DOE/SC-0140

# DOE GENOMIC SCIENCE

## SYSTEMS BIOLOGY FOR ENERGY AND ENVIRONMENT

## Joint Meeting 2011

## Genomic Science Awardee Meeting IX

## and

## USDA-DOE Plant Feedstock Genomics for Bioenergy Awardee Meeting

USDA

U.S. DEPARTMENT OF ENERGY

Office of Science

April 10–13, 2011

# DOE Genomic Science Program

**U.S. Department of Energy**
**Office of Science**
**Office of Biological and Environmental Research**
**genomic.science@science.doe.gov**


Meeting Contact
**Catherine Ronning**
U.S. Department of Energy
Office of Biological and Environmental Research
301.903.9549
catherine.ronning@science.doe.gov

Director, Biological Systems Science Division
**Sharlene Weatherwax**
U.S. Department of Energy
Office of Biological and Environmental Research
301.903.3213
sharlene.weatherwax@science.doe.gov

# USDA-DOE Plant Feedstock Genomics

**Catherine Ronning**
U.S. Department of Energy
Office of Biological and Environmental Research
301.903.9549
catherine.ronning@science.doe.gov

**Ed Kaleikau**
U.S. Department of Agriculture
National Institute of Food and Agriculture
ekaleikau@nifa.usda.gov


United States Department of Agriculture
National Institute of Food and Agriculture

An electronic version of this document is available at the Genomic Science web site:
http://GenomicScience.energy.gov/pubs/2011abstracts/

DOE/SC-0140

U.S. DEPARTMENT OF
ENERGY
Office of Science

# Joint Meeting 2011

## Genomic Science Awardee Meeting IX

**and**

## USDA-DOE Plant Feedstock Genomics for Bioenergy Awardee Meeting

**[Revised: April 14, 2011]**

## Crystal City, Virginia
## April 10-13, 2011

Prepared for the
U.S. Department of Energy
Office of Science
Office of Biological and Environmental Research
Germantown, MD 20874-1290

http://genomicscience.energy.gov

Prepared by
Biological and Environmental Research Information System
Oak Ridge National Laboratory
Oak Ridge, TN 37830
Managed by UT-Battelle, LLC
For the U.S. Department of Energy
Under contract DE-AC05-00OR22725

# Welcome

**Department of Energy**
**Washington, DC 20585**
**April 10, 2011**

Dear Participant:

Welcome to the 2011 Genomic Science Annual Contractor-Grantee Meeting/USDA-DOE Plant Feedstock Genomics for Bioenergy Program Meeting. As the premier workshop for the Genomic Science program, this meeting brings together researchers supported by the program, representatives from the Department of Energy, and colleagues from other federal agencies. The Genomic Science program—within the Biological Systems Science Division of the Office of Biological and Environmental Research (BER)—supports fundamental research to achieve a predictive, systems-level understanding of plants, microbes, and biological communities through the integration of fundamental science and technology development. This program provides the foundation for biological solutions to DOE mission challenges in energy, the environment, and climate. Program objectives are:

- *Determine the genomic properties, molecular and regulatory mechanisms, and resulting functional potential of microbes, plants, and biological communities central to DOE missions.*
- *Develop the experimental capabilities and enabling technologies needed to achieve a genome-based, dynamic systems-level understanding of organism and community functions.*
- *Develop the knowledgebase, computational infrastructure, and modeling capabilities to advance the understanding, prediction, and manipulation of complex biological systems.*

This past year has been exciting and productive for the Genomic Science program as it continues to support groundbreaking research by individual investigators, interdisciplinary research teams, user facilities such as DOE's Joint Genome Institute (JGI) and Environmental Molecular Sciences Laboratory (EMSL), and the Bioenergy Research Centers (BRCs). The BRCs have completed their third year of operations and will report on exciting new results that are paving the way for conversion of cellulosic biomass to biofuels. In 2010, the first projects were awarded under BER's Biological Systems Research on the Role of Microbial Communities in Carbon Cycling program, and this year's meeting will feature a breakout session devoted to linking molecular- and global-scale phenomena to understand the global carbon cycle.

Over the past year, researchers in the Genomics Science program have developed new methods to image, sequence, characterize, and engineer microbial systems and to gain insights about plant-microbial interactions that are key to understanding efficient, sustainable energy systems and that meet DOE missions in energy and the environment. Work in this area will be highlighted in the keynote address by invited presenter Dr. Sharon Long and in plenary sessions dedicated to Bioenergy Research, Plant Feedstocks Genomics for Bioenergy, and Innovative Tools for Genomic Sciences.

As has become tradition in alternate years, this meeting once again convenes researchers supported by the Genomics Science program and by the joint USDA-DOE Plant Feedstock Genomics for Bioenergy program, which focuses on genomics-based research to improve the use of biomass and plant feedstocks for producing biofuels. By bringing together these distinct but complementary research communities, we hope to stimulate the exchange of ideas, sharing of expertise, and formation of new partnerships. Both a plenary session and breakout session will be held to discuss some very important developments in feedstock genomics research that have taken place since these groups last met. This meeting provides an excellent opportunity for members of the plant science community to meet and interact with scientists

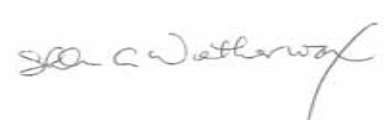
developing the DOE Systems Biology Knowledgebase (Kbase), an important new program within BER. Kbase has the potential to significantly enhance the translation of genomics tools to plant breeding research and thus accelerate development of dedicated bioenergy crops.

The Genomic Science program is committed to supporting research through outstanding DOE user facilities. Plenary talks will highlight two BER-funded national user facilities, JGI and EMSL, and also will describe progress on Kbase—a cyberinfrastructure to integrate, search, and visualize experimental data, metadata, corresponding models, and analysis tools. The Kbase implementation plan, which was developed based on a series of community workshops, articulates scientific objectives in microbial, plant, and metacommunity research and describes infrastructure and architectural details of the system. The implementation plan is available at genomicscience.energy.gov/compbio/kbase_plan/.

The Genomic Science program depends on developments in fundamental research in these technologies. Breakout sessions on databases and functional annotation as well as computational biology and modeling will be held to discuss the current state of the science in these critical areas and to identify needs and resources for the future. We also have added new breakout sessions to discuss recent research advances in Genome Structure and Epigenetics and in Metabolic Pathway Analysis. Based on its continued success, we will again include lunchtime presentations by undergraduate and graduate research investigators. Finally, we are proud to present a plenary session featuring BER awardees of the first DOE Office of Science Early Career Award program. This next generation of promising young scientists represents fields as diverse as biomass and plant feedstocks, microbial processes for bioenergy and the environment, and imaging and instrumentation.

We look forward to an exciting and productive meeting and encourage you to exchange ideas and share your expertise with other researchers in discussions during the plenary presentations, breakout meetings, and poster sessions. We thank you for lending your knowledge, creativity, and vision to the Genomic Science program and wish you continued success in the coming year.

Sincerely,



Sharlene C. Weatherwax, Ph.D.
Director, Biological Systems Science Division, SC-23.2
Office of Biological and Environmental Research
Office of Science

# Contents

# Agenda

**Correct as of March 24, 2011**

Even-numbered posters will be presented on Monday, odd-numbered posters on Tuesday. Please set up the posters no earlier than 5 p.m. on Sunday and leave them up until noon Wednesday.

## Sunday, April 10

| | |
|---|---|
| 5:00–8:00 pm | Early Registration and Poster Setup—Independence Foyer, Independence Level |
| 6:00–8:30 pm | Reception and Scientific Mixer |

## Monday, April 11

| | |
|---|---|
| 7:00-8:30 am | **Registration**—Independence Foyer, Independence Level |
| 8:30–9:00 am | **Welcome and State of Genomic Science Program**<br>**Sharlene Weatherwax**, Director, Biological Systems Science Division |
| 9:00–11:00 am | **Plenary Session: DOE Bioenergy Research Centers**<br>**Moderator**: John Houghton |

**Speakers:**

| | |
|---|---|
| 9:00–9:30 am | **Henrik Scheller, JBEI and Lawrence Berkeley National Laboratory**<br>**Title:** Optimizing Plant Cell Walls for Efficient Biofuel Production |
| 9:30–10:00 am | **Rick Dixon, BESC and Samuel Roberts Noble Foundation**<br>**Title:** Gene Targets for Improving Biomass Deconstruction and Fermentation in Switchgrass |
| 10:00–10:30 am | **Break** |
| 10:30–11:00 am | **Ken Keegstra, GLBRC and Michigan State University**<br>**Title:** Understanding and Manipulating Hemicellulose Biosynthesis |

| | |
|---|---|
| 11:00 am–12:00 pm | **Keynote Presentation:**<br>**Sharon Long**, Stanford University<br>**Title:** Symbiosis: The Once and Future Sustainability? |
| 12:30–2:00 pm | **Lunchtime Student Oral Presentations**<br>**Moderator:** Libby White |

**Breakout Sessions**

2:00–5:00 pm **Breakout Session A: Plant Feedstocks Genomics for Bioenergy**
**Moderator:** Cathy Ronning

**Description of Session:** The joint USDA-DOE Plant Feedstocks Genomics for Bioenergy research program supports genomics-based research leading to the improved use of biomass and plant feedstocks for the production of fuels such as ethanol or renewable chemical feedstocks. This session will focus on fundamental research on plants to improve biomass characteristics, biomass yield, or that will facilitate lignocellulosic degradation. Also, systems biology approaches to identify genetic indicators enabling plants to be efficiently bred or manipulated, or research that yields fundamental knowledge of the structure, function and organization of plant genomes leading to improved feedstock characterization and sustainability.

**Speakers:**

2:00–2:10 pm **Cathy Ronning, DOE-BER**
**Session Introduction**

2:10–2:30 pm **Amy Brunner,** Virginia Tech University
**Title:** Poplar Biomass Protein-Protein Interactions

2:30–2:50 pm **Ed Buckler**, USDA-ARS and Cornell University
**Title:** Developing Genomic Selection (GS) and Genome-Wide Association Studies (GWAS) for Upland Switchgrass

2:50–3:10 pm **Zhaohua Peng**, Mississippi State University
**Title:** Role of Histone Modifications in the Regulation of Cell Wall Synthesis in Rice (*Oryza sativa*)

3:10–3:30 pm **Break**

3:30–3:50 pm **Wilfred Vermerris**, University of Florida
**Title:** Genetic Dissection of Bioenergy Traits in Sorghum

3:50–4:10 pm **Jan Leach**, Colorado State University
**Title:** Identifying Genes and Networks for Increasing Biomass Production in New Energy Grasses by Using Rice as a Model System

4:10–4:30 pm **Robin Buell**, Michigan State University
**Title:** The Biofuel Feedstock Genomics Resource - A Comparative Gene and Transcript Annotation Database for Lignocellulosic Feedstock Species

4:30–4:50 pm **Jose Gonzales**, South Dakota State University
**Title:** Genetic and Genomic Resources in Prairie Cordgrass

2:00–5:00 pm **Breakout Session B: Metabolic Pathway Analysis**
**Moderators:** John Houghton and Roland Hirsch

**Description of Session:** Metabolically engineered microbes and plants are increasing the options for solutions to selected DOE missions, such as production of renewable transportation fuels. Also, predicting a biological response to changed environmental parameters will be critical to other DOE missions. Better understanding of the underlying processes has become increasingly reliant on determining and modeling metabolic pathways. Progress has been facilitated by advances in DNA sequencing; new technologies that allow better characterization of intracellular transcripts, proteins, and metabolites; and comprehensive metabolic and regulatory computer models. Yet many fundamental challenges remain, such as predicting ancillary effects of engineered changes to metabolic pathways and complex responses to environmental perturbations. This breakout session will focus on case studies with an emphasis on sharing challenges and solutions, new techniques, and future research directions. The presentations will include research on metabolic pathway analysis with applications such as the synthesis of next generation liquid transportation fuels as well as a discussion of the ethical, legal, and societal implications.

