. Hydroponic cultivation of switchgrass from cuttings was developed and employed to enable multiple, periodic harvests of deuterated biomass from this perennial species. Standard characterization methods based on GPC, FTIR, and NMR were used to characterize the physical and chemical properties of the harvested lignocellulosic biomass from the plants cultivated in D_2O. To understand the characteristics related to enzymatic hydrolysis of deuterated biomass, we investigated the effects of deuterium substitution on the biomass composition, degree of polymerization (DP) and cellulose crystallinity for each species compared to protiated controls cultivated in H_2O. While the isolated hemicellulose molecular weight and cellulose crystallinity remained unaffected, deuterated annual ryegrass had cellulose weight average and number average molecular weights that were 80% of those of protiated controls. Deuterium incorporation into different biomass components maximizes the power of neutron scattering by making it possible to visualize individual components in native and pretreated states. The level of deuterium incorporation and structural characterization accomplished for these grass species support the effective use of such deuterated biomass for neutron scattering studies.

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.

211. Construction and Characterization of Interspecies Yeast Hybrids Using Newly Discovered Species with Native Biofuel Potential

David Peris,[1] William G. Alexander,[1] Kayla Sylvester,[1] Maria Sardi,[1] Lucas Parreiras,[1] Alex La Reau,[1] Mick McGee,[1] Yaoping Zhang,[1] Trey Sato,[1] and **Chris Todd Hittinger**[1,2,3*] (cthittinger@wisc.edu)

[1]DOE Great Lakes Bioenergy Research Center, University of Wisconsin, Madison; [2]Genome Center of Wisconsin, Madison; [3]Wisconsin Energy Institute, Madison

https://www.glbrc.org
http://hittinger.genetics.wisc.edu

Project Goals: The DOE Great Lakes Bioenergy Research Center performs basic research that generates the technology to convert lignocellulosic biomass to ethanol and advanced biofuels. Within this broad theme, we are engineering yeast to efficiently convert hexoses and pentoses to ethanol in the presence of the inhibitory toxins found in lignocellulosic hydrolysates. Specifically, we have sought 1) to determine the potential of diverse non-*cerevisiae* strains of *Saccharomyces* for cellulosic ethanol production; and 2) to improve their performance through engineering, hybridization, and experimental evolution. We are now focused on identifying specific genes and alleles from yeasts and other fungi that can be engineered into xylose-fermenting GLBRC "chassis" strains of *S. cerevisiae*, as well as other biofuel strains.

Most research in biofuel production involves improvement of strains of *Saccharomyces cerevisiae*[1]. Recent discoveries and genome projects have made it clear that other species of *Saccharomyces* also have many of the same traits useful for fermentation, but these new species harbor novel genes and alleles[2,3]. The maximum pairwise genetic divergences between *Saccharomyces* species are similar to those between humans and birds[4] and ~60x greater than the most divergent *S. cerevisiae* strains. In addition, natural interspecies hybrids between *S. cerevisiae* and other *Saccharomyces* species can tolerate some stressful fermentative environments better than *S. cerevisiae*, such as during certain winemaking and brewing processes[5,6]. Here we explore the growth properties of 457 non-*cerevisiae* strains in AFEX corn stover hydrolysate (ACSH). We also engineered 6 interspecies hybrids by crossing representatives to strains of *S. cerevisiae* that had been engineered and evolved for xylose metabolism. To explore metabolic diversity and fermentation bottlenecks, we selected the best and the worst growing strains from each species for fermentation in ACSH in microaerobic conditions. They were then compared against the engineered interspecies hybrids and GLBRC benchmark strains.

Our results show that several non-*cerevisiae* species of *Saccharomyces* possess novel and native biofuel potential. *S. paradoxus* and *S. mikatae* had unusually high tolerance to the fermentation inhibitors found in ACSH (over 20% faster growth for some strains). Some strains of *S. eubayanus* and *S. paradoxus* had high natural tolerance to the most prevalent lignotoxin, acetamide, and could slowly metabolize xylose without engineering. Many of these traits proved to be dominant in interspecies hybrids constructed by crossing engineered xylose-fermenting strains of *S. cerevisiae* with these new species. In some cases, the engineered hybrid strains displayed hybrid vigor and novel, industrially desirable properties. For example, *S. mikatae x S. cerevisiae* had unusually high tolerance to acetamide and produced ethanol from ACSH faster and with higher yields than the xylose-fermenting *S. cerevisiae* parent. Finally, by using a gene from a filamentous fungus, we have engineered an acetamide-consuming strain. We conclude that several other yeast and fungal species possess genes and alleles that could be used to improve the performance of *S.*

cerevisiae biofuel strains. Our current research focuses on identifying these genes and developing high-throughput engineering techniques to test their effect on lignocellulosic fermentations.

References

1. Hong KK, Nielsen J (2009) Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cell Mol Life Sci* **69**: 2671-90.
2. Scannell DR, Zill OA, Rokas A, Payen C, Dunham M, Eisen MB, Rine J, Johnston M, Hittinger CT (2011) The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. *G3 Genes Genomes Genetics* **1**: 11-25.
3. Hittinger CT (2013) *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet* **29**: 309-17.
4. Dujon B (2006) Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends Genet* **22**: 375-387.
5. Belloch C, Orlić S, Barrio E, Querol A (2008) Fermentative stress adaptation of hybrids within the *Saccharomyces sensu stricto complex*. *Int J Food Microbiol* **122**: 188-195.
6. Libkind D, Hittinger CT, Valério E, Gonçalves C, Dover J, Johnston M, Gonçalves P, Sampaio JP (2011) Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc Natl Acad Sci USA* **108**: 14539-44.

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

212. A functional genomics approach to improving microbial fermentation

Jeff S. Piotrowski[1*] (jpiotrowski@wisc.edu), Li Hinchman[1], Quinn Dickinson[1], Scott Bottoms[1], Sean McIlwain[1], Irene Ong[1], and **Robert Landick**[1,2]

[1]DOE Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, WI; [2]Department of Biochemistry, University of Wisconsin, Madison, WI.

https://www.glbrc.org/

Project Goals

Efficient and sustainable conversion of cellulosic derived sugars to biofuel is dependent on robust microbial fermentation. Fermentation stressors, such as lignocellulosic derived inhibitors, are major hurtles to conversion of sugars. Further, the limited ability of powerful ethanologens like the yeast *Saccharomyces cerevisiae* to convert xylose to ethanol impedes potential yields. To improve biofuel productions, at the DOE Great Lakes Bioenergy Research Center we are using Systems Biology to overcome these roadblocks to create superior biocatalysts

Abstract

In addition to its role as the primary industrial ethanol producing microbe, the yeast *Saccharomyces cerevisiae* is powerful tool for genetics and as such benefits from a suite of systems biology tools, which can be leveraged to improve biofuel production. We have developed new reagent sets for understanding the roadblocks to efficient bioethanol production. Using yeast chemical genomics, we have developed a high-throughput, low-cost method of "biological fingerprinting" lignocellulosic hydrolysates that can be used to identify nutrient limitations as well as lignocellulosic derived inhibitors (LCDIs). Further, through chemical genomic analysis of a novel LCDI we have determined the mode-of-action of an uncharacterized compound that exerts toxicity by attacking the yeast cell wall, and has potential as a value-added bioproduct from cellulosic hydrolysates. Additionally, we have transformed industrially-relevant, xylose-fermenting yeast with a genome-wide, barcoded ORFeome collection (MoBY-ORF 2.0) to allow massively-parallel, gain-of-function studies and chemical genetic interactions. With this system we have identified genes that confer specific tolerance to anaerobic hydrolysate fermentation. These data will serve as the basis for rational engineering of yeast for bioconversion of next-generation biofuels.

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

213. Directed Engineering and Evolution for Xylose Fermentation and Lipid Biofuels from AFEX-Pretreated Corn Stover Hydrolysate by *Saccharomyces cerevisiae*

Trey K. Sato[1*] (tksato@glbrc.wisc.edu), Lucas Parreiras[1], Rebecca Breuer[1], Tam Tran-Nguyen[2], Rago Avanasi[1], Alex La Reau[1], Haibo Li[1], Edward Pohlmann[1], Alan Higbee[1,3], Allison Balloon[3], Joshua Coon[1,3], David Keating[1], Timothy Durrett[2], Yaoping Zhang[1], Audrey P. Gasch[1,4] and Robert Landick[1,5]

[1]DOE Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, WI; [2]Department of Biochemistry and Molecular Biophysics, Kansas State University, Manhattan, KS; [3]Department of Chemistry, University of Wisconsin, Madison, WI; [4]Department of Genetics, University of Wisconsin, Madison, WI; [5]Department of Biochemistry, University of Wisconsin, Madison, WI.

https://www.glbrc.org/

Project Goals

Despite the increased interest and efforts in cellulosic biofuels research, a number of molecular and biochemical barriers remain that prevent the efficient bioconversion of plant feedstocks into ethanol and lipid biofuels. These barriers for the industrial yeast biocatalyst, *Saccharomyces cerevisiae*, include metabolic limitations induced by cellular stress from chemical compounds generated by feedstock pretreatment, which in turn impact fermentation yield and productivity. The second major barrier is the absence of genes required for native *S. cerevisiae* to ferment xylose and produce lipid biofuels. At the DOE Great Lakes Bioenergy Research Center, we have attempted to better understand, address and overcome these barriers in *S. cerevisiae* through evolutionary, synthetic biology and metabolomic approaches.

Abstract

Previously, we screened a large panel of wild and domesticated *S. cerevisiae* strains grown in lignocellulosic hydrolysates generated from Ammonia Fiber Expansion (AFEX™) and alkaline hydrogen peroxide pretreatments. One wild strain, which we named GLBRCY0, was identified for its robust growth properties. Through engineering and experimental evolution, we developed modified derivatives that can ferment xylose into ethanol effectively from AFEX™ corn stover hydrolysate (ACSH). To understand how these strains are able to ferment xylose more rapidly, we performed metabolomic profiling of these strains during ACSH fermentation. This analysis suggested that elevated activities by enzymes in the Pentose Phosphate Pathway are important for this evolved phenotype. We additionally engineered a stress-tolerant, xylose- fermenting yeast strain with the *Euonymus alatus* Diacylglycerol Acetyltransferase (*EaDAcT*) to produce 3-acetyl-1,2-diacyl-*sn*-glycerol (acTAG), which is an oil with diesel like properties. This *S. cerevisiae* strain engineered with *EaDAcT* can produce acTAGs aerobically from both glucose and xylose, as well as from glucose in ACSH. Together, these results indicate that engineered *S. cerevisiae* can effectively produce acTAGs from renewable sugars generated from pretreated plant feedstocks, and our strains have significant promise as industrial biocatalysts for ethanol and lipid biofuels. This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).

214. Overcoming barriers for the conversion of lignocellulose to biofuels by *E. coli*

Eduardo H. Morales[1]* (emorales2@wisc.edu), Yaoping Zhang[1], David H. Keating[1], Irene M. Ong[1], Sean McIlwain[1], Jeffrey A. Grass[1], Donna Bates[1], Mary Tremaine[1], William Bothfeld[1], Alan Higbee[1], Arne Ulbrich[1], Allison Balloon[1], Emma E. Coughlin[1], Michael S. Westphall[1], Cameron Cotton[1], Leonardo da Costa Sousa[2], Mingjie Jin[2], Venkatesh Balan[2], Joonhoon Kim[1], Jennifer L. Reed[1], James Ellinger[1], Oleg Moskvin[1], Yury Bukhman[1], Joshua Coon[1], John Markley[1], Bruce Dale[2], **Patricia J. Kiley[1], Robert Landick[1].**

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

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

Project goals: Our goal is to understand and overcome barriers to development of an economically viable and sustainable conversion of lignocellulose to biofuels by bacteria. Currently, two major barriers are targeted. First, most microbes and chemical processes cannot simultaneously produce fuels from both C_6 (e.g., glucose) and C_5 (e.g., xylose) sugars, thereby limiting their ability to utilize rapidly all of the sugars in the hydrolysate. Second, most microbes are intolerant to stresses experienced during biofuel fermentation, including high ethanol or biofuel concentrations, toxic pretreatment byproducts, and high osmotic strengths experienced in the hydrolysate.

Multiomic and computational analyses define effects of (lignotoxins) LTs on metabolic and regulatory networks. Study of *E. coli* ethanologenesis in AFEX-treated corn stover hydrolysate (ACSH) coupled with analyses of its chemical composition allowed the development of a next generation synthetic medium (SynHv2) containing a cocktail of lignocellulose-derived inhibitors (referred to here as lignotoxins, LTs) that replicates the ACSH growth, sugar utilization, and gene expression profiles of an *E. coli* ethanologen (GLBRCE1). Major stress responses are caused by aromatic aldehyde, amide, and carboxylate LTs derived from lignocellulose. Gene expression patterns in the ethanologen were similar for ACSH and SynH containing the LT cocktail, indicating that the new SynH closely mimics the stresses caused by ACSH. Using this system as a model for ACSH, the effects of LTs on conversion and *E. coli* physiology were studied in detail.