**Speakers:**

2:00-2:10 pm **John Houghton and Roland Hirsch, DOE-BER**
**Session Introduction**

2:10-2:40 pm **Himadri Pakrasi**, Washington University
**Title:** Metabolic Potentials of Cyanobacteria for Advanced Biofuel Production

2:40-3:10 pm **James C. Liao**, University of California, Los Angeles
**Title:** Design, Evolution, and Ensemble Modeling of High Flux Microbial Synthesis Networks

3:10-3:30 pm **Break**

3:30-4:00 pm **Costas Maranas**, Pennsylvania State University
**Title:** Using Computations to Facilitate Metabolic Reconstructions and Guide Strain Optimization

4:00-4:30 pm **Andrzej Joachimiak,** Argonne National Laboratory
**Title:** Exploring Microbiomes for New Enzyme Activities and Reengineering Their Function

4:30-5:00 pm **Nathan Hillson**, JBEI and Lawrence Berkeley National Laboratory
**Title**: ELSI Pilot: Assessing and Mitigating the Risks of Large-Scale Metabolic Engineering

2:00–5:00 pm **Breakout Session C: Databases and Functional Annotation**
**Moderators:** Dan Drell and Susan Gregurick

**Description of Session:** The DOE Genomic Science program supports research to better understand gene functions and to increase access to and use of the masses of data emerging from DOE-funded research. This session will focus on research efforts and experimental activities to improve functional assignments of newly sequenced gene products and genomes and to more effectively store and make available to users the masses of genome sequence data and increasing amounts of associated data ("metadata"). Future opportunities in annotation will be discussed.

**Speakers:**

2:00–2:20 pm **Victor Markowitz**, DOE Joint Genome Institute
**Title:** Improving the Consistency and Completeness of Microbial Genome Annotations in an Integrated Database Context

2:20–2:40 pm **Frank Collart**, Argonne National Laboratory
**Title:** Functional Characterization by Mapping Protein-Ligand Interactions

2:40–3:00 pm **Margie Romine**, Pacific Northwest National Laboratory
**Title:** Use of Protein Localization Prediction in Functional Annotation – Obstacles and Opportunities

3:00–3:20 pm **John Orban**, University of Maryland
**Title:** Functional Annotation of Putative Enzymes in Methanogens

3:20–3:40 pm **Break**

3:40–3:50 pm **João Carlos Setubal**, Virginia Polytechnic Institute and State University
**Title:** GO Consortium Announcement: MENGO - The Microbial ENergy Processes Gene Ontology Workshop

3:50–4:10 pm **Folker Meyer**, Argonne National Laboratory
**Title:** Version 3 of the MG-RAST Server: A Metagenomics Portal for a Fully Democratized Sequencing World

4:10–4:30 pm **Owen White**, University of Maryland
**Title:** The Open Data Framework for Access and Analysis of Microbiome Data

4:30–5:30 pm **Discussion Around Critical Questions:**

- What can we annotate well?
- What annotation challenges are we poor at?
- What tools do we need to do a better job?

5:30–8:00 pm **Poster Session**

# Tuesday, April 12

8:30–10:00 am **Plenary Session: User Facilities**
**Moderator**: Dan Drell

**Speakers:**

8:30–9:00 am **Eddy Rubin**, DOE Joint Genome Institute
**Title:** DOE Joint Genome Institute Update

9:00–9:30 am **Allison Campbell**, Environmental Molecular Sciences Laboratory
**Title:** EMSL: High-Impact Science and State-of-the-Art Capabilities at a National Scientific User Facility

9:30–10:00 am **Susan Gregurick**
**Title:** The DOE Systems Biology Knowledgebase Update

10:00–10:30 am **Break**

10:30 am–12:00 pm **Plenary Session: Plant Feedstocks Genomics for Bioenergy**
**Moderator:** Ed Kaleikau (USDA-NIFA)

**Speakers:**

10:30–11:00 am **Maria Harrison**, Boyce Thompson Institute for Plant Research
**Title:** Toward an Understanding of AM Symbiosis for Growth of Biomass Feedstocks in Low-Input Systems

11:00–11:30 am **Jeff Bennetzen**, University of Georgia
**Title:** *Setaria* as a Model for Grass Genetics and Biofuel Feedstock Genomics

11:30 am-12:00 pm **Christian Tobias,** USDA-ARS Western Regional Research Center
**Title:** Integration of Switchgrass into the Grass Consensus Genome

12:30–2:00 pm **Lunchtime Student Oral Presentations**
**Moderator:** Libby White

**Breakout Sessions**

2:00–5:00 pm **Breakout Session D: Microbes and the Global Carbon Cycle**
**Moderator:** Joe Graber

**Description of Session:** Understanding and predicting processes of the global carbon cycle will require new research approaches aimed at linking global-scale climate phenomena, biogeochemical processes of ecosystems, and functional activities encoded in the genomes of microbes, plants, and complex biological communities. This goal presents a formidable challenge, but emerging systems biology research approaches provide new opportunities to bridge the knowledge gap between molecular- and global-scale phenomena. In this session, new systems-level research approaches will be discussed that incorporate experimentation on model organisms and systems, collection of observational data on communities and ecosystems, and mechanistic modeling of processes ranging from metabolic to global scales.

**Speakers:**

2:00–2:10 pm **Joe Graber, DOE-BER**
**Session Introduction**

2:10–2:35 pm **T.C. Onstott**, Princeton University
**Title:** The Impacts of Global Warming on the Carbon Cycle of Arctic Permafrost

2:35–3:00 pm **Mary Firestone,** University of California Berkeley
**Title:** Carbon Dynamics in Rhizosphere Soil

3:00–3:25 pm **Bill Metcalf**, University of Illinois
**Title:** Systems Biology of Methanosarcinales

3:25–3:45 pm **Break**

3:45–4:10 pm **Don Zak**, University of Michigan
**Title:** Atmospheric N Deposition Slows Litter Decay, Alters Fungal Community Composition, and Suppresses Ligninolytic Gene Expression in a Northern Hardwood Forest

4:10–4:35 pm **Chris Blackwood,** Kent State University
**Title:** Impact of "Cheating" in Microbial Community Carbon Cycle Processes

4:35–5:00 pm **Bob Hettich,** Oak Ridge National Laboratory
**Title:** -Omics Techniques for Functional Analysis of Microbial Carbon Cycling

2:00–5:15 pm **Breakout Session E: Genome Structure and Epigenetics**
**Moderator**: Pablo Rabinowicz

**Description of Session:** Epigenetic phenomena are inheritable genetic changes that occur without DNA sequence alteration. The underlying mechanisms responsible for these epigenetic changes include DNA methylation and hydroxymethylation, multiple covalent histone modifications, interaction between RNA and chromatin, and probably others that remain to be discovered. Epigenetic modifications can be inherited both meiotically and mitotically, and can be altered by environmental effects. Recent advances in our understanding of the molecular basis of epigenetics highlight its critical role in gene regulation and its consequences in all aspects of biology, from development to disease. New genomic and bioinformatics technologies for genome-wide DNA methylation and histone modification analyses are taking this field to a new level. This breakout session will focus on recent advances in epigenomic analyses of different organisms including model plants for bioenergy, as well as current computational tools for epigenomic data analysis and integration.

**Speakers:**

2:00-2:05 pm **Pablo Rabinowicz, DOE-BER**
**Session Introduction**

2:05-2:30 pm **John Dunn**, Brookhaven National Laboratory
**Title**: Fidelity and Dynamics of DNA Methylation in Plants

2:30-2:55 pm **Pamela Green**, University of Delaware
**Title:** Genomic Analysis of miRNAs and target RNAs of *Brachypodium*

2:55-3:20 pm **Matteo Pellegrini,** University of California, Los Angeles
**Title**: DNA Methylation Profiling Using Illumina Sequencers

3:20-3:45 pm **Judith Bender**, Brown University
**Title:** Maintenance of H3 Lysine 9 Methylation in *Arabidopsis*

3:45-4:00 pm **Break**

4:00-4:25 pm **Aleksandar Milosavljevic**, Baylor College of Medicine
**Title:** Epigenome Atlas and Comparative Epigenome Analysis

4:25–4:50 pm **Joseph R. Ecker**, Salk Institute for Biological Studies
**Title:** Linking Genetic, Epigenomic, and Transcriptional Variation in 1,001 *Arabidopsis* Genomes

4:50–5:15 pm **Steven Strauss**, Oregon State University
**Title:** Genome and Developmental Variation in DNA Methylation in Poplar

2:00–5:00 pm **Breakout Session F: Computational Biology – New Mathematical Methods to Model and Understand Biological Processes**
**Moderator:** Susan Gregurick

**Description of Session**: The Computational Biology program, in association with the ASCR SciDAC program, funds research that brings together new mathematical analysis and simulation methods to understand and model biological processes which range from transcription, regulation and metabolism to the fundamental physics of molecular interactions. This session explores the development of new methods to identify RNA binding motifs for transcription regulation, new methods to develop metabolic networks and new methods to analyze the spectroscopic data from proteomics studies. The overall goal of this session is to explore how new methods in mathematical biology can be used to enable our understanding of systems biology.

**Speakers:**

2:00-2:30 pm **Charles Lawrence**, Brown University
**Title:** Computational Identifications of Bacterial RNA Regulatory Motifs Using a Gibbs Sampling Algorithm

2:30–3:00 pm **Dmitry Rodionov**, Burnham Institute
**Title:** Comparative Genomics Approaches for Reconstruction of Transcriptional Regulatory Networks in Bacteria

3:00–3:30 pm **Michael Saunders,** Stanford University
**Title:** Satisfying Flux Balance and Mass-Action Kinetics in a Network of Biochemical Reactions

3:30–3:45 pm **Break**

3:45–4:15 pm **William Cannon**, Pacific Northwest National Laboratory
**Title:** Proteomics of Environmental Samples Using Genome-Based Optimizations and Spectral Library Developments

4:15–4:45 pm **Nagiza Samatova**, North Carolina State University
**Title:** Ultrascale Computational Modeling of Phenotype-Specific Metabolic Processes in Microbial Communities

4:45–5:00 pm **Wrap up and Panel Discussions:** Barriers to Mathematical Modeling of Biological Processes

5:30–8:00 pm **Poster Session**

## Wednesday, April 13

8:30–10:00 am **Plenary Session: Innovative Tools for Genomic Sciences**
**Moderator:** Dean Cole

**Speakers:**

8:30–9:00 am **Bill Moses,** Lawrence Berkeley National Laboratory
**Title:** Radiotracer Imaging and Genomic Science

9:00–9:30 am **Jonathan Sweedler,** University of Illinois, Urbana-Champaign
**Title:** Metabolomics and Chemical Imaging for Probing Cellular Heterogeneity

9:30–10:00 am **Chris Lee**, University of California, Los Angeles
**Title:** Phenotype Sequencing: Identifying the Genes that Cause a Phenotype Directly from Pooled Sequencing of Independent Mutants

10:00–10:30 am **Break**

10:30–11:50am **Plenary Session: Early Career Program**
**Moderator:** Arthur Katz
**Speakers:**

10:30–10:50 am **Matias Kirst**, University of Florida
**Title:** A Systems Biology, Whole-Genome Association Analysis of the Molecular Regulation of Biomass Growth and Composition in *Populus deltoides*

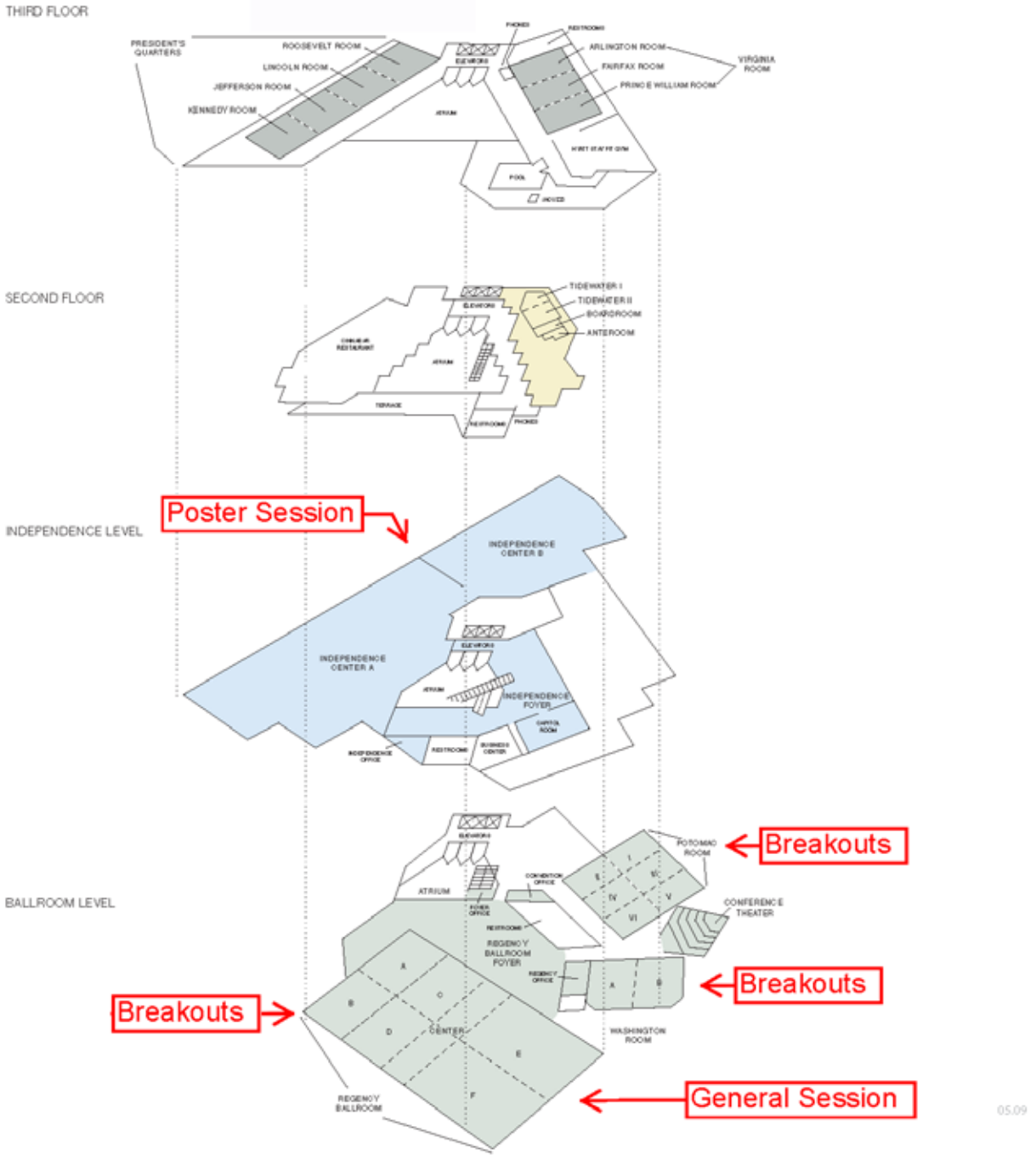
10:50–11:10 am **Victoria Orphan**, California Institute of Technology
**Title:** Insights into the Ecophysiology of Diverse Archaeal-Bacterial Symbioses Mediating the Anaerobic Oxidation of Methane

11:10-11:30 am **Jonathan Schilling**, University of Minnesota
**Title:** Applying the Biology of Brown Rot Fungi to Consolidated Bioprocessing

11:30-12:00 pm **Wei-Jun Qian**, Pacific Northwest National Laboratory
**Title:** Novel Proteomic Approaches for Quantitative Profiling of Important Protein Posttranslational Modifications

12:00 pm **Closeout and Adjournment**

# Hotel Map


THIRD FLOOR
PRESIDENT'S QUARTERS
ROOSEVELT ROOM
LINCOLN ROOM
JEFFERSON ROOM
KENNEDY ROOM
ELEVATORS
PHONES
RESTROOMS
ARLINGTON ROOM
FAIRFAX ROOM
PRINCE WILLIAM ROOM
VIRGINIA ROOM
ATRIUM
POOL
SECOND FLOOR
TIDEWATER I
TIDEWATER II
BOARDROOM
ANTEROOM
TERRACE
INDEPENDENCE LEVEL
Poster Session
INDEPENDENCE CENTER B
INDEPENDENCE CENTER A
INDEPENDENCE FOYER
BALLROOM LEVEL
Breakouts
POTOMAC ROOM
ATRIUM
CONFERENCE THEATER
REGENCY BALLROOM FOYER
Breakouts
Breakouts
CENTER
WASHINGTON ROOM
REGENCY BALLROOM
General Session
05.09


## Hyatt Regency Crystal City

### At Reagan National Airport

DIRECTIONS

From Reagan National Airport (1/2 mile): Follow signs to Crystal City. Take the Rt. 1 South exit and get in the left hand lane. Turn left at first light, 27th Street, the hotel is on the left. From Dulles Airport (32 miles): Take I 66 East to Exit 75 (Rt. 110 South/turns into Rt. 1 South). Proceed on Rt. 1 to 4th light (27th Street). Turn left onto 27th Street. Hotel is on left.

# Abstract Titles

## Joint BioEnergy Institute (JBEI)

## Systems Biology and Metabolic Engineering Approaches for Biological Hydrogen Production

## Computing for Bioenergy

## Biological Systems Research on the Role of Microbial Communities in Carbon Cycling ..... 127

# Workshop Abstracts

## Abstract Organization

Genomic Science program abstracts and posters are organized according to the following research areas important to achieving the program's ultimate scientific goal and objectives. Even-numbered posters will be presented on Monday, odd-numbered posters on Tuesday. Please set up the posters no earlier than 5 p.m. on Sunday and leave them up until noon Wednesday.

**Systems Biology for DOE Energy Missions: Bioenergy**

Bioenergy Research Centers

- BioEnergy Science Center (BESC)
- Great Lakes Bioenergy Research Center (GLBRC)
- Joint BioEnergy Institute (JBEI)

Biofuels: Analytical and Imaging Technologies, Engineering, and Production

Systems Biology and Metabolic Engineering Approaches for Biological Hydrogen Production

Computing for Bioenergy

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

**Joint USDA-DOE Plant Feedstock Genomics for Bioenergy**

**Systems Biology for Environmental and Subsurface Microbiology**

**Biological Systems Research on the Role of Microbial Communities in Carbon Cycling**

**Systems Biology Strategies and Technologies for Understanding Microbes, Plants, and Communities**

Analytical Strategies for the Study of Plants, Microbes, and Microbial Communities

- Biological Systems Interactions
- Plant-Microbe Interfaces
- The Predictive Microbial Biology Consortium (ENIGMA)

Structural Biology, Molecular Interactions, and Protein Complexes

Validation of Genome Sequence Annotation

**Knowledgebase and Computing for Systems Biology**

**Communication and Ethical, Legal, and Societal Issues**

## Genomic Science Goal and Objectives

**Ultimate Scientific Goal**
Achieve a predictive, systems-level understanding of plants, microbes, and biological communities, via integration of fundamental science and technology development, to enable biological solutions to DOE mission challenges in energy, environment, and climate.

**Objective 1:** Determine the genomic properties, molecular and regulatory mechanisms, and resulting functional potential of microbes, plants, and biological communities central to DOE missions.

**Objective 2:** Develop the experimental capabilities and enabling technologies needed to achieve a genome-based, dynamic systems-level understanding of organism and community functions.

**Objective 3:** Develop the knowledgebase, computational infrastructure, and modeling capabilities to advance the understanding, prediction, and manipulation of complex biological systems.

# Systems Biology for DOE Energy Missions: Bioenergy

## Bioenergy Research Centers

# 1

### Inter-BRC Collaboration to Characterize Ultrastructure and Sugar Yields from Three Different Pretreatments of Corn Stover

**Charles E. Wyman**[2,4]* (cewyman@cert.ucr.edu), and Xiadi Gao[3]; Nirmal Uppugundla,[1] Leonardo da Costa Sousa,[1] Shishir P.S. Chundawat,[1] Bruce E. Dale,[1] Venkatesh Balan,[1] Seema Singh,[3] Dean Dibble,[3] Patanjali Varanasi,[3] Gang Cheng,[3] Lan Sun,[3] Chenlin Li,[3] Blake Simmons,[3] **Tim Donohue**[1] (GLBRC PI), **Jay Keasling**[3] (JBEI PI), and **Paul Gilna**[4] (BESC PI)

[1]Great Lakes Bioenergy Research Center, University of Wisconsin, Madison; [2]Bourns College of Engineering, CE-CERT, University of California, Riverside; [3]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, Calif.; and [4]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://glbrc.org
http://jbei.org
http://bioenergycenter.org

**Project Goals: The BioEnergy Science Center (BESC), the Great Lakes Bioenergy Research Center (GLBRC), and the Joint BioEnergy Institute (JBEI) are collaborating to understand how biomass pretreatments with much different deconstruction patterns impact the chemical and ultrastructural features of biomass and its biological conversion to sugars.**

Dilute sulfuric acid (DA), ammonia fiber expansion (AFEX), and ionic liquid (IL) pretreatments are applied to the same source of corn stover by the BESC, GLBRC, and JBEI, respectively. Common sources of cellulase and other accessory enzymes are then employed to release sugars from the solids left after each pretreatment. The GLBRC applies material balances to each overall pretreatment-hydrolysis system to determine the fates of key biomass constituents and also optimizes enzyme formulations for each substrate using their microplate saccharification system. The BESC focuses on characterizing how cellulose accessibility to enzymes, enzyme adsorption and desorption kinetics, and changes in substrate features during hydrolysis vary for the solids resulting from the three pretreatments over the course of enzymatic hydrolysis, especially at low enzyme loadings that are commercially promising. JBEI utilizes various characterization techniques, including SEM, FT-Raman, XRD, NMR, pyrolysis, and high resolution imaging to track ultrastructural and chemical changes resulting from each pretreatment. These results will then be integrated to identify key features influencing enzymatic hydrolysis of solids from these pretreatments and similarities and differences in their impacts on the effectiveness of enzymatic hydrolysis of biomass.

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.

## BioEnergy Science Center (BESC)

# 2

### Integrated Study of *Populus* Tension Stress Response to Understand Mechanisms Underlying Desirable Lignocellulosic Biomass Characteristics

**U.C. Kalluri**[1]* (kalluriudayc@ornl.gov), S. Jawdy,[1] N. Engle,[1] T. Tschaplinski,[1] P. Abraham,[1] R. Adams,[1] R. Hettich,[1] X. Yang, P. Ranjan,[1] R. Sykes,[2] M. Davis,[2] S. Decker,[2] S-Y. Ding,[2] M. Foston,[3] A. Ragauskas,[3] E. Lindquist,[4] Z. Wang,[4] **G. Tuskan**,[1] and **P. Gilna**[1] (BESC PI)

[1]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]National Renewable Energy Laboratory, Golden, Colo.; [3]Georgia Institute of Technology, Atlanta; and [4]Joint Genome Institute, Walnut Creek, Calif.