Lignotoxins significantly perturbed metabolism, causing dramatic elevations of pyruvate and acetaldehyde with concomitant depletion of ATP, NADPH, and NADH. Transcriptomic, proteomic, and bioinformatic analyses revealed 5 major regulatory responses consistent with LT- induced energy stress in *E. coli*. One, mediated by elevated levels of RpoS, reflects a general stress response. Four others, mediated by YqhC, FrmR, AaeR, and MarA/Rob/SoxS, reflect specific responses to LTs. YqhC and FrmR, known regulators of aldehyde detoxification pathways, showed early induction that was ameliorated in stationary phase. Additionally, a decline in hydrolysate aromatic aldehyde levels paralleled by an accumulation of their alcohol forms in the extracellular medium was observed, consistent with detoxification of this class of LTs and the observed down regulation of *yqhC* and *frmR* in stationary phase. AaeR and MarR/MarA/Rob control efflux pumps and metabolic detoxification pathways. Both aromatic carboxylate-specific (AaeAB) and more general "multidrug resistance" pumps were activated; notably, these pumps consume ATP in each efflux cycle and their expression remained elevated throughout all stages of fermentation. Unlike the aromatic aldehydes, aromatic acid and amide LT levels remained constant throughout fermentation, suggesting that they were not

metabolized and might create an energetic challenge for cells that limits their ability to generate adequate reducing equivalents for conversion of pyruvate to ethanol, particularly when catabolizing C5 sugars (eg, xylose) where the energetics are less favorable. In agreement, directed evolution of *E. coli* in glucose-depleted ACSH resulted in a *rob* knockout mutation that improved xylose conversion, consistent with the hypothesis that Rob-mediated stress responses divert energy that can otherwise drive xylose to ethanol conversion.

Identification of strains capable of xylose utilization during growth in lignocellulose hydrolysates. It is known that xylose transport is subject to inducer exclusion, and that this is the predominant mechanism that prevents pentose sugar uptake when glucose is present. To test whether removing a component of the inducer exclusion pathway improved xylose utilization in the presence of glucose, we deleted the gene for the main glucose transporter *ptsG*. Although elimination of PtsG resulted in glucose/xylose co-utilization in SynHv2, xylose consumption was not complete, and only 50% the glucose was consumed. We also examined altered function mutants of Crr (or $EIIA^{GLC}$), a second component of the inducer exclusion pathway that is required for glucose phosphorylation, activation of adenylate cyclase, and inhibits transport of non-PTS sugars by directly binding the relevant transporters. We analyzed mutant variants that we predicted would separately inactivate these various functions and tested whether any of these variants improved xylose utilization. Of the Crr mutants tested, only substitution of His 90 to Ala showed slightly increased xylose utilization, suggesting that this residue plays a role in inhibiting xylose uptake in the presence of glucose, in addition to its expected role in transfer of phosphoryl groups to the glucose transporter PtsG. We are currently testing if mutating His 90 to aspartic acid or glutamic acid, which is expected to mimic the phosphorylated form of Crr and to abolish its interaction with secondary sugar transporters, results in glucose/xylose co-utilization.

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

215. *Streptomyces* isolated from herbivorous insects reveal phylogenetically-linked mechanisms for the degradation of lignocellulose

Adam Book[1,2]* (ajbook@wisc.edu), Gina Lewin,[1,2] Evelyn Wendt-Pienkowski,[1,2] Taichi Takasuka,[1,3] Bradon McDonald,[1,2] Jonathan Klassen,[1,2] Drew Doering,[1,2] Matt Wagar,[1,2] Brian Fox,[1,3] and **Cameron Currie**[1,2]

[1] DOE Great Lakes Bioenergy Research Center, Univ. of Wisconsin – Madison, WI 53706, USA
[2] Department of Bacteriology, University of Wisconsin-Madison, Madison, WI 53706, USA
[3] Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706, USA

http://currielab.wisc.edu

Project Goals: Identify plant biomass-degrading enzymes from highly cellulolytic *Streptomyces*.

Free-living Actinobacteria in the genus *Streptomyces* are well known for degrading plant biomass in the soil. More recently, *Streptomyces* have been found associated with insect herbivores. Despite *Streptomyces'* predicted contributions to carbon cycling in nature the extent of their biomass degrading ability and the enzymatic mechanism of this activity remain largely unknown. Here, we comparatively analyze strains across the *Streptomyces* genus including free-living soil isolates and strains from insect herbivores. While few soil-isolated *Streptomyces* grow on pure cellulose, two clades of insect-associated *Streptomyces* are enriched in cellulolytic activity. Biochemical analysis of secreted enzymes demonstrates significant differences in substrate specificity between cellulolytic strains. Furthermore, phylogenetic and genomic analyses identify key differences between in carbohydrate active enzyme composition between cellulolytic and non-cellulolytic strains. Genome-wide transcriptomic analyses reveal a highly conserved core set of enzymes used to deconstruct plant biomass as well as strain specific gene expression. Finally, we show that targeted-gene deletion of a global cellulose regulator results in constitutive expression of cellulolytic genes. Together, these data identify phylogenetically related clusters of *Streptomyces* with high biomass degrading activity and increase our understanding of how bacteria deconstruct lignocellulosic biomass in insect herbivore systems.

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

216. Highly Cellulolytic Microbial Communities Enriched from Leaf-Cutter Ant Refuse Dumps

Gina R. Lewin[1,2]* (glewin@wisc.edu), Amanda L. Johnson[1], Adam J. Book[1,2], Kailene S. Perry[1,2], and **Cameron R. Currie**[1,2]

[1]University of Wisconsin-Madison and [2]DOE Great Lakes Bioenergy Research Center, Madison, WI

http://currielab.wisc.edu/antcam

Project Goals: This project aims to understand how cellulose is degraded by a community of microbes to inform on soil carbon cycling and on mechanisms of biomass degradation for biofuel production.

Microbial degradation of cellulose is a key driver of the global carbon cycle, but little is known about the role of communities in this process. Here, we study the leaf-cutter ant refuse dump as a model system to explore biomass degradation by a community of microbes using culture- dependent and -independent techniques. Leaf-cutter ants are the dominant herbivore in New World tropical ecosystems. However, harvested leaves are not used by the ants as a primary food source, but rather as nutrients for a fungal cultivar that partially degrades the leaf material and provides energy to the colony. Non-digested leaf material, which is enriched in recalcitrant plant polymers such as cellulose and lignin, is moved to refuse dumps. Previous culture-independent work showed these nutrient-rich refuse dumps contain a diverse microbial community enriched in biomass-degrading enzymes. Here, we show that microbial communities in refuse dumps are highly capable at deconstructing cellulose; in ten days, the most highly cellulolytic communities degrade the same amount of cellulose as *Trichoderma reesei* RUT-C30, a fungus used in the biofuels industry. Our analysis of 16S pyrotag sequencing reveals that cellulose-enriched communities are dominated by Bacteroidetes and Proteobacteria. Additionally, specific operational taxonomic units (OTUs), including *Diaphrobacter* and *Chryseobacterium*, are significantly more abundant in highly cellulolytic communities than in non-cellulolytic communities, while Microbacteriaceae are significantly less abundant. However, there were few OTUs present that are known to degrade cellulose, even in highly cellulolytic samples, suggesting that cellulose degradation may be performed by previously unappreciated cellulolytic organisms or microbial consortia. We expanded on this work with long-term enrichments of highly cellulolytic communities. Characterization of these communities using metagenomic and metatranscriptomic analyses and culture-based experiments is currently underway. The results of this research are important for understanding community-based cellulose degradation in a biomass-enriched environment and for providing insights to improve cellulosic biofuel production.

This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494), UW-Madison CALS Hilldale Fellowship, and an NSF-GRPF (DGE-1256259).

217. Computational Algorithms for Metabolic Engineering Strain Design

Joonhoon Kim[1,2], Cameron Cotten[1,2] , Xiaolin Zhang[3], Mary Tremaine[2], David Keating[2], Robert C. Landick[2,4], Patricia J. Kiley[2,5], and Jennifer L. Reed[1,2]* (reed@engr.wisc.edu)

[1]Dept. of Chemical and Biological Engineering, Univ. of Wisconsin-Madison
[2]Great Lakes Bioenergy Research Center, Univ. of Wisconsin-Madison
[3]Cellular and Molecular Biology Graduate Program, Univ. of Wisconsin-Madison
[4]Dept. of Bacteriology, Univ. of Wisconsin-Madison
[5]Dept. of Biomolecular Chemistry, Univ. of Wisconsin-Madison

https://www.glbrc.org/

Project Goals: Computational models of biological systems can be used to explain observed behaviors, predict un-measurable quantities, and predict cellular behavior arising from environmental, genetic perturbations, or both. Models can be useful in engineering biofuel production strains, where challenges include finding bottlenecks in metabolic pathways and identifying appropriate perturbations to force a microorganism to produce or consume more of a compound of interest. Our research efforts have focused on developing and improving computational tools for designing strains, and applying these tools to models of *Escherichia coli* and *Saccharomyces cerevisiae*.

Metabolic engineering seeks to improve cellular production of valuable biochemicals, such as biofuels, by altering metabolic and regulatory pathways in a host strain. Computational tools are becoming increasingly available to design microbial strains using genome-scale *in silico* models of metabolism that predict the re-distribution of metabolic fluxes after genetic or environmental perturbations. In this poster, we describe some of our recent results involving computational and experimental approaches to improve biofuel production.

First, we developed and applied a new computational tool for identifying fluxes that must change to increase biofuel production or increase sugar utilization. The method determines how much to increase or decrease fluxes to improve production. This tool, which we named CosMos (for continuous modifications of fluxes), was able to determine strategies for up- and down- regulation of metabolic genes in *E. coli* and *S. cerevisiae* that lead to production of a variety of desired chemicals, including biofuels and amino acids. We have compared our results from CosMos to previous results in the experimental metabolic engineering literature, and we found a number of cases wherein our predictions with CosMos closely matched successful experimental strategies. We are currently using this approach to design *E. coli* and yeast strains for advanced biofuel production.

Second, we used our developed algorithm (OptORF) to identify gene deletions needed to improve pyruvate production in *E. coli*. OptORF identifies which metabolic and regulatory genes to delete or over express so that chemical production is coupled to cellular growth [1]. We engineered a number of strains and have achieved yields of more than 0.88 g pyruvate per g of glucose (~90% theoretical yield). To re-engineer the pyruvate strains to produce ethanol, pyruvate formate-lyase (PflB) was deleted and pyruvate decarboxylase (Pdc) and alcohol dehydrogenase II (AdhB) from *Zymomonas mobilis* were highly expressed. These re-engineered strains fermented glucose to ethanol with a yield of 0.35 g ethanol per g of glucose (~70% of theoretical yield). While higher ethanol yields have been

achieved, these results illustrate that pyruvate over-producing strains can serve as a platform to generate other biofuels derived from pyruvate.

Third, we used our previously developed algorithms (OptORF and RELATCH [2]) to identify gene deletions needed to improve xylose fermentation in *E. coli* when glucose is present. Knockout mutant strains were constructed and adaptively evolved under anaerobic conditions. Some of the resulting strains were able to co-utilize glucose and xylose when grown anaerobically in minimal medium and were able to consume xylose at an increased rate (albeit at the expense of glucose consumption). Together these three studies illustrate how computational tools can be used to facilitate the design of strains for converting lignocellulosic biomass into biofuels.

References

1. Kim, J. and J.L. Reed, *OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains.* BMC Syst Biol, 2010. **4**: p. 53.
2. Kim, J. and J.L. Reed, *RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations.* Genome Biol, 2012. **13**(9): p. R78.

This work was funded by the U.S. Department of Energy Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494) and in part by the W.M. Keck Foundation.

218. Using Chemical Pretreatment to Understand Bacterial Biomass Deconstruction

Stephanie A. Eichorst[1,2], Chijioke Joshua[1,2], Noppdon Sathitsuksanoh[1,2], Yu-Wei Wu[1,2], Seema Singh[1,3], Blake A. Simmons[1, 3], and **Steven W. Singer**[1,4]

[1]Joint BioEnergy Institute, Emeryville, California 94608; [2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [3]Biological and Materials Sciences Center, Sandia National Laboratories, Livermore, California, 94551; [4]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, 94720

http://www.jbei.org/research/divisions/deconstruction/microbial-communities/

Project Goals: Natural microbial communities are rich sources of microbes, enzymes and transporters to improve the conversion of biomass to biofuels. The Microbial Communities group at the Joint BioEnergy Institute discovers these activities and translates them for applications in biofuel production.

Abstract: Microbial communities that deconstruct plant biomass have broad relevance in biofuel production and global carbon cycling. Understanding the response of these communities to variations in biomass composition is confounded by the dual complexities of natural microbial communities and diverse plant cell wall structures. Chemical pretreatment of biomass, used to reduce plant biomass recalcitrance for enzymatic hydrolysis, provides a predictable method to vary the structure and composition of plant cell walls. This property of chemical pretreatment has been exploited to assess how thermophilic bacterial consortia adapt to deconstruct biomass of varying compositions, through parallel cultivations of compost-derived consortia on intact and pretreated switchgrass (ammonia-fiber expansion and ionic liquid pretreatment). This adaptation selected for three dominant taxa that were representatives of the *Firmicutes*, *Bacteroidetes*, and *Deinoccocus/Thermus* phyla and demonstrated that community complexity was correlated with the chemical and physical structure of the biomass. Gravimetric analysis of the residual biomass demonstrated that pretreatment enhanced the digestion of biomass and 2D-NMR indicated that polysaccharide hydrolysis was the dominant process occurring during microbial biomass deconstruction and lignin was largely unaltered. Metagenomic sequencing of a switchgrass-adapted consortium enabled reconstructions of novel thermophilic *Chitinophaga* and *Truepera* genomes that revealed broad capabilities to deconstruct plant polysaccharides.

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.

219. Developing New Fatty Acid-Derived Biofuels at JBEI: Methyl Ketones and Ladderanes

Pouya Javidpour[1,*] (PJavidpour@lbl.gov), Ee-Been Goh[1], Edward Baidoo[1], Vivek Mutalik[1], Sam Deutsch[2], Jay D. Keasling[1], and **Harry R. Beller**[1]

[1]Joint BioEnergy Institute (JBEI), Emeryville, CA; [2]Joint Genome Institute (JGI), Walnut Creek, California

Project Goals: The Joint BioEnergy Institute (JBEI) aims to produce a chemically diverse suite of biofuels from lignocellulosic biomass. Some JBEI fuels use fatty acids as precursors, as these biomolecules are highly reduced, aliphatic compounds that, when modified (e.g., decarboxylated), can have properties comparable to those of petroleum-derived fuel components. The goals of the two projects presented here are to: (1) engineer *E. coli* to produce diesel-range methyl ketones in the gram-per-liter range with yields of at least 40% of maximum theoretical yield and (2) elucidate the biosynthetic pathway for unique and highly energetic ladderane fatty acids, which contain linearly concatenated cyclobutane rings, are synthesized by anaerobic ammonium-oxidizing (anammox) bacteria, and have promise as a new class of fatty acid-derived biofuels.

Methyl Ketones

We have engineered *Escherichia coli* to overproduce saturated and monounsaturated aliphatic methyl ketones in the C_{11} to C_{15} (diesel) range; this group of methyl ketones includes 2- undecanone and 2-tridecanone, which have favorable cetane numbers and are also of importance to the flavor and fragrance industry. We have made specific improvements that resulted in more than 10,000-fold enhancement in methyl ketone titer relative to that of a fatty acid-overproducing *coli* strain, including the following: (a) overproduction of beta-ketoacyl-coenzyme A (CoA) thioesters achieved by modification of the beta-oxidation pathway (specifically, overexpression of a heterologous acyl-CoA oxidase and native FadB, and chromosomal deletion of *fadA*) and (b) overexpression of a native thioesterase (FadM). The first generation of engineered *E. coli* (Goh et al. 2012) produced ~380 mg/L of methyl ketones in rich medium. We have subsequently made additional genetic modifications, including consolidation of all pathway genes onto a single plasmid (they were originally borne on two plasmids). We have also conducted *in vitro* assays with purified pathway enzymes, which revealed that FadM is promiscuous and has thioesterase activity not only toward beta-ketoacyl-CoAs but also toward other CoA thioester intermediates in the pathway (such as *trans*-2-enoyl-CoAs). These *in vitro* results have provided insight on how to fine-tune expression of pathway genes for further optimization of methyl ketone production. Our current methyl ketone titer is >1.5 g/L with 1% glucose, which represents 40-45% of maximum theoretical yield; these are the highest titer and yield values reported to date for methyl ketones.

Ladderanes

Ladderanes (or their derivatives) hold potential as a novel next-generation biofuel and for material science applications. These highly energetic molecules, with a carbon backbone consisting of linearly concatenated cyclobutane rings attached to an acyl chain, are incorporated into organelle (anammoxosome) membranes by anammox bacteria. To date, metagenomic data analysis (Rattray et al. 2009) identified 34 genes from an anammox bacterium, *Kuenenia stuttgartiensis*, hypothesized to be candidates involved in ladderane biosynthesis. However, the exact pathway and role of each gene product are unknown. We have developed a strategy to test these candidate ladderane biosynthesis genes in *E. coli* because *K. stuttgartiensis* is not amenable to biochemical and genetic studies, as it has a two-week doubling time and is not available as an isolate. The genes were synthesized and divided into non-native operonic groups, each gene with a unique inducible promoter, translation initiation element, and

terminator, based on putative function as well as reported expression levels from metatranscriptome data (Kartal et al. 2011).

Strategies and tools for DNA assembly, gene expression, and fatty acid (conventional and ladderane) analysis using GC/MS were developed and, to date, 17 genes have been simultaneously expressed in *E. coli* under anaerobic conditions. Based on GC/MS data, the strain expressing the potential ladderane pathway genes exhibited higher ratios of monounsaturated to saturated C_{14}, C_{16}, and C_{18} fatty acids compared to the control strain. Additionally, compared to the control, the experimental strain produced higher levels of C_{16} and C_{18} cyclopropane fatty acids relative to palmitic acid (*n*-C_{16}). Incorporation and expression of the remaining putative pathway genes is ongoing. This work will lay the foundation for a novel advanced biofuel pathway.

References

1. Goh EB, Baidoo EE, Keasling JD, Beller HR. 2012. Engineering of bacterial methyl ketone synthesis for biofuels. *Appl. Environ. Microbiol.* **78:**70-80.
2. Rattray JE, Strous M, Op den Camp HJ, et al. 2009. A comparative genomics study of genetic products potentially encoding ladderane lipid biosynthesis. *Biol. Direct* **4:**8.
3. Kartal B, Maalcke WJ, de Almeida NM, et al. 2011. Molecular mechanism of anaerobic ammonium oxidation. *Nature* **479:**127-130.

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.

220. High-Throughput Nanostructure-Initiator Mass Spectrometry (HT-NIMS) Metabolite Screening at JBEI

Xiaoliang Cheng[1,2], Kai Deng[1,3], Ee-Been Goh[1,2], Satoshi Yuzawa[1,2,6], Tristan de Rond[1,6], Richard Heins[1,3], Samuel Deutsch[2,5], Brian G. Fox[4], Blake A. Simmons[1,3], Anup K. Singh[1,3], **Paul D. Adams[1,2,6], Jay Keasling[1,2,6], Trent R. Northen[1,2*]** (TRNorthen@lbl.gov)

[1]Joint BioEnergy Institute, Emeryville, CA; [2]Lawrence Berkeley National Laboratory, Berkeley, CA; [3]Sandia National Laboratories, Livermore, CA; [4]GLBRC, University of Wisconsin, Madison, WI; [5]Joint Genome Institute, Walnut Creek, CA; [6]University of California, Berkeley, CA

http://www.jbei.org

Project Goals: The Joint BioEnergy Institute (JBEI) is a United States Department of Energy Bioenergy Research Center dedicated to developing advanced biofuels—liquid fuels derived from the solar energy stored in plant biomass that can replace gasoline, diesel and jet fuels. JBEI's approach to biomass derived biofuels uses a combination of novel feedstock engineering, ionic liquid biomass pretreatment and high performance enzymes to effectively produce low-cost fermentable sugars. In parallel, synthetic biology approaches are being pioneered to enable conversion of these sugars into high-performance drop-in biofuels. In this case, the disconnect between the rate of clone production vs. our ability to measure biofuel yield and composition presents a major technical challenge. Typically, this requires either non-specific assays or radical down selection prior to chemically specific analysis using mass spectrometry. In this project we are pioneering HT-NIMS as a tool to rapidly screen metabolite composition in support of fuels engineering efforts at JBEI.

We have recently developed high throughput enzyme activity analysis capabilities using the novel integration of nanoliter-scale acoustic sample deposition with nanostructure-initiator mass spectrometry (NIMS) analysis platform to rapidly detect and characterize screen large libraries of glucosidases. Central to this approach has been the use of bioconjugate chemistries enabling modification of glycans for high sensitivity and quantitative from complex mixtures. This HT- NIMS technology enables rapid screening of enzyme libraries across relevant process conditions to support the rapid develop enzyme cocktails for a given plant biomass.

Here we describe recent efforts to extend this platform to synthetic biology based fuel engineering efforts at JBEI. In some cases, relevant molecules (e.g. fatty acids) can be detected without modification enabling rapid screening of lipid composition. In other cases, reactive chemical probes are used to improve the mass spectrometry characteristics of target molecules for NIMS analysis. Probes can be added using oxime chemistry and are found to effectively label ketone containing molecules enabling screening of both fatty acid methyl ketones and polyketides. Fatty acid methyl ketones can have high cetane numbers making them highly desirable as diesel fuels. Reduced polyketides provide a largely unexplored approach to producing biofuels and has great potential for producing renewable gasoline and jet fuels.
Overall, HT-NIMS has tremendous potential to enable direct screening that will help optimize the yield and composition for both yield and composition.

The DOE Joint BioEnergy Institute and DOE Great Lakes Bioenergy Research Center are supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 and through contract DE-FC02-07ER64494, respectively. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's Nuclear Security Administration under contract DE-AC04-94AL85000

221. Production of anteiso-branched fatty acids in *Escherichia coli*, next generation biofuels with improved cold-flow properties

Robert W. Haushalter*[1], Woncheol Kim[1], Megan E. Garber[2], Ted A. Chavkin[2], Leonard Katz[2], and **Jay D. Keasling**[1,2].

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, 5885 Hollis Street, 4th Floor, Emeryville, CA 94608; [2]QB3 Institute, University of California-Berkeley, 5885 Hollis Street, 4th Floor, Emeryville, CA 94608.

E-mail: jdkeasling@lbl.gov

Abstract
A major disadvantage of fuels derived from biological sources is their undesirable physical properties such as high cloud and pour points and high viscosity. Here we report the development of an *Escherichia coli* strain that efficiently produces anteiso-branched fatty acids, which can be converted into downstream products with lower cloud and pour points compared to less heterogeneous mixtures produced via the native metabolism of the cell. This was achieved through the deletion of *metA, tdh, ilvB*, and *ilvN* and the overexpression of *thrABC* from *E. coli*, *ilvA* from *Corynebacterium glutamicum*, *ilvGMCD* from *Salmonella typhimurium*, as well as *bFabH2* and the *bkd* operon from *Bacillus subtilis*, which together promote the synthesis of the 2-methylbutyryl-CoA and use this metabolite to prime fatty acid synthesis. When these genetic manipulations are coupled with those that promote free fatty acid synthesis and accumulation, 22% of the free fatty acids produced in the engineered *E.coli* cells were anteiso-branched. This work addresses a serious limitation that must be overcome in order to produce renewable biodiesel and oleochemicals that perform as well as their petroleum-based counterparts.

URL: http://www.jbei.org/

222. Novel Mechanisms for Ionic Liquid Resistance in Microbial Biofuel Production

Thomas Rüegg,[1,2,6] Douglas Higgins,[1,6] Eun-Mi Kim,[1,3] Blake Simmons,[1,4] Jay Keasling,[1,3,5] Steven Singer,[1,3] Taek Lee,[1,3] and **Michael Thelen**[1,6]* (mthelen@llnl.gov)

[1]Joint BioEnergy Institute, Emeryville, California; [2]University of Basel, Switzerland; [3]Lawrence Berkeley National Laboratory, Berkeley, California; [4]Sandia National Laboratories, Livermore, California; [5]University of California, Berkeley, California; [6]Lawrence Livermore National Laboratory, Livermore, California

http://www.jbei.org/research/divisions/deconstruction/microbial-communities

Project Goals: Natural microbial communities are rich sources of microbes, enzymes and transporters to improve the conversion of biomass to biofuels. The Microbial Communities group at the Joint BioEnergy Institute discovers these activities and translates them for applications in biofuel production.

Microorganisms have developed innovative ways to stay alive in extreme conditions, and many of these adaptive mechanisms operate to their advantage in anthropogenic milieux. Many forest soil microbes, for example, secrete highly active enzymes and can withstand fluctuations in temperature, pH, ionic strength, and hydration. Thus, using environmental bacteria could improve biomass degradation and reduce microbial growth inhibition from toxic byproducts. An alternative involves harnessing the positive traits of these native microbes to engineer existing laboratory strains for increased efficiency in cellulosic biofuel production.

We isolated a novel halotolerant bacterium, *Enterobacter lignolyticus*, from a tropical cloud forest soil community[1]. *E. lignolyticus* digests lignin[2] and moreover grows in the presence of relatively high concentrations of an ionic liquid used in pretreating biomass[3]. This chemical, 1-ethyl-3-methylimidazolium chloride, or [C_2mim]Cl, is an effective cellulose solvent but the imidazolium cation is especially toxic to biofuel producing microbes.

We examined changes in the global transcriptome of *E. lignolyticus* after exposure to [C_2mim]Cl, using RNA-seq transcriptomics to identify differentially expressed genes and the associated pathways[3]. In a parallel study we screened a library of *E. lignolyticus* genome fragments, and found an IL resistance system consisting of two adjacent genes[4]. One of these corresponded to the highest IL-induced transcript in *E. lignolyticus*. The two genes retain their full functionality when transferred to an *E. coli* biofuel host, with IL resistance established by an inner membrane transporter that is regulated by an IL inducible repressor. Expression of the transporter, a MFS multidrug efflux pump, is dynamically adjusted by a TetR-type regulator in direct response to IL concentration, enabling growth and biofuel production at levels of IL that are toxic to native strains (**Fig. 1**).