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). Addressing the roadblock of biomass recalcitrance will require a greater understanding of plant cell walls from synthesis to deconstruction. This understanding would generate models, theories and finally processes that will be used to understand and overcome biomass recalcitrance. This grand challenge calls for an integrated research approach, which is illustrated by**

**the project described below. BESC research is multidisciplinary by design and multi-institutional in composition.**

Plant cell walls with higher cellulose/lower lignin levels, increased wall thickness and increased cell number constitute some desirable characteristics of feedstock materials for biochemical ethanol production. Understanding the underlying molecular, biochemical and phenotypic properties of the cell wall is critical to enabling the design of suitably tailored feedstocks. Tension wood formed in response to tension stress on the outer side of the bent stem of a woody angiosperm is characterized by these desirable feedstock properties. Here we present results from an integrated study of tension stress response in *Populus* stems. Xylem and phloem samples were collected from two genotypes of *Populus* following a 14-day bending stress treatment. For molecular studies, Illumina-based transcriptome profiling, targeted qRT-PCR and LC-MS/MS-based proteome profiling was performed in triplicate runs. For phenotypic characterization, MBMS cell wall compositional analysis, SEM and CARS imaging, and $^{13}C$ cross-polarization magic angle spinning (CP/MAS) NMR, gel permeation chromatography (GPC) were used to characterize the ultrastructure of cellulose, microstructure of wood and the chemical profiles of lignin and cellulose. MBMS revealed that the tension wood samples have higher cellulose and lower lignin levels. This was correlated to the higher relative glucose and lower relative xylose and lignin contents as determined by HPLC carbohydrate and Klason analysis in the tension wood sample when compared to the normal wood samples. The molecular weight of cellulose in the G-layer of tension wood was higher than normal wood and this was accompanied by changes in the H:G:S ratio. Furthermore, CARS microscopy revealed a cell-specific distribution of lignin and metabolic profiling revealed depletions of soluble carbohydrates, and an accumulation of specific phenolic glycosides in the lower lignin samples. To gain a perspective on whether the properties of tension wood translate to reduced recalcitrance of these lignocellulosic materials, sugar-release assays were performed. Molecular studies identified several known and as yet unknown cell wall pathway genes that appear to be significantly differentially up-regulated under tension stress. The present integrated study has shed further light on underpinnings of tension stress response and presented new gene targets for further evaluation in feedstock improvement efforts.

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.

# 3

## Genome-Wide Association Mapping for Characterization of the Genetic Architecture of Recalcitrance and Biomass Productivity in *Populus*

Wellington Muchero[1]* (mucherow@ornl.gov), Lee Gunter,[1] Gancho Slavov,[2] Stephen DiFazio,[2] Ann Wymore,[1] Priya Ranjan,[1] Michael Studer,[4] Jaclyn DeMartini,[4] Charles Wyman,[4] Mark Davis,[3] Robert Sykes,[3] Steve Decker,[3] **Gerald Tuskan,**[1] and **Paul Gilna**[1] (BESC PI)

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]West Virginia University, Morgantown; [3]National Renewable Energy Laboratory, Golden, Colo.; and [4]University of California-Riverside

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). Addressing the roadblock of biomass recalcitrance will require a greater understanding of plant cell walls from synthesis to deconstruction. This understanding would generate models, theories and finally processes that will be used to understand and overcome biomass recalcitrance. This grand challenge calls for an integrated research approach, which is illustrated by the project described below. BESC research is multidisciplinary by design and multi-institutional in composition.**

### Background

*Populus* is an important feedstock for biofuel production that exhibit wide variation in phenotypes targeted for increased biomass productivity and reduced recalcitrance during lignocellulose deconstruction. Such variation in germplasm provides important genetic resources for use in addressing challenges in biofuel production. However, to date only pedigrees derived from interspecific hybridization with limited diversity, genetic marker information and QTL resolution have been used to identify genomic regions harboring genes controlling such traits. The consequence being a dearth in number of genes cloned and lack of knowledge on the molecular mechanisms underlying the expression of economically important traits. The aim of this study was to utilize high-throughput genotyping and an extensive collection of diverse *Populus* genotypes to conduct a linkage disequilibrium (LD)-based association mapping analysis to identify genomic loci associated with economically important traits.

### Methods

Forty-seven extreme phenotypes were selected across measured lignin content and ratio of syringyl and guaiacyl units

(S/G ratio) from a sample set of wood cores representing 1100 individual undomesticated *Populus trichocarpa* trees. Total sugar release through enzymatic hydrolysis alone, combined hot water pretreatment, and enzymatic hydrolysis using a high-throughput screening method were used to characterize phenotypic differences among samples.

High-confidence single nucleotide polymorphisms (SNPs) were identified based on alignment of genome sequence information from 15 re-sequenced *Populus* genotypes. SNPs targeted for use in the association study were chosen based on presence within promoter regions of candidate genes as well as dense coverage of known high-value QTLs. Selected SNPs were used to develop a 34K *Populus* Infinium SNP array with capability to assay 34,132 unique SNP loci. The array was used to characterize 1100 *Populus* genotypes that were clonally propagated and set up in replicated field trials across three geographical locations for phenotypic characterization. Genotypes were initially collected from diverse ecological zones in Pacific Northwest. Samples from one of the field trials in Corvallis, OR were evaluated for recalcitrance based on the sugar release assay, lignin content and syringyl/guaiacyl ratio. In addition, biomass productivity traits including height, diameter, 1st- and 2nd-year growth, crown architecture, and internode length were collected from three replicates for use in marker-trait association.

**Results**

Sugar yield varied greatly among samples with the total amount of glucan and xylan released up to 92% of the theoretical maximum. In fact, several genotypes displayed relatively high sugar release 9up to ca. 64% maximum) under no pretreatment. For S/G ratios <2.0 there was a strong negative correlation between sugar release and lignin content. Interestingly, for S/G ratios ≥2.0, lignin content had a less pronounced effect on the total amount of glucan and xylan released. In all cases, those samples with S/G ratio ≥2.0 released more sugar than those samples with S/G ratio <2.0.

For SNP analysis, 31,769 SNPs, representing 93% of the array, were successfully genotyped on 1100 *Populus* genotypes with no-call rates <10%. Among these 29,594 (87%) were informative and 23,305 (68%) had minor allele frequencies ≥0.1. At the phenotypic level, substantial variation, reproducible across replicates, was observed on all recalcitrance and productivity traits suggesting adequate genetic variability for successful marker-trait association. This cumulative data is being used to conduct LD-based gene cloning and/or QTL fine mapping for application genetic improvement of *Populus*.

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.

# 4

## High-Throughput Development and Characterization of Cell Wall Cisgenic Mutants in *Populus*

Xiaohan Yang[1]* (yangx@ornl.gov), **Gerald A. Tuskan,[1]** Lee E. Gunter,[1] Priya Ranjan,[1] Chu-Yu Ye,[1] Udaya C. Kalluri,[1] Sara Jawdy,[1] Timothy J. Tschaplinski,[1] Zackary T. Moore,[1] Shujun Chang,[2] **Maud Hinchee,[2]** Will Rottmann,[2] Kim Winkeler,[2] Mark Davis,[3] and **Paul Gilna[1]** (BESC PI)

[1]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]ArborGen LLC, Summerville, S.C.; and [3]National Renewable Energy Laboratory, Golden, Colo.

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.**

The purpose of the *Populus* Transformation effort is to identify all candidate genes related to the construction of primary and secondary plant cell walls and to validate the function and utility of candidate cell wall biosynthesis genes via transformation and regeneration of cisgenic *Populus*.

Regulatory features controlling biomass recalcitrance may originate within the nucleus at the transcription level or within organelle and/or membrane-associated proteins or at the point of polymerization, or at all three. As such, based on our expression data, information in published literature and a general knowledge of *Populus* growth and development, we targeted all three putative regulatory arenas for the transformation pipeline for the first two years of the project and have scrutinized thousands of potential gene candidates. Our current focus is on candidates from gene discovery tasks, including QTL mapping, association studies, and expression studies. We submitted our first-round of candidate genes for transformation to the BESC Transformation Pipeline Committee in April 2008. Since then, we have submitted gene candidates in eight additional rounds, resulting in a total of 288 approved constructs in rounds 1-9.

Once a gene is approved by the committee primers are designed at Oak Ridge National Laboratory (ORNL) for

creating over-expression or knockdown constructs. The primers are then used to amplify the target gene fragments by PCR at ArborGen from cDNA libraries made at ORNL from leaf and xylem tissue in *P. trichocarpa* clone '93-968'. The amplified gene fragments are cloned into ArborGen's transformation vectors and transformed into a *P. deltoides* clone that exhibits good growth characteristics, disease/pest resistance and ease of propagation through vegetative cuttings. The *Populus* Transformation effort is fully integrated with the BESC LIMS database from candidate gene selection to primer design to cisgenic delivery to phenotypic data collection.

To date, 8,822 plants representing 1,240 transformed/control lines and 68 constructs have been delivered to ORNL. Plantlets delivered in tissue culture media have been transplanted to soilless media and propagated in the greenhouses for phenotypic and genotypic analysis related to the goals stated above. Leaf, stem and root tissues collected from the first three sets of cisgenics (18 constructs and 1382 plants) are currently being phenotyped at ORNL and archived while cell wall biomass is assayed for lignin content and syringyl/guaiacyl monomer ratios, as well as for glucose and xylose release by enzymatic hydrolysis at National Renewable Energy Laboratory (NREL). Such data is being used to predict the impact of each gene of interest on fermentation efficiencies.

Moreover, seven constructs representing 87 transgenic lines (522 trees) have been planted in the stool beds at ArborGen. Additional plants from 16 constructs (461 lines and about 3,500 trees) will be planted in ArborGen stool beds in spring 2011. Finally, approximately 180 constructs (3,556 lines and more than 40,000 plants) are in tissue culture awaiting regulatory permit approval for transfer to the University of Georgia stool bed facility.

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.

# 5

## Investigation of the Fate of Lignin Structures of Poplar and Switchgrass During Various Pretreatments to Understand its Impact to Biomass Recalcitrance

Y. Pu,[1,3] S. Cao,[1,3] R. Samuel,[1,3] N. Jaing,[1,3] M. Foston,[1,3] M. Studer,[2,3] **C. Wyman**,[2,3] **A.J. Ragauskas**[1,3]* (Arthur.ragauskas@chemistry.gatech.edu), and **P. Gilna**[3] (BESC PI)

[1]Georgia Institute of Technology, Atlanta; [2]CE-CERT, Chemical and Environmental Engineering Department, Bourns College of Engineering, University of California, Riverside; and [3]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

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 researchers in characterization, modeling, and data management areas are engaged in 1) applying advanced technologies to analyze chemical and structural changes within biomass and 2) storing, tracking, analyzing and integrating data and understanding across the center.**

The utilization of lignocellulosic biomass for biofuels production is considered as one of the most promising alternative sources of renewable energy, providing a sustainable option for fuels production in an environmentally compatible manner. The biological conversion of biomass to biofuels typically consists of three main steps including pretreatment, enzymatic hydrolysis and fermentation. The effective utilization of biomass is largely predicated on pretreatment technologies that can reduce biomass recalcitrance. The objective of pretreating lignocellulosics is to alter the structure of plant cell wall lignin and polysaccharides, thereby rendering the polysaccharides more accessible/amenable to hydrolytic enzymes. In this study, we present our results of characterizing lignin structures of poplar and switchgrass before and after various pretreatments including steam, dilute acid and aqueous lime. For poplar, ball-milled lignin was isolated after pretreatment and characterized using gel permeation chromatography (GPC) and nuclear magnetic resonance (NMR) techniques. For switchgrass, the impact of pretreatment was evaluated by using whole cell dissolution solvent system of perdeuterated pyridinium chloride/DMSO coupled with $^{13}C$-$^{1}H$ HSQC NMR analysis. The aryl-*O*-ether linkage (*β*-*O*-4) of poplar lignin was extensively cleaved and lignin condensation occurred during acid pretreatment. The poplar lignin was also observed to have a decreased carboxylic group and methoxyl group content and these changes were accompanied with increases in condensed lignin. Steam pretreatment of switchgrass resulted in a slight degradation in lignin. Major structural changes were observed due to the lime pretreatment of switchgrass including substantial reduction in the *β*-ether linkages, along with decrease of other lignin subunits. The lignin structural changes were examined from a perspective of future improvements in biomass pretreatment for enhanced sugar production for biofuels.

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.

# 6

## Studies with *Arabidopsis* Mutants Reveal that Xylan Sidechain Substitution Affects Recalcitrance

Breeanna Urbanowicz,[1,2] Maria Peña,[1,2] Jason Backe,[1,2] Supriya Ratnaparkhe,[1,2] Utku Avci,[1,2] **Malcolm O'Neill**,[1,2] **William S. York**[1,2]* (will@ccrc.uga.edu), and **Paul Gilna**[2] (BESC PI)

[1]University of Georgia, Complex Carbohydrate Research Center, Athens; and [2]BioEnergy Science Center, Oak Ridge, Tenn.