To further understand and overcome the growth inhibition caused by pre-treatment ionic liquids, we have isolated several other environmental microbes – Gram-negative and Gram-positive bacteria and fungi – that tolerate [C_2mim]Cl and [C_2mim][OAc], a more potent cellulose solvent. We identified other efflux pumps in both the MFS and SMR families that are effective in providing ionic liquid resistance (**Fig. 2**). The SMR pump genes occur in tandem, and some are adjacent to transcriptional regulators that may adjust pump levels in response to changing IL concentrations. Further studies are

aimed at fine-tuning these pumps and their regulators for use in biofuel-producing *E. coli* and *S. cerevisiae*. The discovery of IL-tolerant microbes and their resistance mechanisms will overcome a major hurdle in the use of these cellulose solvents, and will advance our progress towards readily and economically converting lignocellulose biomass to biofuels and renewable chemicals.

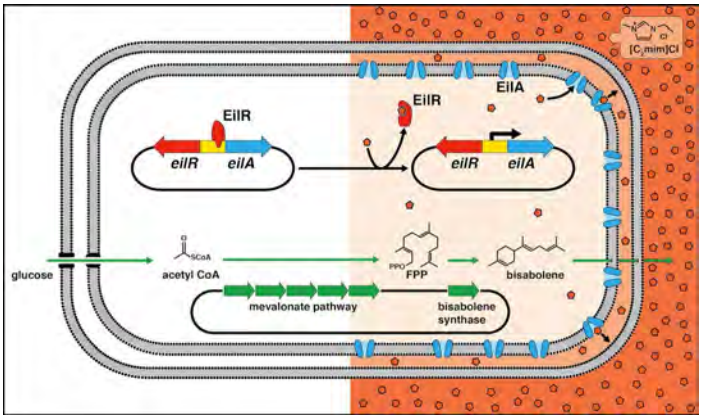
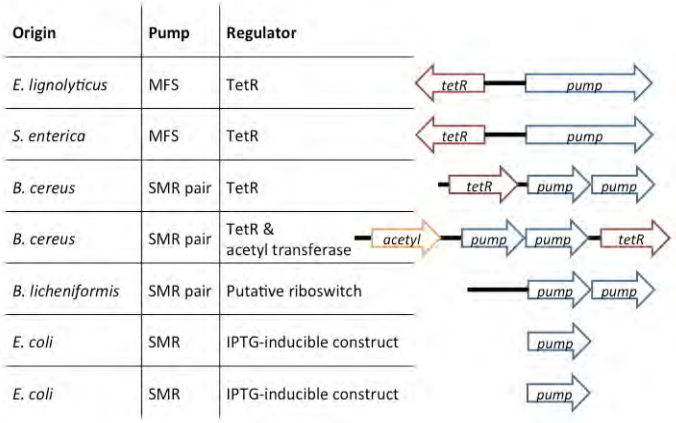


Figure 1 | Model of the IL tolerance mechanism engineered in a bacterium that converts hydrolyzed biomass into biofuel under two different conditions, either in the absence (left half) or presence (right half) of ionic liquid. *key*: bisabolene biosynthetic genes (green) and biofuel production pathway (green arrows), efflux pump EilA (blue), repressor EilR (red), intergenic region (yellow), imidazolium cation (orange pentagons).

Figure 2 | Library of resistance genes originating from several bacteria, including Salmonella and Bacillus strains, identified by screening genomic fragments in E. coli for growth in medium containing IL. Membrane transporters are either Multi- facilitator Superfamily (MFS) or Small Multidrug Resistance (SMR) efflux pumps, and some are associated with TetR-type regulatory genes. Although E. coli does not contain an MFS ortholog, two SMR homologs are present but only confer IL tolerance when expression is induced.



[1] KM Deangelis et al. (2011) Genome sequence of "*Enterobacter lignolyticus*" SCF1. *Stand Genomic Sci* 5, 69. http://dx.doi.org/10.4056%2Fsigs.2104875

[2] KM Deangelis et al. (2013) Evidence supporting dissimilatory and assimilatory lignin degradation in *Enterobacter lignolyticus* SCF1. *Frontiers Microbiol* 4, 280. http://dx.doi.org/10.3389%2Ffmicb.2013.00280

[3] JI Khudyakov et al. (2012) Global transcriptome response to ionic liquid by a tropical rain forest soil bacterium, *Enterobacter lignolyticus*. *Proc Natl Acad Sci USA* 109, E2173. http://dx.doi.org/10.1073/pnas.1112750109

[4] TL Rüegg et al. (2014) An inducible mechanism for ionic liquid resistance in microbial biofuel production. *Nature Communications* (accepted)

This work was performed at the Joint BioEnergy Institute with funding from the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 with Lawrence Berkeley National Laboratory.

223. High-throughput cellulase activity profiling on ionic liquid pretreated biomass

Taya Feldman[1,3]* (tfeldman@lbl.gov), Joel Guenther[2,3], Huu Tran[2,3], Anup Singh[3], Paul Adams[2,4], Blake Simmons[1,3], **Ken Sale[1,3]**

[1]Deconstruction and [2]Technology Divisions, Joint BioEnergy Institute, Emeryville, CA; [3]Sandia National Laboratory, Livermore, CA; [4]Lawrence Berkeley National Laboratory, Berkeley, CA.

Project Goals: To develop a database of lignocellulolytic enzymes whose activities have been characterized across a matrix of conditions (T, pH, ionic liquid concentration) to identify the enzyme features that govern activity.

Abstract

Technoeconomic modeling at JBEI shows that an industrial cellulosic biofuel process based on the ionic liquid 1-ethyl-3-methylimidazolium acetate can be cost-competitive with fossil fuels provided that cellulases can be found with activity at ≥70 °C in ≥20% ionic liquid[1]. To identify cellulases compatible with ionic liquids, our high-throughput pipeline must rapidly and thoroughly profile candidates. Discovery of cellulases with activity in the presence of ionic liquids has been hampered by labor-intensive, low-throughput assays and poor model substrates. Using standard liquid handling robots, we have developed jSALSA – an automated, high- throughput process for precisely dispensing sub-milligram quantities of ionic liquid pretreated biomass. With this solid substrate in microplate format, we are thoroughly profiling cellulase activities as a function of temperature, pH and ionic liquid concentration. This dataset will allow us to (1) understand the response of enzyme activity to ionic liquid pretreatment conditions and downstream saccharfication requirements (2) explore correlations between protein sequence, thermotolerance and ionic liquid tolerance and (3) select the best matched set of candidate enzymes for optimization of multi-component cellulase mixtures.

References

1. Klein Marcuschamer D, Simmons BA, Blanch HW (2011) Techno-economic analysis of a lignocellulosic ethanol biorefinery with ionic liquid pre-treatment. Biofuels, Bioproducts and Biorefining; 5: 562–569.

224. Identification of six Golgi localized bi-functional UDP-Rhamnose/UDP-Galactose transporters in Arabidopsis

Carsten Rautengarten[1], Berit Ebert[1,2], Ignacio Moreno[3], Henry Temple[3,] Thomas Herter[1], Bruce Link[4], Jennifer C. Mortimer[5], Alex Schultink[6], Wolf-Dieter Reiter[4], Paul Dupree[5], Markus Pauly[6], Ariel Orellana[3], Joshua L. Heazlewood[1], and **Henrik Vibe Scheller**[1,6]* (hscheller@lbl.gov)

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California; [2]Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Denmark; [3]Centro de Biotecnología Vegetal, Universidad Andrés Bello, Santiago, Chile; [4]Department of Molecular and Cell Biology, University of Connecticut; [5]Department of Biochemistry, University of Cambridge, UK; [6]Department of Plant and Microbial Biology, University of California, Berkeley.

http://www.jbei.org

Project Goals: To determine the mechanisms of cell wall biosynthesis in plants. A better understanding of the metabolic processes responsible for accumulation of polysaccharides in the cell wall will provide tools for engineering of plants with improved properties as biofuel feedstocks.

Plant cell walls are composed of polysaccharides most of which are synthesized in the Golgi apparatus from nucleotide sugars that are transferred from the cytosol into the Golgi lumen by nucleotide sugar transporters (NSTs). By using a novel approach that combines reconstitution of NSTs into liposomes and subsequent LC-MS/MS analysis of nucleotide sugar uptake, we identified six bi-functional UDP-rhamnose / UDP-galactose transporters. All six transporters are localized in the Golgi. Mutants in URGT1 have reduced galactose in the cell wall whereas overexpressors accumulate up to 50% more galactose than wild-type plants, mainly in the pectin- rich fraction. In contrast, galactose in xyloglucan is unaffected in mutants and overexpressors. Mutants in URGT2 exhibit lower levels of rhamnose and galacturonic acid in seed mucilage, a structure rich in rhamnogalacturonan I while URGT2 overexpressors have no change in leaf cell wall composition. Our results suggest that some NSTs may be channeling nucleotide sugars to specific glycosyltransferases.

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

225. GUX and IPUT Members of Arabidopsis Glycosyltransferase Family 8 are Glucuronosyltransferases Involved in Cell Wall and Glycosphingolipid Synthesis

Emilie A. Rennie,[1,2,3,*] (earennie@lbl.gov), Paul Hussey,[1,2,3] Edward E.K. Baidoo,[1,2] Berit Ebert,[1,2] Katy M. Christiansen,[1,2] Solomon Stonebloom,[1,2] Christopher J. Petzold,[1,2] Paul D. Adams,[1,2,4] Jay D. Keasling,[1,2,4,5] Joshua L. Heazlewood,[1,2] and **Henrik V. Scheller**[1,2,3]

[1]Joint BioEnergy Institute, Emeryville, California
[2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California
[3]Department of Plant & Microbial Biology, University of California, Berkeley
[4]Department of Bioengineering, University of California, Berkeley
[5]Department of Chemical Engineering, University of California, Berkeley

http://www.jbei.org/

Project Goals: The Cell Wall Biosynthesis group at the Joint BioEnergy Institute is focused on developing a better understanding of plant cell wall biosynthesis and the role of cell wall polymers in plant growth and development. One of our specific aims is to identify and study the structures and functions of glycosyltransferases and other enzymes involved in cell wall synthesis. Our eventual goal is to use this knowledge to modify the composition of plant cell walls for better plant growth and biofuel yields.

Plant cell walls are a major sink for fixed carbon and a potentially important source of feedstocks for renewable energy. Cellulose microfibrils, the main load-bearing component of cell walls, are cross-linked by highly branched polysaccharides called hemicelluloses. In biofuel crops such as switchgrass, Miscanthus, and poplar, the major hemicellulosic polymer is xylan (Figure 1).

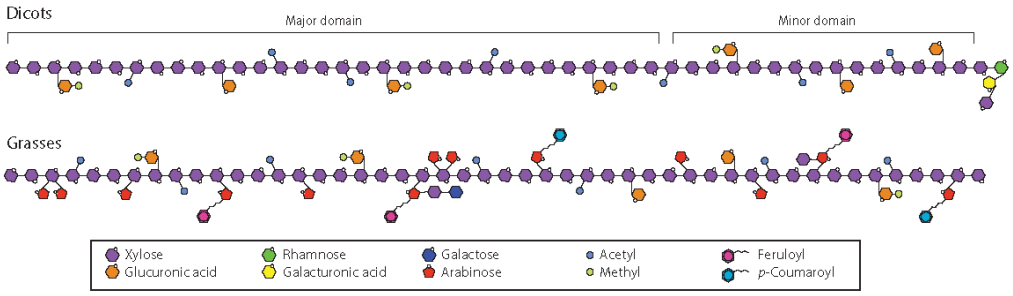


Figure 1. Structures of xylan in dicots and grasses. From (1)

We have previously shown that three members of Arabidopsis Glycosyltransferase Family 8, named Glucuronic Acid Substitution of Xylan (GUX)1, GUX2, and GUX4, are glucuronosyltransferases responsible for synthesizing glucuronic acid side chains on xylan.[2] Here, we characterized the substitution patterns with which GUX proteins transfer glucuronic acid to xylan. This analysis showed that *in vitro*, GUX1 transfers glucuronic acid to alternating xylose residues while GUX2 and GUX4 do not distinguish between odd or evenly spaced residues. These preferences closely match proposed roles for GUX1 and GUX2 *in planta*, where they are thought to synthesize distinct glucuronoxylan domains with different substitution patterns.[3] These results indicate that the GUX proteins are capable of recognizing subtle differences in xylan backbone structure, leading to related but separate enzymatic functions. Further

characterization of these proteins will help us understand how GUX proteins alter specific domains of the xylan backbone and how these alterations affect xylan functions in plants. In addition, we characterized a closely related protein that we have named Inositol Phosphorylceramide Glucuronosyltransferase 1 (IPUT1). Surprisingly, we found that this protein does not synthesize cell wall polymers but instead glycosylates an abundant class of sphingolipids called glycosyl inositol phosphorylceramides (GIPCs). We used a synthetic biology approach to reconstruct the plant GIPC synthesis pathway in yeast, allowing us to demonstrate the function of IPUT1. The identification of this new enzymatic activity sheds light on the synthesis of an important group of plant lipids. In addition, it will help us understand how the GUX/IPUT clade of proteins is able to recognize and differentiate between specific acceptor substrates such as glycolipids and xylan polymers.

References

1. **Rennie, E.A. and Scheller, H.V.** (2014). Xylan Biosynthesis. *Curr. Opin. Biotechnol.* **26:** 100-107.
2. **Rennie, E.A., Hansen, S.F., Baidoo, E.E.K., Hadi, M.Z., Keasling, J.D., and Scheller, H.V.** (2012). Three Members of the Arabidopsis Glycosyltransferase Family 8 Are Xylan Glucuronosyltransferases. *Plant Physiol.* **159:** 1408-1417.
3. **Bromley JR, Busse-Wicher M, Tryfona T, Mortimer JC, Zhang Z, Brown DM, Dupree P** (2013). GUX1 and GUX2 Glucuronyltransferases Decorate Distinct Domains of Glucuronoxylan with Different Substitution Patterns. *Plant J.* **74:** 423-434.

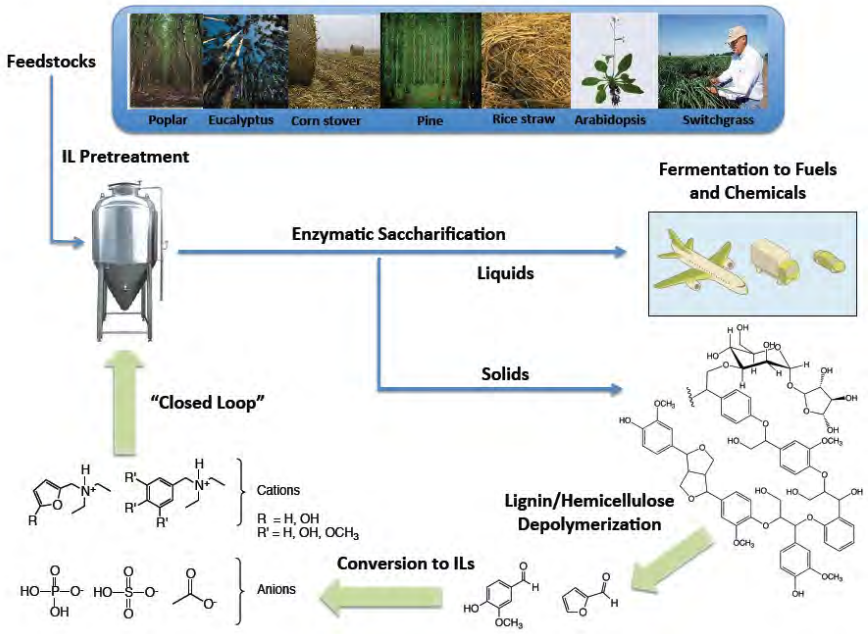
This work was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Lab and the U.S. Department of Energy.

226. Efficient Biomass Pretreatment using renewable Ionic Liquids derived from Lignin and Hemicellulose

Seema Singh[1,4*], Aaron M. Socha[1,2], Ramakrishnan Parthasarathi[1,4], Jian Shi[1,4], Dorian Whyte[1,2], Maxime Bergeron[1], Sivakumar Pattathil[3], Sivasankari Venkatachalam[3], **Blake A. Simmons**[1,4]

Presenter email: ssingh@lbl.gov, seesing@sandia.gov

Abstract: Ionic liquids (ILs), solvents composed entirely of paired ions, have been utilized in a wide variety of applications. For Biomass pretreatment, imidazolium based ILs have shown remarkable potential as a solvent and reaction medium. Although very efficient, imidazolium cations are currently derived from petroleum, making this IL expensive and limiting its larger scale use and industrial deployment. In an attempt to replace imidazlium based ILs with ILs derived from renewable sources that retain their efficiency for biomass pretreatment, we synthesized a series of novel ILs from monomers derived from lignin and hemicellulose, the major byproducts of lignocellulosic biofuel production. Molecular modeling allowed for the selection of three ILs for biomass pretreatment studies. A comprehensive glycome profile of extractable cell wall carbohydrates and sugar yields from raw switchgrass and switchgrass pretreated with ILs derived from vanillin, *p*-anisaldehyde and furfural confirmed new ILs effectiveness for biomass pretreatment and comparable to those observed after pretreatment with 1-ethyl-3-methylimidazolium acetate. Our concept of deriving ILs from lignocellulosic biomass show potential of a closed loop process for future lignocellulosic biorefineries and has far- reaching economic impact for other IL based processes currently using ILs synthesized from petroleum sources.



Hypothetical process flow for a "closed-loop" bio-refinery using efficient renewable ionic liquids derived from lignocellulosic biomass.

227. High-throughput Microfluidic Platforms for Bioenergy Research

Anup K. Singh*[1,4] (aksingh@sandia.gov), Jess Sustarich[1,4], Chao Shih[1,4], Philip Gach[1,4], Joel Guenther[1,4], Randy Drevlin[2,4], Ken Sale[2,4], Nathan Hillson[3,5], Blake Simmons[2,4], Paul Adams[1,5], **Jay Keasling**[3,5]

[1]Technology Division, [2]Deconstruction Division, [3]Fuels Synthesis Division, Joint BioEnergy Institute, Emeryville, CA; [4]Sandia National Laboratory, Livermore, CA; [5]Lawrence Berkeley National Laboratory, Berkeley, CA

http://www.jbei.org

Project Goals: We are developing next-generation assays for biofuels research using microfluidic technologies to provide significant improvement over conventional platforms in throughput, sensitivity, multiplexing and speed of analysis. These platforms are useful for numerous aspects of biofuels R&D including biomass deconstruction, feedstock development, and fuel synthesis.

Microfluidic platforms are finding widespread applications in biochemical analysis relevant to bioenergy research. Examples that we are exploring include screening of genetically-engineered cellulases and glycosyltransferases, assessing performance of pretreatment processes, optimization of enzyme cocktails for hydrolysis of biomass, and combinatorial screening of gene variants for optimization of metabolic pathways. For activity screening of glycosyl hydrolases and transferases, we have developed a microfluidic electrophoretic glycan analysis protocol that can be performed in a commercial microfluidic instrument. The chip allows 10-fold faster analysis than HPLC using 100-fold smaller amounts of reagents. Recently, we have been developing a droplet microfluidic platform for carrying out hundreds of reactions in parallel for two applications- a) screening of enzyme cocktails and b) combinatorial assembly of genes. Performing multi-step biochemical assays require the ability to perform functions such as droplet merging for addition of reagents and droplet sorting for selective isolation of desired reactions. We have developed innovative schemes to reproducibly merge, sort and array droplets. The droplet chip was used to screen combinations of cellulases with real insoluble substrates and the results show that the chip-based screening is an excellent agreement with conventional screening methods while offering advantages of throughput, speed and lower reagent consumption.

228. A nature inspired platform for production of acetyl-CoA derived mevalonate pathway products in *S. cerevisiae*

Sarah Rodriguez*[1, 2], Charles Denby & **Jay Keasling**[1,2,3,4]

[1]California Institute of Quantitative Biosciences (QB3), University of California, Berkeley, California, USA; [2]Joint BioEnergy Institute, Emeryville, California, USA; [3]Department of Chemical & Biomolecular Engineering, University of California, Berkeley, California, USA; [4]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA.

URL: http://www.jbei.org/

Project Goals:
We aim to develop a robust *S. cerevisiae* system to drive the production of isoprenoids and other industrial products derived from acetyl-CoA. Our strategy to increase the abundance of cytosolic acetyl-coA is modeled after the well-studied, endogenous mechanisms of oleaginous, lipid accumulating yeast. Enzymes responsible for oleaginicity divert excess acetyl-coA from the mitochondria while providing a sufficient supply of redox potential and ATP units required for lipid biosynthesis. Specifically, cytoplasmic acetyl-CoA is generated through the activity of ATP:citrate lyase (ACL), which cleaves citrate to form acetyl-CoA and oxaloacetate. The combination of expression of mevalonate producing enzymes, a heterologous ACL from the oleaginous yeast, *Aspergillus nidulans*, and modifications to increase precursor supply, have demonstrated the utility of a robust host engineered with strong acetyl-CoA production for improved flux though the mevalonate pathway.

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

229. Engineering a Balanced MEV Pathway for Biofuels Production

Jorge Alonso-Gutierrez[1], Hector Garcia-Martin[1], Eun-Mi Kim[1], Robert H. Dahl[1], Edward Baidoo[1], Tanveer S. Batth[2], Christopher J. Petzold[2], Paul D. Adams[2], Jay D. Keasling[1], **Taek Soon Lee** (tslee@lbl.gov)[1,*]

[1]Fuels Synthesis and [2]Technology Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California

Project Goals

The mevalonate pathway (MEV) has the potential to produce biofuels such as limonene and bisabolene. The heterologous expression of MEV in *E. coli* removes undesired regulatory points and enables higher titers of the target terpene. However, the unregulated over-expression of enzymes leads to an accumulation of metabolites with toxic and inhibitory effects. Therefore, it is essential to coordinate the relative levels of the pathway enzymes to improve production titers. The goal of this project is to develop tools for pathway balance, and therefore, to achieve high titer production of biofuel compounds.

Abstract

In the present study we showed two different approaches for balancing heterologous MEV gene expression in E.coli and improve production from 1% glucose: i) Engineer a dynamic regulation of the pathway using promoters responding to an accumulation of the toxic intermediate FPP,[1] which improved bisabolene titer to 875 mg/L ii) Explore expression of the MEV pathway enzymes using targeted proteomics and its associated limonene production[2] to define engineering strategies; a principal component analysis (PCA) of the proteomics and the production at various conditions suggests that an over-expression of the bottom portion of the pathway and a moderate expression of the top portion are keys to improve the production. Indeed, extra copies of the terpene synthase and a modified top portion led to 665 mg/L and 1150 mg/L of limonene and bisabolene productions, respectively. This represents a 40% increase in production titer over previous, unbalanced systems using the same genes2.

References

1. Dahl, R.H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T.S., Redding-Johanson, A.M, Petzold, C.J., Mukhopadhyay, A., Lee, T.S., Adams, P.D., Keasling, J.D., Engineering dynamic pathway regulation using stress-response promoters. *Nature Biotech.* **2013**, (doi: 10.1038/nbt.2689)
2. Alonso-Gutierrez, J., Chan, R., Batth, T.S., Adams, P.D., Keasling, J.D., Petzold, C.J., Lee, T.S., Metabolic engineering of Escherichia coli for limonene and perillyl alcohol production. *Metabol. Eng*. **2013**, *19*, 33-41 (doi: 10.1016/j.ymben.2013.05.004)

230. Rapid Kinetic Characterization of Glycosyl Hydrolases (GHs) Based on Oxime Derivatization and Nanostructure-Initiator Mass Spectrometry (NIMS)

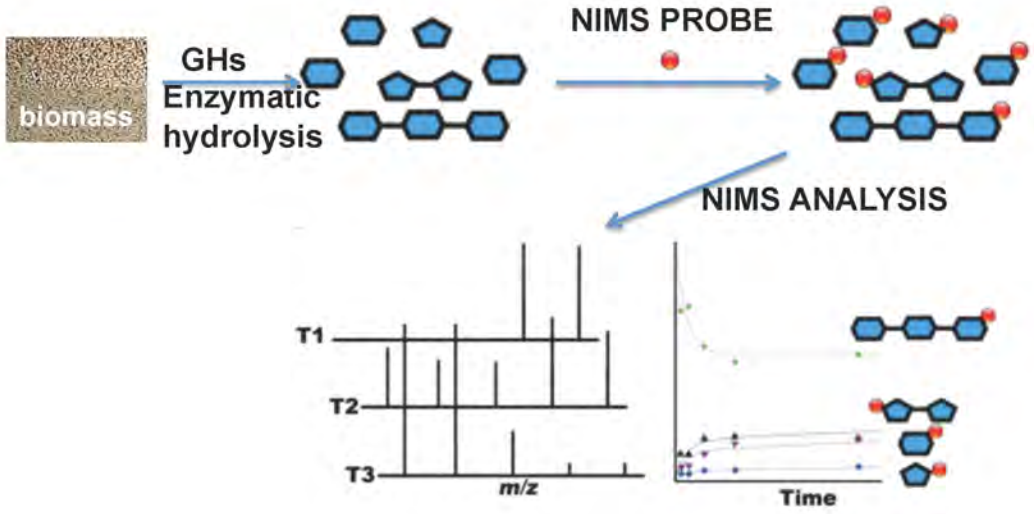
Kai Deng[1,2*](kdeng@lbl.gov), Taichi E. Takasuka[3], Richard Heins[1,2], Xiaoliang Cheng[1,4], Lai F. Bergeman[3], Jian Shi[1,2], Sam Deutsch[4,5], Seema Singh[1,2], Kenneth L. Sale[1,2], **Blake Simmons**[1,2], **Paul D. Adams**[1,4,6], **Anup K. Singh**[1,2], **Brian G. Fox**[3,], **Trent R. Northen**[1,4]

[1]Joint BioEnergy Institute, Emeryville, CA; [2]Sandia National Laboratories, Livermore, CA; [3]Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, WI; [4]Lawrence Berkeley National Laboratory, Berkeley, CA; [5]Joint Genome Institute, Walnut Creek, CA; [6]University of California, Berkeley, CA

http://www.jbei.org

Project Goals: Glycoside hydrolases (GHs) are ubiquitous enzymes in nature and play significant roles in many biological and industrial processes. Unfortunately only a small fraction of the hundreds of thousands of putative GHs have been functionally characterized. Thus methods are urgently needed to rapidly provide direct biochemical evidence. Here we describe an oxime chemistry for nanostructure-initiator mass spectrometry characterization of in situ expressed mono and multifunctional GHs relevant to biomass conversion to biofuels. This approach enabled simultaneous detection of hexose and pentose product cascades from the biomass hydrolysis to understand enzyme kinetics. Overall, this integrated platform of robotic cell-free translation with mass spectrometry analysis can rapidly provide kinetic information for GHs to improve our functional understanding of this important class of enzymes.