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.**

The thick recalcitrant secondary cell walls of woody tissues are principally composed of cellulose, glucuronoxylan (GX), and lignin. The GX present in hardwoods including poplar and in mature stems of the model plant *Arabidopsis thaliana* have 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). *Arabidopsis* GX has approximately one uronic acid residue for every eight xylosyl residues and a GlcA:MeGlcA ratio of 1 to 3. We are using *Arabidopsis* as a model system to identify genes involved in xylan biosynthesis and modification, as the identities and functions of many of the enzymes involved in this process are poorly understood. The information gained with *Arabidopsis* and with model monocot species will provide a solid basis for decreasing recalcitrance by targeted modification of wall composition and architecture in biofuel crops including poplar and switchgrass.

We have identified a series of *Arabidopsis* plants carrying mutations in genes predicted by *in silico* expression analysis to be highly expressed during secondary cell wall formation. The effects of these mutations on the structure of secondary cell wall xylan and cell wall recalcitrance are under investigation. Here we describe our studies of a gene which belongs to a ten member family in *Arabidopsis* and encodes a protein containing a Domain of Unknown Function (DUF) 579. Plants carrying a mutation in this gene are phenotypically indistinguishable from wild-type plants. However, these mutant plants have walls with altered recalcitrance and synthesize xylan in which the degree of GlcA methylation is reduced to ~20% of wild-type levels. These data, when taken together with our results obtained using other *Arabidopsis* xylan mutants, have revealed that biomass recalcitrance is correlated with altered xylan sidechain substitution.

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.

# 7

## A Novel Method for Assigning Holes in Metabolic Pathways with Predicted Enzyme-Encoding Genes Based on Genomic Locations

Fengfeng Zhou[1]* (ffzhou@csbl.bmb.uga.edu), Xizeng Mao,[1] Mark Schell,[2] Yanbin Yin,[1] **Ying Xu**,[1] and **Paul Gilna**[3] (BESC PI)

[1]Computational Systems Biology Lab, Department of Biochemistry and Molecular Biology; [2]Department of Microbiology, University of Georgia-Athens; and [3]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

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 researchers in characterization, modeling, and data management areas are engaged in 1) applying advanced technologies to analyze chemical and structural changes within biomass and 2) storing, tracking, analyzing and integrating data and understanding across the center.**

We have previously discovered that genomic locations of genes in bacteria are tightly constrained by the pathways in which they participate; and the more pathways a gene is involved in, the smaller the range ("*comfort zone*") of the genomic locations the gene can possibly have.[1] Overall each bacterial genome has evolved to have all its genes arranged in genomic locations so that all genes are in their "comfort zones". This discovery laid the foundation for development of a suite of new tools for solving a number of very challenging genome analysis problems, such as (1) assigning genes to holes in partially elucidated metabolic pathways and (2) assembly of large contigs into bacterial genomes. Basically we discovered a new information source and a new tool, the global arrangement of genes, to study bacterial genomes.

Filling holes in partially elucidated metabolic pathways represents a very important and highly challenging problem.

Current techniques generally rely on possible association relationships among genes to suggest candidates to fill the pathway holes on based their pathway neighbors with assigned genes. Here we present a novel way to fill the pathway holes based on the predicted enzyme-encoding genes, their genomic locations and the newly discovered pathway-gene location relationships. Using this method, we have assigned over a hundred holes in *E. coli* KEGG pathways. Our simulation studies (through removing portions of the assigned KEGG pathways and treating them as holes) indicate ~50% of the predicted gene-assignment are highly reliable. We anticipate that this method will become a widely used technique for assigning genes to specific pathways among numerous other applications.

**Reference**

1. YB Yin, H Zhang, V Olman and Y Xu, Genomic arrangement of bacterial operons is constrained by biological pathways encoded in the genome, *Proc Natl Acad Sci USA*, 107 (14), 6310 – 6315, 2010.

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

# 8

## Glycome Profiling Using Glycan-Directed Monoclonal Antibodies: Multiple Applications to Plant Cell Wall/Biomass Characterization

Sivakumar Pattathil[1,2]* (siva@ccrc.uga.edu), Heather McCormick,[1,2] Jeffrey Miller,[1,2] Virginia Brown,[1,2] Jaclyn D. DeMartini,[1,3] Irina Kataeva,[1,4] Ajaya Biswal,[1,2] Supriya Ratnaparkhe,[1,2] Trina Saffold,[1,2] Malcolm O'Neill,[1,2] Debra Mohnen,[1,2] William S. York,[1,2] Charles E. Wyman,[1,3] Michael W.W. Adams,[1,4] Huanzhong Wang,[1,5] Fang Chen,[1,5] Richard A. Dixon,[1,5] **Michael G. Hahn**,[1,2,6] and **Paul Gilna**[1] (BESC PI)

[1]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Complex Carbohydrate Research Center, The University of Georgia-Athens; [3]Bourns College of Engineering, CE-CERT, University of California, Riverside; [4]Department of Biochemistry, The University of Georgia-Athens; [5]Samuel Roberts Noble Foundation, Ardmore, Oklahoma; and [6]Department of Plant Biology, The University of Georgia, Athens

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). Addressing the roadblock of biomass recalcitrance will require a greater understanding of plant cell walls from synthesis to deconstruction. This understanding would generate models, theories and finally processes that will be used to understand and overcome biomass recalcitrance. This grand challenge calls for an integrated research approach, which is illustrated by the project described below. BESC research is multidisciplinary by design and multi-institutional in composition.**

The complexity of various types and forms of plant cell walls makes a detailed understanding of their overall structure and biosynthesis a major challenge. As a result, one of the focus areas of cell wall research has been the development of advanced tools that enable easy, rapid, high-throughput and reliable characterization of cell walls. Cell wall directed monoclonal antibodies have now emerged as very powerful probes to assist in the detailed structural analysis of plant cell walls. The worldwide collection of plant cell wall glycan-directed monoclonal antibodies is now sufficiently large and comprehensive to be used to monitor the majority of cell wall glycans. Here we report an ELISA-based method that employs a toolkit of ~150 monoclonal antibodies (representing most major classes of plant polysaccharides) for characterizing diverse plant cell walls. The method involves sequential extraction of cell wall materials with increasingly harsh reagents to obtain fractions enriched in diverse wall-polysaccharides, such as pectins, arabinogalactans, xylans, xyloglucans, and mannans. Each of these fractions is subjected to ELISA against the toolkit of antibodies and the binding responses are depicted as heat maps. Our studies show that glycome profiling is highly reproducible and has the potential to pinpoint structural differences in cell walls from various plant tissues and to monitor cell wall changes that occur during development and differentiation. Glycome profiling has also proven useful for comparative glycomics of diverse plants. Studies of mutants in pectin and lignin biosynthetic pathways using glycome profiling, revealed cell wall structural alterations in these plants that pertain to their recalcitrance. Lastly, glycome profiling can be used to monitor cell wall structural alterations in plant biomass samples undergoing different pretreatment regimes and microbial degradations.

This work is supported by a grant from U.S. Department of Energy BioEnergy Science Center and grant DE-PS02-06ER64304. 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. Generation of the monoclonal antibodies used in this work was supported by the NSF Plant Genome Program (DBI-0421683).

# 9

## Understanding Cellulose Structure

Michael F. Crowley[1]* (Michael.Crowley@nrel.gov), Marcus Foston,[2] James F. Matthews,[1] John Brady,[3] **Michael E. Himmel**,[1] **Arthur J. Ragauskas**,[2] and **Paul Gilna**[4] (BESC PI)

[1]Biosciences, National Renewable Energy Laboratory, Golden, Colo.; [2]Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta; [3]Cornell University, Ithaca, N.Y.; and [4]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

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). Addressing the roadblock of biomass recalcitrance will require a greater understanding of plant cell walls from synthesis to deconstruction. This understanding would generate models, theories and finally processes that will be used to understand and overcome biomass recalcitrance. This grand challenge calls for an integrated research approach, which is illustrated by the project described below. BESC research is multidisciplinary by design and multi-institutional in composition.**

Understanding the origins of biomass recalcitrance and the structural changes that occur during biosynthesis and deconstruction are vital for improving current processing and conversion methods for biofuel production. Efficient enzymatic hydrolysis of lignocellulose has been directly related to cellulase enzyme activity and the potential effect of cellulose characteristics, such as crystallinity. Native, pretreated, enzymatically, and microbially deconstructed biomass samples were subjected to $^{13}C$ solid state NMR techniques designed to probe the ultrastructure of cellulose. We have also modeled cellulose microfibrils using molecular mechanics and dynamics at several temperatures to determine the dependence of its crystallinity on shape and temperature; as well as the features of a microfibril from its core to its surface. New ideas regarding hydrogen bonding patterns and microfibril twist resulted from these modeling and NMR studies.

This work was supported by DOE BER BioEnergy Science Center (cellulose structure by NMR) and by the DOE ASCR SciDAC program (cellulose modeling)
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.

# 10

## Cell Wall Recalcitrance Traits in Grasses Down-Regulated for BESC Target Genes by Virus-Induced Gene Silencing (VIGS)

Hema Ramanna[1]* (hramanna@noble.org), Crissa Doeppke,[2] Ivana Gelineo-Albersheim,[3] Sakae Hisano,[1] Angela Ziebell,[2] Steve Decker,[2] Cindy Crane,[1] Mark Davis,[2] Udaya Kalluri,[4] Debra Mohnen,[3] **Richard Nelson**,[1] and **Paul Gilna**[4] (BESC PI)

[1]The Samuel Roberts Noble Foundation, Ardmore, Okla.; [2]National Renewable Energy Laboratory, Golden, Colo.; [3]University of Georgia-Athens, Complex Carbohydrate Research Center, Athens, Ga.; and [4]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

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). Addressing the roadblock of biomass recalcitrance will require a greater understanding of plant cell walls from synthesis to deconstruction. This understanding would generate models, theories and finally processes that will be used to understand and overcome biomass recalcitrance. This grand challenge calls for an integrated research approach, which is illustrated by the project described below. BESC research is multidisciplinary by design and multi-institutional in composition.**

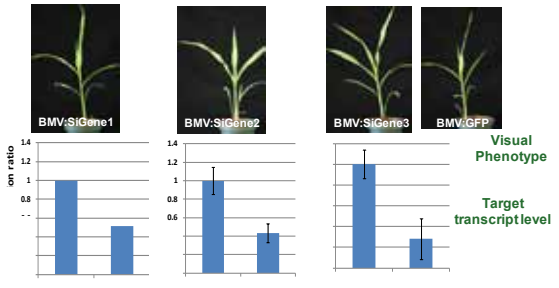
| | |
|---|---|
| Number of gene fragments cloned into VIGS vector | 51 |
| Number of genes analyzed by VIGS | 37 |
| Number of genes showing silencing | 18 |
| Number of genes analyzed for sugar release after VIGS study | 4 |
| Number of remaining genes with NREL for sugar release analysis | 14 |

**Fig. 1: BESC VIGS progress to date**

Perennial grasses (i.e. switchgrass) represent cellulosic biomass with great potential as a high energy and renewable fuel source. The recalcitrance of this biomass to efficient ethanol conversion, however, must be overcome if it is to fulfill this potential. In addition, switchgrass transformation is a time consuming process that slows our ability to evaluate ways to improve this biomass for conversion. Therefore to speed the process BESC has pursued a reverse genetics

RNA silencing technology, virus-induced gene silencing (VIGS), with the closely related grass, foxtail millet, to rapidly identify genes whose modified expression leads to enhanced sugar release. Researchers from many institutions within the center are involved in this work, crossing multiple focus areas. Our findings will allow the rapid identification of genes to focus on in our switchgrass and *Populus* transformation pipelines for enhanced biofuel production.