Abstract: Glycoside hydrolases (GHs) are critical to cycling of plant biomass in the environment, digestion of complex polysaccharides by the human gut microbiome, and industrial activities such as deployment of cellulosic biofuels. High throughput sequencing methods show tremendous sequence diversity among GHs, yet relatively few examples from the over 150,000 unique domain arrangements containing GHs have been functionally characterized. Here, we show how cell-free expression, bio-conjugate chemistry, and surface-based mass spectrometry allows detection of new glycoside hydrolase specificities and enables kinetic analysis of their activities with plant biomass. Detection of soluble products is achieved by coupling a unique chemical probe to the reducing end of oligosaccharides in a stable oxime linkage, while the use of ^{13}C-labeled monosaccharide standards (xylose and glucose) allows quantitation of the derivatized glycans. We apply this oxime-based nanostructure-initiator mass spectrometry (NIMS) method to characterize the functional diversity of GHs secreted by *Clostridium thermocellum*, a model cellulolytic organism. New reactions are identified for previously unassigned enzymes, and differences in rates and yields of individual enzymes are demonstrated in reactions with biomass substrates. Numerical analyses of time series data suggests that synergistic combinations of mono- and multi-functional GHs can decrease the complexity of enzymes needed for the hydrolysis of plant biomass during the production of biofuels.



The DOE Joint BioEnergy Institute and DOE Great Lakes Bioenergy Research Center are supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 and through contract DE-FC02-07ER64494, respectively. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's Nuclear Security Administration under contract DE-AC04-94AL85000.

231. Improving lignocellulose saccharification by high throughput mutant screening and whole-genome sequencing of the mutants

Guotian Li[1, 2]* (gtli@ucdavis.edu), Mawsheng Chern[1, 2], Rashmi Jain[1], Liangrong Jiang[1], Miguel Vega-Sanchez[1, 2], Patrick E. Canlas[1, 2], Wendy Schackwitz[3], Nicholas Santoro[4], and **Pamela C. Ronald**[1, 2]

[1]Department of Plant Pathology and the Genome Center, University of California, Davis, California 95616; [2]Joint BioEnergy Institute, Emeryville, California 94710; [3]U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA; [4]Department of Energy Great Lakes Bioenergy Research Center, Michigan State University, 162 Food and Safety Toxicology Building, East Lansing, MI 48824, USA.

http://www.jbei.org

Project Goals: We are using rice as the model to identify genes involved in grass cell wall biosynthesis and modification.

To achieve cost efficient conversion of lignocellulosic biomass into biofuels, basic knowledge on the genes that control cell wall recalcitrance and enhanced saccharification is needed. We are using rice, a tractable model grass species, to identify genes involved cell wall biosynthesis and modification. We screened 5,000 rice fast neutron irradiated mutants (M2 lines) using a high throughput rice lignocellulose saccharification assay. From this screen, we identified 95 mutant candidates with altered saccharification efficiency. Of these, 45 mutant candidates showed increased saccharification efficiency and 50 mutant candidates showed reduced saccharification efficiency. Two of the mutants, were further characterized. The *rcs60* mutant segregated for reduced saccharification efficiency in the M2 generation. The *rcs60* progeny were dwarf and impaired in fertility. The *rcs815* mutant displayed 60% increase in saccharification.

We have initiated a whole genome sequencing project of 2,000 of the fast-neutron mutants in collaboration with the Joint Genome Institute (JGI). To date we have analyzed the whole- genome sequence for 45 lines and have detected 300 DNA changes. In the *rcs815* mutant, a single 13 bp homozygous deletion was detected. The deletion caused a frameshift mutation in the bZIP domain of a putative transcription factor that is conserved in switchgrass, sorghum and maize. This research provides an efficient approach to identify novel saccharification-related regulators in rice. The knowledge gained will be useful in improving other biofuel crops, including switchgrass, miscanthus and sorghum.

This work is supported by the Office of Science of the U.S. Department of Energy Contract No. DE-AC02-05CH11231 to the Joint Genome Institute and a National Science Foundation grant (IOS-1237975) to PCR.

232. Mining post translational modifications in Arabidopsis using the ModHunter

Greg W. Mann, Hiren J. Joshi, A. Michelle Smith-Moritz, Harriet T Parsons, Chris J. Petzold and **Joshua L. Heazlewood*** (jlheazlewood@lbl.gov)

Joint BioEnergy Institute and Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA

http://www.jbei.org/

Project Goals: To leverage the array of proteomics data in the reference plant Arabidopsis to identify likely regions within proteins that contain post-translational modifications.

The past decade has seen the development of a plethora of online proteomics resources in Arabidopsis reflecting multiple large-scale studies. These resources exist independently and lack a level of integration. The Multinational Arabidopsis Steering Committee, Proteomics (MASCP) has addressed this issue through the development of a proteomics aggregation portal, MASCP Gator (http://gator.masc-proteomics.org/). The portal provides a summary of proteomics and protein information aggregated directly from ten online resources. The development of this portal has enabled us to develop a bioinformatics technique to identify likely regions of post-translational modifications in proteins of Arabidopsis. The ability to locate and identify post translational modifications experimentally by mass spectrometry is extremely challenging and there is a requirement for complementary techniques. Virtually all large-scale proteomics analyses in Arabidopsis have identified proteins with unmodified peptides. Collectively, these data reveal modified regions of a protein as unmatched areas within a protein model. Using a recent large- scale *N*-linked glycosylation survey as a test set, we could demonstrate that unmatched regions represent modification hotspots in proteins. These sites can be further targeted for investigation and characterization. We have now developed a method to locate putative regions with modifications by exploiting mass spectral data in the public domain and are attempting to develop this into a functional portal for the assessment modifications in proteomic datasets (http://modhunter.masc-proteomics.org/).

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.

233. Introduction of a Transgene Regulation Strategy to Improve Lignin Engineering

Yan Liang[1]*(yliang@lbl.gov), Tania L. Gonzalez[2], Aymerick Eudes[1], Millie Chan[1], Ming C Hammond[2], **Dominique Loqué**[1]

[1]Feedstock Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA and [2]Depts of Chemistry and MCB, University of California, Berkeley, CA

http://www.jbei.org/research/divisions/feedstock/cell-wall-engineering/

Project Goals: Lignin is the one of the major components that contribute to cell wall recalcitrance to degradation or cost-effective biofuel production. To achieve lignin reduction, while retaining some essential physiological functions, the development of sophisticated strategies for lignin engineering is required. One of such strategies is to ensure stringent spacio-temporal control of lignin modifying genes.

In the current study, we explored the possibility of employing a novel genetic regulatory devise to facilitate lignin engineering. By incorporating a synthetic intron cassette[1] into the transgene or its leading sequence, an artificial transcript editing control is introduced to add a regulatory switch in our engineered metabolic pathway in addition to promoter controls. The transcript editing control has been shown to improve the expression tuning by avoiding leaky expression of the transgene and has been applied to achieve promoter stacking for lignin engineering in a tissue specific manner.

Reference

1. Hickey, S. F., Sridhar, M., Westermann. A.J., Qin, Q. Vijayendra, P. L, Geoffrey Hammond, M.C. (2012) Nucleic Acids Research 40, 4701-4710

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.

234. Contribution of multifunctional enzymes to biomass hydrolysis

Christopher M. Bianchetti,[1,2*] (bianchetti@wisc.edu), Taichi E. Takasuka,[1,2] Kai Deng,[3,4] Hannah Udell,[1,2] Lai Bergeman,[1,2] Richard Heins,[3,4] Sam Deutsch,[5,6] Seema Singh,[3,4] Kenneth L. Sale,[3,4] Blake A. Simmons,[3,4] Paul D. Adams,[3,5,7] Anup K. Singh,[3,4] Trent R. Northen,[3,4] and Brian G. Fox,[1,2]

Great Lakes Bioenergy Research Center[1] and the Department of Biochemistry[2] University of Wisconsin–Madison, Joint BioEnergy Institute[3], Sandia National Laboratories[4], Lawrence Berkeley National Laboratory[5], Joint Genome Institute[6], and University of California Berkeley[7]

Project Goals: We are using genomic, biochemical, computational and structural approaches to understand the rate-limiting constraints on enzymatic hydrolysis of biomass prepared by pretreatment methods developed in the GLBRC and the other BRCs. Independent of the pretreatment used, it would be desirable to decrease the complexity of enzyme cocktails required. To address this goal, we are investigating how multi-functional enzymes might be used to simplify the composition of enzyme cocktails. Recent work in collaboration with the Joint Genome Institute (JGI) and the Joint BioEnergy Institute (JBEI) has greatly expanded the number of multi-functional enzymes known and has also given a new method for time-resolved studies of biomass.

Abstract: Higher plants are composed of a complex mixture of cellulose, xylan and mannan in addition to other recalcitrant polymers such as lignin. This complexity provides an effective barrier to deconstruction of plant cell walls. In order to overcome this complex structure, microbes secrete enzyme cocktails containing a wide range of different substrate activities. One potential way to reduce the number of enzymes needed, while retaining deconstruction ability, is by the incorporation of high activity enzymes with broad substrate specificity. The glycoside hydrolase (GH) 5 family, one of the largest GH families, offers a promising starting point for further investigation of the relationships between substrate specificity and hydrolytic activity. Our initial work showed that CelE (Cthe_0797), a GH5 enzyme from Clostridium thermocellum, can hydrolyze all three major polysaccharides found in higher plant cell walls. Using this enzyme as an anchor point, we explored phylogenetically related GH5 enzymes using a combination of gene synthesis, automated protein translation, biochemical assays, and structure determination. This work has revealed additional naturally evolved multifunctional GH5 enzymes. The enzymes exhibit a breadth of catalytic properties, thermal stabilities, and pH optima. Structural analysis has suggested several residues that may play a role in multifunctionality. Cocktails including these enzymes are also effective in the deconstruction of pretreated biomass. The unique structural and catalytic properties of CelE and other newly discovered multifunctional enzymes suggest a new approach for targeted improvement of enzyme cocktails for biomass deconstruction.

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

Appendix A: Additional Abstracts

- Breakout Abstracts
- Plenary Abstracts

BREAKOUT ABSTRACTS

Stimulated Raman Scattering (SRS) and atomic force microscopy (AFM) of enzymatic deconstruction of plant cell walls

X. Sunney Xie, Harvard University; Shi-You Ding, National Renewable Energy Laboratory

A biochemical platform holds the most promising route toward sustainable production of biofuels from lignocellulosic biomass. In this process, lignin is removed from or delocalized in the plant cell walls by a thermochemical pretreatment step; polysaccharides are then hydrolyzed by cellulase enzymes to simple sugars that can be fermented by microbes to produce liquid fuels. Our research is aimed at mechanistic understanding of enzymatic deconstruction of the plant cell walls by means of development and application of non-destructive imaging techniques. We will report observations made using recently developed micro-spectroscopic techniques for real-time correlative imaging of the nanoscale architecture of the plant cell walls and in situ quantification of lignin and polysaccharides during pretreatment and enzymatic hydrolysis. We conclude that lignin plays a negative role in enzyme digestion by physically impeding the penetration of cellulase enzymes into the nanoscale sandwich-like layered structures found in plant cell walls. Removing lignin from these layers is the key to enabling rapid enzymatic digestion *(Zeng et al., 2014 Curr. Opin. Biotechnol.; Ding et al., 2012, Science).*

Comparative fungal genomics for improved biofuel production

Audrey P. Gasch (agasch@wisc.edu)[*1,2], Trey Sato[1], Dana Wohlbach1,[2], Jeff Lewis[1,2], Maria Sardi[1,2], Nikolay Rovinskiy[1,2], Kevin Myers [1,2], Jeff Pietrowski[1,2], Nick Riley[1,3], Joshua J. Coon[1,3]

[1] Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison WI 53706
[2] Laboratory of Genetics, University of Wisconsin-Madison, Madison WI 53706
[3] Department of Chemistry, University of Wisconsin-Madison, Madison WI 53706

Project Goals: Our goal is to engineering Saccharomyces cerevisiae for efficient xylose fermentation in the presence of ligno-toxins present in plant hydrolyates. We are using comparative and functional genomics to first understand bottlenecks in xylose fermentation and ligno-toxin survival and then use this information to improve strains for industrial biofuel production.

Abstract: Plant cellulosic material represents an important source of biomass for biofuel production. However, native strains of Sacchaomyces cerevisiae – the current workhorse of bioethanol production – cannot use pentose sugars that comprise a significant fraction of hemicellulose. Furthermore, toxic byproducts generated during biomass pretreatment induce significant stress responses in yeast, limiting biofuel production. These bottlenecks exacerbate each other, since cells minimally engineered for xylose metabolism display a particular defect in xylose consumption in the presence of ligno-toxins, even when metabolism of glucose continues unabated.

We are using multi-omic analyses to understand these bottlenecks and implicate genes and processes for directed engineering for strain improvement. In this talk, I will summarize several complementary projects that are uncovering targets for directed engineering. A key feature of our approach is comparing and contrasting yeast strains with different capabilities through 1) Genome-wide association studies in hundreds of wild strains, 2) transcriptome analyses of stress-resistant wild strains responding to stresses relevant to biofuel production, and 3) transcriptomics and proteomic analysis of strains serially evolved for anaerobic xylose fermentation. The knowledge garnered through these collaborative projects is significantly expanding our understanding of xylose metabolism and stress tolerance, that will also foster improvement of other strains used in industrial processes.

Funding Statement. This work is being funded by a grant from the Department of Energy to the Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).

Design Principles Controlling Hydrogen Metabolism in Phototrophic Organisms

Alexandra Dubini[1*] (alexandra.dubini@nrel.gov), Matthew Wecker[2] (matt.wecker@nrel.gov), Maria L. Ghirardi [1*](maria.ghirardi@nrel.gov)

[1]National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO 80401
[2]Genebiologics, LLC, Boulder, CO 80303

*Presenting authors

Project goals: To obtain a systems-level understanding of the biological barriers to sustained H_2 photoproduction by *Chlamydomonas reinhardtii*.

Photobiological H_2 production from water is a clean, non-polluting and renewable technology. Although the potential light conversion efficiency to H_2 by biological organisms is theoretically high (about 10%), the system is currently limited by biochemical and engineering constraints. The specific objectives of this research are: (i) development, testing, validation and utilization of novel high-throughput assays to identify photosynthetic organisms with altered H_2-producing activities, thus leading to the discovery of novel strategies to circumvent known biochemical limitations; and (ii) deconvolution of the network of metabolic pathways centered on six ferrredoxin homologs found in Chlamydomonas, aimed at understanding reductant flux in photobiological hydrogen production, and identifying targets for future metabolic pathway engineering strategies to reduce flux to non-productive pathways.

In 2013, we used our petri plate-based H_2 detection assay to screen ~500,000 insertional mutants of *C. reinhardtii* for strains capable of high-light H_2 production. Using the assay, five such strains were isolated and verified for H_2 production under 300µE m^{-2} s^{-1} light in comparison to negative H_2 production for the wild type strain. Characterization of H_2 production is on-going, as is identification of the mutational insertion location within the genome. We are further modifying the assay to be based on fluorescence-activated cell sorting. This will allow specific manipulation of growth conditions for single cells (e.g. light intensity and O_2 tension) as opposed to the current assay where heterogeneous conditions exist within colonies. This assay modification will also allow increased throughput in both screening rate and strain isolation. Toward this goal, we are linking the H_2-sensing Rhodobacter cells to the H_2-producing Chlamydomonas cells, either directly or through attachment to microspheres.

We had previously shown that *Chlamydomonas* FDX1 and FDX2 are capable of mediating redox reactions involving either HYDA1 or FNR, but at different rates. In an effort to better understand the differences in property and function of these proteins, we over-expressed them in *E. coli*, purified their mature versions, and used them for spectroscopic and crystallographic studies. We report the first 3D structure of a *Chlamydomans reinhardtii* ferredoxin, FDX2, at atomic resolution of 1.18 Å, and show that its folding motif is similar to previously published plant-type FDX structures. The FDX2 crystal structure allowed us to refine the homologous FDX1 model structure and to simulate the interaction surfaces of both FDXs with the HYDA1 hydrogenase and the FNR1 protein. Moreover, based on earlier findings from the literature, we used site-directed mutagenesis to mutate two residues in FDX2 to the equivalent residues in FDX1 (M61F, -94Y and the double mutant). Our results demonstrate the mutations resulted in additively higher rates of H_2 production, with the double mutant photo-producing H_2 at 54% of the rate of the FDX1-catalyzed reaction, from 17.5% with WT FDX2. On the other hand, when NADPH photo-production was assayed using these FDX2 mutant proteins, we showed that the catalytic rates (Vmax) for FDX1 and FDX2 were fairly similar, and neither of the mutations significantly affected the rates.

Publications:

1. Peden EA, Boehm M, Mulder DW, Davis R, Old WM, King PW, Ghirardi GL, Dubini A (2013). Identification of global ferredoxin interaction networks in *Chlamydomonas reinhardtii*. J Biol Chem. 288(49):35192-209.

2. Wecker MSA, and Ghirardi ML. (2014) "High-Throughput Biosensor Discriminates Between Different Algal H_2-Photoproducing Strains ". Biotechnol. Bioeng., *in press*.

3. Boehm, M, Alahuhta, M, Long, H, Old, W.M, Peden, E.A, Mulder, D, Brunecky, R, Lunin, V, King, P, Ghirardi, M.L and Dubini, A. "Ferredoxin 2 of *Chlamydomonas reinhardtii*: properties, crystal structure at 1.2 Å resolution and its role in hydrogen and NADPH photo-production". *In preparation.*

Confocal fluorescence microscopy/modeling of lignolytic mechanisms

Chris Hunt (US Forest Service)

Reactive oxygen species (ROS) are powerful one-electron biodegradative oxidants produced by fungi during wood degradation. We present results showing that for the soft rot fungus ascomycete *Daldinia concentrica*, oxidation is contained within the wood and is not observed in the aqueous medium around it. That is, the biodegradative oxidants produced by this fungus are not diffusible. This contrasts sharply with our results for the basidiomycetes *Gloeophyllum trabeum* (brown rot) and *Ceriporiopsis subvermispora* (white rot), which clearly produce diffusible, water-soluble oxidants.

We also discuss the development and potential use of novel fluorescent probes, based on the loss of FRET (fluorescent resonance energy transfer) between linked dyes, which occurs via oxidative cleavage of a polyethylene glycol (PEG) linker. This reaction appears to be specific to highly oxidative, hydrogen-abstracting ROS. These new probes will likely prove useful, in that they should only respond to high energy oxidants that are able to cleave lignin.

Neutron and X-Ray Experiments and Computational Modeling of Pretreatment of Biomass

Paul Langan, [1] Loukas Petridis, [2] Hugh M. O'Neill, [1] Sai Venkatesh Pingali, [1] Marcus Foston,[3] Roland Schulz, [2] Benjamin Lindner, [2] Volker Urban, [1] Barbara R. Evans, [4] Arthur J. Ragauskas, [3] Jeremy C. Smith,[2] Brian Davison[2]

[1]Biology and Soft Matter Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831
[2]Bioscience Division, Oak Ridge National Laboratory, Oak Ridge TN 37831
[3] Institute of Paper Science and Technology, School of Chemistry and Biochemistry and Georgia Institute of Technology, Atlanta, GA, 30332
[4]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge TN 37831

Project Goals:
Lignocellulosic biomass comprises the vast majority of biomass on Earth and has the potential to play a major role in generation of renewable biofuels if cost-effective conversion can be achieved. Largely composed of plant cell walls, it is a complex biological composite material that is recalcitrant to the structural deconstruction and enzymatic hydrolysis into sugars that is necessary for fermentation to bioethanol. The Scientific Focus Area in Biofuels is developing "Dynamic Visualization of Lignocellulose Degradation by Integration of Neutron Scattering Imaging and Computer Simulation" for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis. This is providing fundamental information about the structure and deconstruction of plant cell walls that is needed to drive improvements in the conversion of renewable lignocellulosic biomass to biofuels.

Abstract
Lignocellulosic biomass is potentially an important non-edible renewable organic source of energy and chemical feedstock. However, biomass is difficult to exploit in practice because it resists degradation to glucose in industrial hydrolysis processes and thus requires expensive thermochemical pretreatments. Understanding the mechanism of breakdown of the hydrated matrix of hemicellulose and lignin polymers that encrusts the glucose-rich fibrous cellulose during these thermochemical pretreatments will lead to more efficient use of biomass. However, characterizing the morphology of biomass as it evolves in response to thermochemical pretreatment has been difficult because of its hierarchical structure and the complexity of its components. By combining multiple probes of structure, sensitive to different length scales, with molecular dynamics simulations, we reveal some fundamental processes responsible for the morphological changes in biomass during steam explosion pretreatment. We further show that the basic driving forces are the same in some other leading thermochemical pretreatments, such as dilute acid pretreatment. The compensation between the entropy and enthalpy of hydration drives the cell wall components over kinetic barriers, destabilizing them, and causing irreversible morphological changes. Our findings suggest that new pretreatments and plant modifications that promote lignin and hemicellulose phase separation and increase the porosity of the cell wall matrix while preventing increases in cellulose crystallization as a result of dehydration will improve biomass conversion.

Using MetRxn for metabolic model reconstruction, flux elucidation and redesign

Costas D. Maranas (costas@psu.edu)[1]

[1]Department of Chemical Engineering, Pennsylvania State University, University Park, PA

http://maranas.che.psu.edu/

Project Goals: This project aims to organize and disseminate standardized metabolite and reaction information to improve metabolic modeling by accurately describing reaction stoichiometry, directionality, atom mapping from reactants to products, and gene to protein to reaction relations. This resource is deployed for microbial, multi-tissue and multi-organism metabolic model reconstruction, metabolic flux elucidation using labeled substrates (MFA) and computational strain design.

MetRxn is a standardized non-redundant searchable collection of published metabolic models and databases from a wide variety of organisms. The current MetRxn 2.0 update includes recently published metabolic data for a total of 112 metabolic models and 8 metabolic databases. The number of distinct reactions that have been mapped is greater than 20,000 and MetRxn contains tools that allow users to download atom mapping data for each reaction. In this talk, we will elaborate on new features of MetRxn 2.0 (http://www.metrxn.che.psu.edu/) including atom mapping information across all reactions and enhanced integration with other databases. We will describe how this resource can impact genome-scale metabolic model reconstruction by providing curated reaction and metabolite content. Progress towards the development of a multi-tissue metabolic model for maize, rapid generation of cyanobacterial models and metabolic modeling of microbial communities will be briefly highlighted. Reaction atom transition information in MetRxn can rapidly be leveraged to create genome-scale atom mapping models. Efforts towards resolving metabolic fluxes at a genome-scale and computational challenges with current flux elucidation tools will be described and the impact on flux elucidation fidelity will be quantified.

Existing computational strain design approaches relying solely on stoichiometry and rudimentary constraint-based regulation overlook the effect of metabolite concentrations and substrate-level enzyme regulation while identifying metabolic interventions. This may lead to suggested interventions that cannot be implemented. To remedy this, we developed the k-OptForce procedure which integrates all available kinetic descriptions of metabolic reactions with stoichiometric models to sharpen the prediction of intervention strategies for improving the bio-production of a chemical of interest. In addition, we have used the Ensemble Modeling (EM) procedure for constructing kinetic models of core *E. coli* metabolism consistent with available measurements (metabolomic & fluxomic).

Supported by funding from the U.S. Department of Energy to Dr. Costas D. Maranas grant DE-FG02-05ER25684.

Metal Uptake by Methanotrophs: Genetic Basis for the Biosynthesis of A Novel Chalkophore and Molecular Spectroscopic Analyses of Mercury Detoxification

Jeremy D. Semrau[1*] (jsemrau@umich.edu), Alexey Vorobev[1], Sheeja Jagedevan[1], Bipin Baral[2], Nathan Bandow[2], and Alan A. DiSpirito[2],

[1]University of Michigan, Ann Arbor, Michigan; [2]Iowa State University, Ames, Iowa

http://sitemaker.umich.edu/methanobactin

Project Goals: Methanotrophs are ubiquitous in the environment, and despite their critical function in many different ecosystems, the biogeochemical factors that affect their activity and community structure are poorly understood. It is known that copper plays a key role in methanotrophic physiology, but the mechanism used by these microbes for copper acquisition was only recently discovered. This compound, methanobactin (mb), is the first example of a "copper-siderophore" or chalkopore. Mb binds many different metals, including mercury. Further, recent data show that different methanotrophs make different forms of mb that have varying metal affinities. The general objectives of this proposal are thus to consider how mb made by different methanotrophs alters the bioavailability of metals of concern to the DOE and how this affects: (1) the physiology, metabolism and gene expression in methanotrophs; (2) the broader microbial community structure and activity in laboratory soil columns, and; (3) the bioavailability of different metals in subsurface environments.