VIGS is a transient RNA silencing method that provides very rapid results: target gene knockdown and analysis of silenced tissue within two months of inoculation. A *Brome mosaic virus* (BMV) clone was modified to serve as the vector to express plant gene fragments targeted for silencing in our VIGS analysis pipeline. Thirty-seven genes within the VIGS pipeline have been inoculated to plants and target transcript levels determined (~50% of the total submitted to the pipeline: Figure 1). The BMV vector induced significant silencing of 18 target gene transcripts (49% of those inoculated) as determined by quantitative RT-PCR. Multiple target transcripts were silenced more than 70% compared with control tissue and one target transcript was silenced greater than 90%. The visible and cell wall biochemical phenotypes of the silenced tissue from inoculated plants are being evaluated and compared with baseline values observed from plants inoculated with a BMV vector expressing a fragment of GFP, a sequence that would not induce silencing in the plants. Of the silenced target transcripts whose tissue has been analyzed for modified recalcitrance, three had no change in sugar release (glucose or xylose) while one had a 20% increase in total sugar release, entirely accounted for by an increase in glucose levels (Figure 2 and data not shown). Target gene transcript levels and sugar release values will be reported for additional targeted "recalcitrance" genes. This procedure shows promise for rapidly identifying genes whose down-regulation will result in enhanced ethanol production.




| Treatment | Glucose (g/g bm) | Xylose (g/g bm) | Total sugar | Lignin S/G Ratio |
|---|---|---|---|---|
| BMV:GFP | 0.23a ± 0.01 | 0.041 ± 0.005 | 0.27a ± 0.01 | 0.73 ± 0.03 |
| BMV:SiGene1 | 0.23a ± 0.02 | 0.042 ± 0.004 | 0.28a ± 0.02 | 0.72 ± 0.05 |
| BMV:SiGene2 | 0.28b ± 0.05 | 0.041 ± 0.010 | 0.32b ± 0.04 | 0.68 ± 0.06 |
| BMV:SiGene3 | 0.22a ± 0.01 | 0.043 ± 0.005 | 0.26a ± 0.02 | 0.69 ± 0.04 |

Sugar release and lignin S/G values in silenced plants

ANOVA for glucose levels: $LSD_{0.05}$ level=0.04
ANOVA for total sugar: $LSD_{0.05}$ level=0.03

Fig.2 Silencing wall-modulating genes through VIGS and wall characteristics

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.

# 11

## Strategy to Identify Plant Genes that Affect Biomass Recalcitrance

Ajaya K. Biswal,[1,2] Sivakumar Pattathil,[1,2] Michael Hahn,[1,2,4] Mark Davis,[2,5] Steve Decker,[2,5] Robert Sykes,[2,5] Michael Selig,[2,5] Irina Kataeva,[2,3] Mike Adams,[2,3] **Debra Mohnen[1,2,3]*** (dmohnen@ccrc.uga.edu), and **Paul Gilna[1]** (BESC PI)

[1]Complex Carbohydrate Research Center, University of Georgia, Athens; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]Department of Biochemistry and Molecular Biology, University of Georgia, Athens; [4]Department of Plant Biology, University of Georgia, Athens; and [5]National Renewable Energy Laboratory, National Bioenergy Center, Golden, Colo.

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.**

Plants have evolved increasingly complex cell walls to withstand biotic and abiotic stress. Current efforts to use plant cell wall biomass for biofuel production are thwarted by the resistance of plant cell walls to facile deconstruction by enzymes and microbes, thus necessitating expensive pretreatments. Here we describe the use of multiple strategies to identify putative recalcitrance genes. The identification of such genes will provide information on the structural underpinnings of plant cell wall biomass recalcitrance to deconstruction and identify plant cell wall biosynthetic proteins whose modified expression may reduce biomass recalcitrance. We investigated the possible role of a putative glycosyltransferase (GT14) as a recalcitrance gene in *Arabidopsis*. GT14 was shown to be expressed in all tissues with high expression in stems. Two independent homozygous mutant Salk lines (named as gt*14-1* and gt*14-2*) were identified. Five week old gt*14-1* and *gt4-2* mutants had reduced rosette leaf size and stem length. Interestingly, within the subsequent two weeks both mutants caught up to WT in their growth phenotype. ELISA analysis of cell walls and 1M KOH- and 4M KOH-cell wall extracts from wild type (WT) versus *gt14-1* and *gt14-2* mutants using cell wall-directed monoclonal antibodies revealed differential binding to mutant versus WT walls and suggested more

easily extractable walls in *gt14* mutants. Standard pretreatment and enzyme hydrolysis sugar release assays, however, revealed no change in glucose/xylose release in *gt14* mutants compared to *Arabidopsis* WT. Conversely, two bacteria, Cbes (*Caldicellulosiruptor bescii*) and Csac (*C. saccharolyticus),* both grew more efficiently on the *gt14* mutant biomass than on WT biomass. A comparison of the cell wall glycome profile of WT, *gt14-1* and *gt14-2* biomass during the bacterial growth and biomass deconstruction process showed that both Cbes and Csac were able to grow on, and thus deconstruct, more polysaccharides in *gt14* mutant biomass than in WT. Taken together the results indicate that GT14 is a recalcitrance target and that different assay systems provide unique information about the deconstruction process. Bioenergy crop plants with modified GT14 expression are being generated to determine the effects of modified GT14 expression on the recalcitrance of these species.

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.

# 12

## Towards Identification of Plant Genes Involved in Regulation and Synthesis of Recalcitrant Acetylated-Polysaccharides

James Smith[1,2*] (jamess@uga.edu), Miles Ingwers,[1,2] Viviana Martinez,[1,2] **Maor Bar-Peled,[1,2]** and **Paul Gilna[1]** (BESC PI)

[1]Complex Carbohydrate Research Center, University of Georgia-Athens; and [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

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.**

Acetylated glucomannan (Ac-GlcM) is a common glycan in the secondary wall of wood, and considered a recalcitrance polysaccharide. While enzymes and genes involved in the backbone synthesis of GlcM are known, the addition of acetyl groups to the structure remains elusive. We use *Amorphophallus rivieri* corm tissue as a model system to determine acetylation of GlcM. Here we provide data for the degree of acetylation and of GlcM polymerization as a function of its deposition within the corm.

*A. rivieri* corm was sectioned into cubes and GlcM and Ac-GlcM were extracted and used to evaluate the distribution, structure, and acetylation profile throughout the corm tissue. Recombinant mannanases were used to hydrolyze GlcM. The resulting heterogenic populations of shorter-chain Ac-GlcM oligomers were structurally characterized by MALDI-TOF MS, HPLC and $^{1}$H-NMR.

To determine the nature of proteins that accumulate in Ac-GlcM tissues, polypeptides were isolated from corm regions, separated by SDS-PAGE, peptides were gel extracted, chromatographed on C-18, and their sequences were obtained using MS/MS.

In addition, total RNA from specific corm regions was isolated, converted to cDNA and submitted for sequencing. The combined "Ac-GlcM biochemical mapping," proteomic, EST analysis of various tissues, gene clustering methodologies, and computational analyses have already identified new genes involved in Ac-GlcM synthesis and regulation. Many of these genes are annotated to contain different types of DUFs (protein of unknown function with conserved motif).

The collective approach along with genetic manipulation of these genes will be used to reveal the role secondary wall acetylation plays in GlcM metabolism and recalcitrance in *populus* and other plants that are considered as feedstock for the biofuel industry.

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.

# 13

## Unprecedented Advancements in Switchgrass Biotechnology for Improved Biomass Quality Towards Decreasing Recalcitrance

D.G.J. Mann[1*] (dmann1@utk.edu), L.L. Abercrombie,[1] H. Baxter,[1] J.N. Burris,[1] F. Chen,[2] C. Fu,[3] F. Hardin,[3] J.S. Hawkins,[4] T. Hernandez,[2] H. Hisano,[3] Z. King,[5] P.R. LaFayette,[5] Y. Lee,[2] M. Mazarei,[1] R.J. Percifield,[4] C. Poovaiah,[1] M. Rudis,[1] H. Shen,[2] I. Torres-Jerez,[2] W. Wuddineh,[1] X. Xiao,[3] J. Zhang,[2] **J.L. Bennetzen,[4] R.A. Dixon,[2] W.A. Parrott,[5] M.K. Udvardi,[2] Z.-Y. Wang,[3] C.N. Stewart Jr.,**[1] and **P. Gilna[6]** (BESC PI)

[1]Department of Plant Sciences, University of Tennessee, Knoxville; [2]Plant Biology Division, The Samuel Roberts Noble Foundation, Ardmore, Okla.; [3]Forage Improvement Division, The Samuel Roberts Noble Foundation, Ardmore, Okla.; [4]Department of Genetics, University of Georgia, Athens; [5]Center for Applied

Genetic Technologies, University of Georgia, Athens; and [6]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

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.**

Switchgrass (*Panicum virgatum*) is a leading candidate feedstock for biofuels in the United States and is a crucial model feedstock component of the BioEnergy Science Center (BESC). Biotechnology of switchgrass is important for screening potential cell wall biosynthesis genes with the ultimate goal of generating designer feedstocks with improved traits for the biorefinery. This work is being performed multi-collaboratively among six laboratories in three institutions within BESC. Within our lab, biotechnological approaches towards enhancing switchgrass traits include improving tissue culture and transformation systems, isolating novel switchgrass promoters for expression of transgenes, developing a versatile DNA vector set for genetic transformation, and subsequently altering lignin, cellulose and hemicellulose content within switchgrass towards decreasing the recalcitrance of lignocellulose. Switchgrass cell suspension cultures have been produced and characterized for mutant selection, mass propagation, gene transfer experiments via protoplast isolation and cell biology in view of cell wall trait assessment. In order to coordinate switchgrass transformation within BESC and to facilitate rapid screening of genes of interest, we have developed a Gateway-compatible monocot transformation vector set (pANIC) for overexpression and RNAi-mediated knockdown with visual and selectable markers. Overall, BESC has facilitated the coordination of scientific expertise and research in switchgrass biotechnology that would have been otherwise impossible by one investigator with funding under a traditional grant.

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.

# 14

## Development of Resources for Switchgrass Functional Genomics

Ji-Yi Zhang,[1,7] Yi-Ching Lee,[1,7] Ivone Torres-Jerez,[1] Mingyi Wang,[1] Ji He,[1] Christa Pennacchio,[3] Erika Lindquist,[3] Yanbin Yin,[4,7] Wen-Chi Chou,[4,7] Hui Shen,[1,7] Ying Xu,[4,7] Jane Grimwood,[5] Jeremy Schmutz,[5] Laura E. Bartley,[6] Pamela Ronald,[6] Malay Saha,[2,7] Richard Dixon,[1,7] **Yuhong Tang***[1,7] (ytang@noble.org), **Michael Udvardi**,[1,7] and **Paul Gilna**[7] (BESC PI)

[1]Plant Biology Division; [2]Forage Improvement Division, The Samuel Roberts Noble Foundation, Ardmore, Okla.; [3]DOE Joint Genome Institute, Walnut Creek, Calif.; [4]Department of Biochemistry and Molecular Biology, University of Georgia-Athens; [5]HudsonAlpha Genome Sequencing Center, Huntsville, Ala.; [6]Plant Pathology Department, University of California-Davis; and [7]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

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.**

Switchgrass (*P. virgatum L.*) is a perennial C4 grass native to North America. It has been used as forage and for soil conservation and has the potential to become a major source of biomass for biofuel production. To realize this potential, breeding and genetic engineering efforts are underway to improve existing germplasm. As the first step towards developing a set of functional genomics resources that are essential for gene discovery, basic biology research, and molecular breeding efforts, large numbers of expressed sequence tags (EST) have been generated for two tetraploid switchgrass genotypes, AP13 a lowland "Alamo" genotype and VS16 a genotype of upland "Summer". In addition to over 11.5 million high quality ESTs generated by 454/Roche pyrosequencing technology, three full-length enriched cDNA libraries were constructed with RNA from multiple AP13 tissues grown under optimal and stress conditions. About 100,000 clones were sequenced from both ends with the Sanger method and over 69,000 high quality reads were produced. To optimize sequence assembly strategies, different programs including the classical CAP3 were tested,

and a two-stage approach was finally selected to assemble AP13 uni-transcripts. First, 454 ESTs were assembled into 102,000 isotig/contigs using the Newbler program with stringent parameters (overlap 100 bp and identity at 99%). PAVE was then used to assemble Sanger reads and the processed 454 isotig/contigs into ~80,000 unique transcript sequences. Separately, the VS16 454 ESTs were assembled into ~34,000 isotig/contigs using Newbler with the same parameters. To create a switchgrass gene index for gene annotation and Affymetrix cDNA chip design, a total of 545,000 Sanger ESTs of other genotypes in the public domain were downloaded, grouped, and assembled using the PAVE program. A final 132,000 unigene set (PviUT1.2) was generated from existing ESTs with priority order of AP13, Alamo, Kanlow, VS16, and other sequences including about 1502 virtual transcripts predicted from AP13 BAC sequences. The Affymetrix cDNA microarray chip (Pvi_cDNAa520831) based on PviUT1.2 contains ~122,400 probe sets. This chip has an 11µm feature size, with 11 probes for each transcript without mismatch probes. These represent 104,871 switchgrass unitranscript sequences with one or two probe sets. The chip is available to the public through Affymetrix Inc. A switchgrass gene expression atlas is being generated with this platform. The sequence resources will be used for gene annotation, prediction of transcription factor and other gene families of interest, and SNP identification. All switchgrass ESTs generated by this project and the assembled unigene set PviUT1.2 have been deposited to the Switchgrass Genomics database hosted by the Noble Foundation and accessible through this web link: http://switchgrassgenomics.noble.org

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.