One of the persistent and substantial problems in remediation of hazardous waste sites is the mobilization and uncontrollable transport of radionuclides and heavy metals from these sites to surrounding areas. Some microbially-mediated processes can at least temporarily immobilize and reduce the toxicity of these materials through dissimilatory reduction that leads to precipitation and sorption under anaerobic conditions. As such, microbial-mediated processes can limit the dispersal of these materials and thus also limit the exposure of surrounding areas. Microorganisms, however, have effective and ubiquitous mechanisms to solubilize different metals and that non-specific binding of metals by these biogenic metal chelators may increase their solubility, mobility, and bioavailability. Here we are examining how the expression of metal chelating agents analogous to siderophores in methane-oxidizing bacteria i.e., methanotrophs, alters the bioavailability of various metals (copper and mercury) and how this: affects the (1) physiology, metabolism and gene expression in methanotrophs; (2) broader microbial community structure and meta-transcriptome, and; (3) bioavailability and risk associated with various metals. Such studies will enable us to determine how methanotrophic activity may affect the structure of subsurface microbial communities as well as the sustainability of subsurface waters, including at DOE sites.

Recent work in our laboratories has identified the genetic basis of mb and that many, but not all methanotrophs can synthesize mb. Interestingly, mb contains two heterocyclic rings, either imidazole, oxazolone or pyrazinedione rings with an associated enethiol group, which together are responsible for metal binding. Given the structure of the rings, it is quite possible that mb can also bind toxic metals such as mercury and that mb made by one methanotroph may affect the bioavailability of metals to other methanotrophs. Our findings show that mb from *Methylosinus trichosporium* OB3b does indeed bind mercury in addition to copper, and in doing so, reduced toxicity associated with Hg(II) to both α- and γ-Proteobacteria methanotrophs. Interestingly, mercury binding by mb was evident both in the presence and absence of copper, despite the fact that mb had a much higher affinity for copper due to the rapid and irreversible binding of mercury by mb. Metal analyses indicated that Hg(II), after bound by mb, may have been reduced to Hg(0) but was not volatilized. Rather, mercury remained associated with mb, and also was found associated with methanotrophic biomass. It thus appears, although the mercury-mb complex was cell-associated, mercury was not removed from mb.

It was also found that the amount of biomass-associated mercury in the presence of methanobactin from *M. trichosporium* OB3b was greatest for *M. trichosporium* OB3b and least for the tested γ-Proteobacteria methanotroph *(Methylomicrobium album* BG8), suggesting that methanotrophs may have selective mb uptake systems that may be based on TonB-dependent transporters, but that such uptake systems exhibit a degree of infidelity. Further, it was found that the addition of mb from *M. trichosporium* OB3b stimulated the growth of other methanotrophs in the absence of mercury but in the presence of copper. As methanotrophs expressing the particulate methane monooxygenase (the predominant form of methane monooxygenase expressed) require copper for high activity, it may be that methanobactin from *M. trichosporium* OB3b increased the bioavailability of copper, thereby increasing activity of pMMO in other methanotrophs. If so, this suggests that mb made by one methanotroph may actually be taken up by others (as also suggested by the mercury uptake data). Collectively, these studies raise several interesting questions, including do all methanotrophs in a mixed community produce mb, or do some species act as "cheaters" and rely on mb made by other microbes to meet copper requirements for metabolism or for detoxification of metals such as mercury? How do methanotrophs that make mb ensure that they are able to effectively compete with such cheaters for copper? It appears that the mb uptake systems have some infidelity, and that *in situ*, such infidelity may be optimized as a general competition strategy.

Ongoing work is characterizing the transcriptome of both α- and γ-Proteobacteria methanotrophs under a range of copper concentrations to determine how copper affects overall gene expression in methanotrophs. We have also sequenced the genome and transcriptome of the novel facultative methanotroph, *Methylocystis* strain SB2, grown with different carbon sources. We have also recently collected soil samples from the Integrated Demonstration Site of the DOE Savannah River Site where methanotrophs are known to exist. We are in the process of characterizing the initial microbial community composition and will also construct a series of soil columns to characterize how mb affects: (1) copper and mercury mobility and bioavailability in the presence of soils from this site, and; (2) dissolution of soil-associated minerals. The resultant effects on the broader microbial community structure and function will be determined via metagenomics and metatranscriptomics.

Publications (1) Bandow, et al. 2012. *J. Inorgan. Biochem*. 110:72-82. (2) Semrau, et al. 2013. *Environ Microbiol*. 15:3077-3086. (3). Vorobev, et al. 2013. *Appl. Environ Microbiol*. 79:5918-5926. (4) Jagadevan & Semrau. 2013. *Appl. Microbiol. Biotechnol*. 97:5089-5096. (5) Vorobev, et al., 2014. *Appl. Environ. Microbiol* (submitted).

This research was supported by the Office of Science (BER), U.S. Department of Energy.

3D Multi-resolution visualization of the early stages of cellular uptake of peptide-coated nanoparticles

Kevin Welsher and **Haw Yang*** (hawyang@princeton.edu)
Princeton University, Princeton

http://www.princeton.edu/~yanglab/research/energy-biol.shtml

Project Goals: The project, "Development of Quantum Dot Probes for Studies of Synergy between Components of the Wood-Degrading Fungal Enzymes," aims to develop quantum dot-based tagging and imaging technologies tailored for simultaneously monitoring, in real time and in the natural fungal / lignocellulose environment, the mode of action of several lignocellulosic enzymes at the single-molecule level. It is designed to be the first project of a long-term research program for which the overarching goal is to bridge the knowledge gap between the in vitro molecular biochemistry and the naturally occurring biodegradation pathways by a quantitative determination of the biochemical and biophysical properties of these fungal enzymes in realistic plant biomass-microbe milieus.

A major aim of the research program in this funding period is to develop a new way of directly visualizing the cellulase molecular actions (fast, nanometer scale) in the context of the biological environment (slower, micron to millimeter scale) in full three-dimensional (3D) fidelity. Indeed, biological processes, including the production of biofuels either by biodegradation or by biosynthesis, are inherently multiscale. In this breakout presentation, we will summarize the currently available methods, point out exciting opportunities for future developments, and describe a new visualization modality that will help to bridge the gap. As proof of concept, we applied the new instrument to visualizing the early stages of cellular uptake of Tat-peptide coated nanoparticles.[1] A detailed understanding of the cellular uptake process is essential to the development of cellular delivery strategies and to the study of viral trafficking; however, visualization of the entire process, encompassing the fast dynamics—local to the freely diffusing nanoparticle—as well the state of the larger-scale cellular environment, remains challenging. Using the new 3D multi-resolution apparatus, we were able to capture, in real time, the transient events leading to cellular binding and uptake of HIV1-Tat-modified nanoparticles. The direct observation of the nanoparticles landing on the cellular contour in 3D revealed long-ranged deceleration of the delivery particle possibly due to interactions with cellular receptors. Furthermore, by using the nanoparticle as a nanoscale "dynamics pen," we discovered an unexpected correlation between small membrane terrain structures and local nanoparticle dynamics. We are currently applying this new approach to unravel the hidden mechanistic steps in the molecular actions of cellulosic biodegradation.

Publications

1. Kevin Welsher and Haw Yang, *Nat. Nanotechnol.* (2014) in press

We gratefully acknowledge the generous support by the US Department of Energy under DE-SC0006838.

PLENARY ABSTRACTS

The DOE Systems Biology Knowledgebase (KBase): Progress towards a system for collaborative and reproducible inference and modeling of biological function

Adam P. Arkin[*1] (aparkin@lbl.gov), Jason Baumohl[1], Aaron Best[2], Jared Bischof[2], Ben Bowen[1], Tom Brettin[2], Tom Brown[2], Shane Canon[1], Stephen Chan[1], John-Marc Chandonia[1], Dylan Chivian[1], Ric Colasanti[2], Neal Conrad[2],, Brian Davison[3], Matt DeJongh[6], Paramvir Dehal[1], Narayan Desai[2], Scott Devoid[2], Terry Disz[2], Meghan Drake[3], Janaka Edirisinghe[2], Gang Fang[7], José Pedro Lopes Faria[2], Mark Gerstein[7], Elizabeth M. Glass[2], Annette Greiner[1], Dan Gunter[1], James Gurtowski[5], Nomi Harris[1], Travis Harrison[2], Fei He[4], Matt Henderson[1], Chris Henry[2], Adina Howe[2], Marcin Joachimiak[1], Kevin Keegan[2], Keith Keller[1], Guruprasad Kora[3], Sunita Kumari[5], Miriam Land[3], Folker Meyer[2], Steve Moulton[3], Pavel Novichkov[1], Taeyun Oh[8], Gary Olsen[9], Bob Olson[2], Dan Olson[2], Ross Overbeek[2], Tobias Paczian[2], Bruce Parrello[2], Shiran Pasternak[5], Sarah Poon[1], Gavin Price[1], Srivdya Ramakrishnan[5], Priya Ranjan[3], Bill Riehl1, Pamela Ronald[8], Michael Schatz[5], Lynn Schriml[10], Sam Seaver[2], Michael W. Sneddon[1], Roman Sutormin[1], Mustafa Syed[3], James Thomason[5], Nathan Tintle[6], Will Trimble[2], Daifeng Wang[7], Doreen Ware5, David Weston[3], Andreas Wilke[2], Fangfang Xia[2], Shinjae Yoo[4], Dantong Yu[4], Bob Cottingham[3] , Sergei Maslov[4] , Rick Stevens[2]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA, [2]Argonne National Laboratory, Argonne, IL, [3] Oak Ridge National Laboratory, Oak Ridge, TN, [4]Brookhaven National Laboratory, Upton, NY, [5]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, [6]Hope College, Holland, MI, [7]Yale University, New Haven, CT, [8]University of California, Davis, CA, [9]University of Illinois at Champaign-Urbana, Champaign, IL, [10]University of Maryland, College Park, MD

http://kbase.us

Project Goals: The KBase project aims to provide the computational capabilities needed to address the grand challenge of systems biology: to predict and ultimately design biological function. KBase enables users to collaboratively integrate the array of heterogeneous datasets, analysis tools and workflows needed to achieve a predictive understanding of biological systems. It incorporates functional genomic and metagenomic data for thousands of organisms, and diverse tools including (meta)genomic assembly, annotation, network inference and modeling, thereby allowing researchers to combine diverse lines of evidence to create increasingly accurate models of the physiology and community dynamics of microbes and plants. KBase will soon allow models to be compared to observations and dynamically revised. A new prototype Narrative interface lets users create a reproducible record of the data, computational steps and thought process leading from hypothesis to result in the form of interactive publications.

Systems biology is driven by the ever-increasing wealth of data resulting from new generations of genomics-based technologies. With the success of genome sequencing, biology began to generate and accumulate data at an exponential rate. In addition to the massive stream of sequencing data, each type of technology that researchers use to analyze a sequenced organism adds another layer of complexity to the challenge of understanding how different biological components work together to form a functional living system. Achieving this systemslevel understanding of biology will enable researchers to predict and ultimately design how biological systems will function under certain conditions. A collaborative computational environment is needed to bring researchers together so they can share and integrate large, heterogeneous datasets and readily use this information to develop predictive models that drive scientific discovery.

The advancement of systems biology relies not only on sharing the results of projects through traditional

methods of peer review and publication, but also on sharing the datasets, workflows, software, models, best practices, and other essential knowledge that made those published results possible. Establishing a common framework for managing this knowledge could save time, reduce duplicative effort, and increase the scientific return on investment in systems biology research. This framework for precisely tracking what was done to achieve a certain outcome also will empower researchers to reproduce published results and review projects more effectively. This ability to reproduce experiments described in publications is a key principle of the growing Open Science movement.

The KBase team is developing an open-source, open-architecture framework for reproducible and collaborative computational systems biology. KBase's primary scientific aim is to push multiple types of functional data towards increasingly specific models of metabolic and regulatory behavior of microbes, plants and their communities. We have brought together data and tools that allow probabilistic modeling of gene function, which can be used in turn to produce experimentally testable models of cellular metabolism and gene regulation. The system will soon allow models to be compared to observations and dynamically revised.

One of the new components that will enable these complex, iterative analyses is KBase's new prototype Narrative Interface, which provides a transparent, reproducible, and persistent view of the computational steps and thought processes leading to a particular conclusion or hypothesis. These “active publications” enable researchers anywhere in the world to re-use a workflow, follow chains of logic, and experiment with changing input data and parameters. Augmented by the KBase Workspace, users can upload their own data, integrate and analyze them with information from the public data stores, and share their data and results with colleagues. As more researchers are able to test the algorithms, datasets, and workflows that are shared with the larger community, they can begin to apply these diverse approaches to their own datasets and provide more informed feedback for improvement. In addition to sharing the information and tools essential to reproducing results, KBase's computational environment is open and extensible, allowing users to add tools and functions to meet their particular research needs.

By enabling members of the community to integrate and use a wide spectrum of analysis tools and datasets, KBase will serve as a catalyst for biological research, accelerating discovery for DOE missions and providing insights and benefits that can ultimately serve numerous application areas.

KBase is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

Index