# 15

## Redesigning Lignocellulosic Feedstocks: Genetic Modification of COMT in Switchgrass Significantly Reduces Recalcitrance and Improves Ethanol Production

Chunxiang Fu,[1a] Jonathan R. Mielenz,[3a] Xirong Xiao,[1,3] Yaxin Ge,[1] Choo Y. Hamilton,[3] Miguel Rodriguez Jr.,[3] Fang Chen,[2,3] Marcus Foston,[3,4] Art J. Ragauskas,[3,4] Joseph Bouton,[1] **Richard A. Dixon,**[2,3] **Zeng-Yu Wang**[1,3]* (zywang@noble.org), and **Paul Gilna** (BESC PI)

[1]Forage Improvement Division, The Samuel Roberts Noble Foundation, Ardmore, Okla.; [2]Plant Biology Division, The Samuel Roberts Noble Foundation, Ardmore, Okla.; [3]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [4]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta

[a]C. Fu and J. Mielenz 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 multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). Addressing the roadblock of biomass recalcitrance will require a greater understanding of plant cell walls from synthesis to deconstruction. This understanding would generate models, theories and finally processes that will be used to understand and overcome biomass recalcitrance. This grand challenge calls for an integrated research approach, which is illustrated by the project described below. BESC research is multidisciplinary by design and multi-institutional in composition.**

Switchgrass is a leading dedicated bioenergy feedstock in the United States because it is a native, high yielding, perennial prairie grass with broad cultivation range and low agronomic input requirements. Biomass conversion research has developed processes for production of ethanol and other biofuels but they remain costly primarily due to the intrinsic recalcitrance of biomass. We show here that switchgrass genetic modification can produce phenotypically normal plants that have reduced thermal-chemical (≤180°C), enzymatic and microbial recalcitrance. Downregulation of the switchgrass caffeic acid *O*-methyltransferase (COMT) gene decreases lignin content modestly, reduces the syringyl to guaiacyl lignin monomer ratio, improves sugar release and more importantly, increases the ethanol yield by up to 38% using conventional biomass fermentation processes. In addition to increased ethanol production, the transgenic switchgrass also showed increased forage quality, which is very beneficial for farmers since switchgrass can serve as a dual purpose (bioenergy/forage) crop. The reduced lignin content has minor or negligible impact on cellulose content or structure. The only phenotypic change observed between the control and COMT down-regulated lines was the brownish to reddish color in the basal internode and its cross sections in the severely downregulated lines. This trait can be used as a phenotypic maker during breeding and selection process.

The downregulated lines require reduced pretreatment severity and 300-400% lower cellulase dosages for equivalent product yields using simultaneous saccharification and fermentation with yeast. The increased susceptibility of the transgenic lines to commercial cellulases is not a result of a specific pretreatment condition because reducing the severity of the pretreatment of both non-transgenic and transgenic switchgrass did not affect the improved susceptibility of the transgenics. Furthermore, fermentation of diluted acid pretreated transgenic switchgrass using *Clostridium thermocellum* with no added enzymes showed superior product yields compared to unmodified switchgrass.

Regardless of the processes (with or without pretreatment, various pretreatment conditions, different enzyme dosages, fermentation by a CBP microorganism) used, and irrespective of the materials (stem vs. whole tiller) analyzed, the transgenics consistently showed significantly improved fermentation yields. Thus, these switchgrass lines can improve

the economic viability of various bio-based fermentation-derived fuels and chemicals by greatly improving the energy, cost, and land-use efficiency of their production. The innovative transgenic switchgrass with superior processing properties illustrates the feasibility and potential of developing energy crops specifically designed for industrial processing to biofuel.

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.

# 16

## Integrated Computational Biology Capabilities at BESC

Jeremy Smith,[1,2] Ed Uberbacher[2]* (uberbacherec@ornl.gov), Eberhard Voit,[3] **Ying Xu**,[4] Brian Davison,[2] and **Paul Gilna**[2] (BESC PI)

[1]The University of Tennessee, Knoxville; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]Georgia Institute of Technology, Atlanta; and [4]University of Georgia-Athens

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). Addressing the roadblock of biomass recalcitrance will require a greater understanding of plant cell walls from synthesis to deconstruction. This understanding would generate models, theories and finally processes that will be used to understand and overcome biomass recalcitrance. This grand challenge calls for an integrated research approach, which is illustrated by the project described below. BESC research is multidisciplinary by design and multi-institutional in composition.**

The Computational Biology Team of BESC has been developing a suite of computational biology and bioinformatics tools in support of the needs of researchers at BESC and beyond. These tools address *omics* data analysis and interpretation, genome assembly and annotation, protein and assembly structure/functional prediction and analysis, phylogenetic tree reconstruction and analysis, as well as inference, analysis and modeling of regulatory and metabolic pathway systems. In addition, a number of databases have been developed for organizing data generated to support BESC research. These databases include the DOOR operon database, genome Barcode server cBAR, pDAWG for plant cell wall genes, and the GASdb database for glycosyl hydrolysis systems. Most of these tools and databases are in the process of being integrated into the BESC Knowledge Base (KBase).

Currently the core KB for plants consists of 21 plant genomes including six algal genomes, along with a rich set of annotated data and computed information for: (a) gene structures; (b) protein products; (c) homology-based functional prediction; (d) domain structures; (e) ortholog and paralog prediction; (f) gene ontology; and (g) metabolic and enzymatic pathways. Currently, the KB's reference plant data consist of over 500,000 coding genes from which nearly 400,000 protein coding genes with function prediction have been identified. The Plant KB maintains available gene model variations, alternative gene models (including alternative splicing), and historical versioning. The CompBio team has also collected rich set of omics data from external resources like NCBI GEO and EMBL ArrayExpress that are related to the assembly of the cell wall pectic matrix, cellulose synthesis and cell elongation, cellulose synthase mutants, cell wall stress, primary to secondary stem development and many other experiments for *Arabidopsis*, rice and popular.

The core KB for microbes consists of information regarding 37 microbes, including biomass degraders, fuel producers, endophytes and model organisms. Data collected on these organisms include genome annotation, biochemical data including enzymes, ligands and pathway annotation generated using the Pathway Tools software and provided in KEGG, carbohydrate active enzyme data generated by CAT and provided by CAZy, operon predictions from BeoCyc and DOOR, protein functional domain predictions from CDD, Pfam, COGS, TigrFam, SMART. The KB also contains omics data on the effects of alcohol on cells, fermentation time courses, growth on model substrates found in lignocelluloses, growth on simple and complex sugars, biomass deconstruction, and many other experiments from several *Clostridium* and other microbes relevant in bioenergy research.

Using these capabilities, the CompBio team has carried out a variety of computational studies in support of the BESC mission. Examples include: (1) a molecular dynamics simulation of lignocellulosic biomass; (2) simulation of enzyme reaction mechanisms using combined quantum mechanical and molecular mechanical models; (3) genome annotation of *Caldicellulosiruptor bescii*; (4) detailed functional inference and phylogeny analyses of a number of gene families associated with plant cell walls, such as GT8 genes, the cellulose synthase superfamily, UDP-4-keto-pentose/UDP-xylose synthase, nucleotide sugar interconversion enzymes and syringyl lignin biosynthesis genes; (5) modeling of lignin biosynthesis in *Populus* and *Medicago*; and (6) systems biology studies of *C. thermo*.

Our ultimate goal in developing these tools and the KBase is to provide one unified interface and computational environment to facilitate integrated analyses of plant cell-wall related genes, pathways and enzymes, and microbes capable of degrading lignocellulosic materials.

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.

# 17

## High-Throughput Compositional Analysis of Lignocellulosic Biomass

Jaclyn DeMartini[1]* ( jdema001@ucr.edu), Michael J. Selig,[4] Melvin P. Tucker,[2] Geoffrey Turner,[3] Michael H. Studer,[2] Michael E. Himmel,[3] **Charles E. Wyman**,[1] **Stephen R. Decker**,[3] and **Paul Gilna**[5] (BESC PI)

[1]Center for Environmental Research and Technology, University of California-Riverside; [2]National BioEnergy Center, National Renewable Energy Laboratory, Golden, Colo.; [3]BioSciences Center, National Renewable Energy Laboratory, Golden, Colo.; [4]Forestry and Wood Products, Faculty of Life Sciences, Copenhagen University, Copenhagen, Denmark; and [5]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

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 researchers in characterization, modeling, and data management areas are engaged in 1) applying advanced technologies to analyze chemical and structural changes within biomass and 2) storing, tracking, analyzing and integrating data and understanding across the center.**

The determination of the composition of lignocellulosic materials is a crucial experimental step in the study of lignocellulose recalcitrance and the conversion to fuels and chemicals. Typical protocols require significant amounts of material, as well as cumbersome equipment and time consuming steps to produce even adequately accurate results. Building on our earlier high-throughput recalcitrance platforms, we have developed HTP methods for determining the key structural carbohydrates, lignin, and ash present in lignocellulosic materials. One system utilizes deepwell 96-well reactors to perform a scaled down version of the most commonly cited two-stage sulfuric acid hydrolysis protocol. Rapid determination of glucose and xylose released by the process is performed via glucose oxidase and xylose dehydrogenase coupled assays. HPLC analysis can be used for more precise quantification of all sugars. Alternatively, another HTP compositional analysis system we have developed utilizes 1.5 mL high recovery HPLC vials as reactors. This system also follows the standard two-stage sulfuric acid hydrolysis protocol, but in addition to measuring carbohydrate content via HPLC, it also enables the estimation of Klason lignin content through the measurement of acid insoluble residue (AcIR). Furthermore, if desired, this system can provide the whole ash content in a separate step.

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.

# 18

## BioEnergy Science Center Education and Outreach

Sue Kral,[1] J. Griffin,[1] H. Schulson,[1] Suzy Fowler,[2] Brian Davison,[2] **Jan Westpheling**[2,3]* (JanWest@uga.edu), and **Paul Gilna**[2] (BESC PI)

[1]The Creative Discovery Museum, Chattanooga, Tenn.; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [3]University of Georgia, Athens

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). Education and outreach to the general public is critical in the acceptance and deployment of bioenergy. In addition to leveraging successful education and training programs already in place at our partner institutions, BESC has developed educational lessons and activities that target elementary and middle school children.**

In addition to our efforts to prepare a new generation of scientists for the emerging fields of bioenergy through the interdisciplinary training of graduate students and postdocs, our center has taken a novel approach in that our education efforts begin with fifth graders. We have developed lesson plans aimed at 4th, 5th and 6th grades to educate and inform students about the basics of energy production and utilization. These lessons include basic concepts such as the carbon cycle, lignocellulosic biomass as substrate for the production of biofuels as well as technical and economic obstacles to a bio-based fuel economy. Lessons were piloted in schools in North Georgia and Tennessee in the 2008-2009 school year and have, to date, reached more than 5000 students. In addition to the in school lessons, we have established "science night" programs offered to students, their parents and the general public through local schools, museums and community centers reaching more than 8000 students and parents. Through summer workshops we have established national outreach through museums in 6 states and in the 2010-2011 school year will reach more than 10,000 students nationwide with family nights anticipated to reach more than 25,000 students and parents. In addition, a traveling exhibit will be completed by the end of 2011 suitable for outreach to schools, museums, libraries and community centers throughout the country.

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.

# 19

## Defining the Minimum Components of the Chromosomal Origin of Replication of the Hyperthermophilic Anaerobe, *Pyrococcus furiosus*: Use for Construction of a Stable Replicating Shuttle Vector

Joel Farkas,[1,3] Daehwan Chung,[1,3] Megan DeBarry,[1,3] Michael W.W. Adams,[2,3] **Janet Westpheling**[1,3]* (JanWest@uga.edu), and **Paul Gilna**[3] (BESC PI)

[1]Departments of Genetics and Biochemistry, University of Georgia-Athens; [2]Molecular Biology University of Georgia-Athens; and [3]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

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 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.**

We report the construction of a replicating shuttle vector that consists of a low copy number cloning vector from *E. coli* and functional components the origin of replication (*oriC*) of the *Pyrococcus furiosus* chromosome. In the process of identifying the minimal origin sequences required for autonomous plasmid replication in *P. furiosus,* we discovered that several features and structures of the origin predicted by bioinformatics analysis and *in vitro* binding studies were, in fact, not essential for stable autonomous plasmid replication. A minimal region required to promote plasmid DNA replication was identified and plasmids based on this sequence readily transformed *P. furiosus*. The plasmids replicated autonomously and existed in single copy. In contrast to shuttle vectors based on a *P. abyssi* plasmid for use in *P. furiosus*, plasmids based on the *P. furiosus* chromosomal origin were structurally unchanged after DNA transformation and were stable without selection for more than one hundred generations.

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.

# 20

## Genetic Tool Development and Deployment in *Clostridium thermocellum*

Adam M. Guss[1,2]* (gussam@ornl.gov), Daniel G. Olson,[1,2] D. Aaron Argyros,[1,3] Shital A. Tripathi,[1,3] Trisha F. Barrett,[1,3] Bethany B. Miller,[1,3] David A. Hogsett,[1,3] Nicky C. Caiazza,[1,3] **Lee R. Lynd**,[1,2,3] and **Paul Gilna**[1] (BESC PI)

[1]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Thayer School of Engineering, Dartmouth College, Hanover, N.H.; and [3]Mascoma Corporation, Lebanon, N.H.

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 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 thermophilic anaerobic bacterium that specializes in the rapid solubilization and fermentation of crystalline cellulose to products that include ethanol. As such, it is considered a prime candidate for consolidated bioprocessing, the process by which a single organism is responsible both for the hydrolysis and fermentation of plant biomass for biofuel production. However, the dearth of genetic tools available in this organism has hindered its development into a practical industrial strain. Here we describe the recent advances leading to a suite of genetic tools for the manipulation of *C. thermocellu*m. Positive and negative selectable markers have been developed that, when used together, can lead to the generation of unmarked deletions. Due to the scarcity of thermophilic selectable markers, this technology is essential for the genetic modification of *C. thermocellum* for both fundamental and applied research. Furthermore, investigation into the cause of differences in transformation efficiency of similar plasmids led to new understanding of factors that affect the efficiency of transformation. Using these new techniques, greater than 20 genes have been deleted or heterologously expressed singly or in combination in *C. thermocellum*. Targeted genes include ones involved in cellulose solubilization, central metabolism, electron transport, and gene regulation. Analysis of these mutants has led to a better understand the physiology of *C. thermocellum*, as well as increased yield of ethanol from cellulose.

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.

# 21

## Reconstruction of Transcription and Metabolic Networks in *Clostridium thermocellum*

Andrey A. Gorin[1]* (agor@ornl.gov), Israel J. Huff,[1] Tatiana Karpinets,[1] Nikita D. Arnold,[1] Steven D. Brown,[1] Tamah Fridman,[1] Loren J. Hauser,[1] Michael Leuze,[1] Guruprasad Kora,[1] Mustafa Syed,[1] Byung Park,[1] Daniel J. Quest,[1] Shihui Yang,[1] **Robert W. Cottingham**,[1] **Edward Uberbacher**,[1] and **Paul Gilna** (BESC PI)[1,2]

[1]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

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 researchers in characterization, modeling, and data management areas are engaged in 1) applying advanced technologies to analyze chemical and structural changes within biomass and 2) storing, tracking, analyzing and integrating data and understanding across the center.**

*Clostridium thermocellum* (Ct) is one of the most intensively studied targets in the quest for the organisms suitable to industrial-scale production of biofuels from cellulosic biomass. Ct is considered as one of the model bacterial organisms for consolidated bioprocessing. The ongoing efforts to improve various metabolic characteristics of Ct require much better characterization of its regulatory and metabolic networks. The network elucidation faces many experimental and algorithmic obstacles, including the metabolic complexity of Ct (only ~50% of genes have any functional annotation), the lack of systematic experimental assays for unknown enzymatic functions, and a limited availability of the gene expression profiles obtained under different conditions. We have begun to address these challenges by the following approaches: (a) utilizing new experimental technologies — Roche NibleGen High Density Tiling Arrays (HDTAs) and RNA high-throughput sequencing (RNAseq); (b) developing novel mathematical algorithms for processing and reliable interpretation of the expression profiles; (c) assembling individual Knowledge Base (KB) tools into an integrated pipeline for the discovery of Ct metabolic pathways and evaluation of genetic variation effects on the metabolic capabilities.

We have demonstrated that the HDTAs provide important advantages in comparison to standard expression arrays: (1) the large number of probes eliminates many types errors by analyzing consistency between probe measurements belonging to the same gene; (2) other sources of noise (e.g. occurring during preparation and handling of the RNA samples) can be detected and accurately quantified; (3) probes to the intergenic space significantly help with an identification of transcription units. Essentially, the large number of probes can be interpreted as a presence of multiple technical replicas collected for the same sample. These virtual "replicas" can be exploited to assign reliable error bars to the measurements of the differential expression. The results are significantly improved with a careful preprocessing of the raw data, and we have implemented the required tools to remove unreliable probes, to subtract background signals, and to compensate for saturation effects. We developed and evaluated several algorithms for gene expression quantification together with the estimation of "error bar" values for these measurements. The developed algorithms are applicable to both HDTA and RNAseq data flows and provide quality guidance for experimental protocol evaluations. The KB will provide a graphical interface tool — HDTA/RNAseq Browser — to compare different expression profiles and to overlap them with current genome annotations.

A consistent analytical approach was developed to integrate information from multiple conditions (3 time points for the Ct growth under the ethanol stress, 2 time points from the switchgrass growth curve, and 3 time points from the cellobiose growth curve) as well as across multiple signals that can indicate a transcription unit boundary between two neighboring genes. We found over 1200 distinct transcription units ranging from 1 to 16 genes that by our estimates closely correspond to the underlying operon structure (some of the units may contain several operons). Remarkably, a large majority of the units could be reliably determined from any pair wise comparison of the investigated conditions. This observation leads to several important conclusions. First, it illustrates a tremendous superiority of HDTA data over standard arrays. Under standard expression arrays technology any experimental investigation of the operon structure would require a collection of at least 30 or more expression profiles obtained under sharply different conditions. Second, and more importantly, the observation reveals that a very similar "palitra of transcription" is utilized in two different scenarios: same transcription units are regulated in response to a dramatic change of the growth media (switchgrass) and to relatively mild environmental stress (ethanol). The magnitudes and directions of the responses are vastly different, but the bacterial genome seems to be divided into the same units in both cases. It probably implies that (1) these groups of genes indeed reflect tightly coupled processes (and, correspondingly, coupled biological functions) and (2) even under mild regulatory rearrangements we can observe differential expression responses of the individual units.

On another track of the same project we have investigated a feasibility to map exact positions of the transcription starts in the Ct genome. The experimental identification of positions yields valuable insights into consensus promotor sequences, which allows more localized searches of the transcription factor binding sites, helps to refine annotations of the genes, etc. We found that the best HDTAs do contain information that can be used to map start positions. In more

than 800 cases we were able to locate such starts with an estimated precision of only 5-10 nucleotides. A customized version of Welch *t*-test was used to find exact locations and provide an analytical estimate of the detection reliability. In preliminary studies over 300 cases were found with the expected False Discovery Rate (FDR) under $10^{-3}$ and over 500 cases with FDR values between $10^{-2}$ and $10^{-3}$. This work will be continued toward more comprehensive reconstruction of Ct Gene Regulatory Network. The established transcription units could be assembled into regulon clusters, and the corresponding calculations would benefit from an acquired understanding of the measurement errors and data variability.

Finally, we will utilize our KB tools for the reconstruction of Ct metabolic pathways activated in the observed transcription changes. Such work may lead to new functional hypothesis for specific genes, better understanding of metabolite and energy flows, and more clear picture of how different flows are coupled together. The working hypotheses will be refined by the analysis of proteome and metabolome data already deposited in KB, giving new leads for the engineering efforts to improve metabolic capabilities of this important template organism.

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.

# 22

## Mutant Alcohol Dehydrogenase Leads to Improved Ethanol Tolerance in *Clostridium thermocellum*

**Steven D. Brown[1]*** (brownsd@ornl.gov), Xiongjun Shao,[1,2] Adam M. Guss,[1,2] Ashwini Bhandiwad,[1,2] Tatiana V. Karpinets,[1] Jerry M. Parks,[1] Nikolai Smolin,[1] Shihui Yang,[1] Miriam L. Land,[1] Dawn M. Klingeman,[1] Miguel Rodriguez Jr.,[1] Babu Raman,[1] Mingjun Zhu,[4] Jonathan R. Mielenz,[1] Jeremy C. Smith,[1,3] Timothy J. Tschaplinski,[1] Nancy L. Engle,[1] Richard J. Giannone,[1] Robert L. Hettich,[1] Martin Keller,[1] **Lee R. Lynd,[1,2]** and **Paul Gilna[1]** (BESC PI)

[1]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Thayer School of Engineering, Dartmouth College, Hanover, N.H.; [3]Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville; and [4]School of Bioscience and Bioengineering, South China University of Technology, University Town, Guangzhou, China

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 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.**

Fuels from cellulosic biomass are among the leading options to meet sustainability and energy security challenges associated with fossil fuels, and conversion processes featuring biological fermentation are among the leading options for producing cellulosic biofuels. Among fermentation-based conversion processes, use of cellulose-fermenting microorganisms without added enzymes—consolidated bioprocessing or CBP—has strong potential and a variety of microorganisms are under development. *Clostridium thermocellum* is a model thermophilic bacterium that can rapidly solubilize biomass and utilize cellulose as a carbon and energy source. Wild-type strains produce ethanol as well as organic acids but growth is inhibited by relatively low ethanol concentrations (<10 g/L). Cultures of *C. thermocellum* have been adapted to tolerate ethanol concentrations as high as 80 g/L, and while greater ethanol production has been reported for tolerant strains the highest concentration of ethanol production reported for this organism is <30 g/L. We have developed and characterized *C. thermocellum* mutant strains that can grow in the presence of up to 50 g/L ethanol. One study utilized a distinctive strategy of alternating between increasingly stringent selections for greater ethanol tolerance and relaxation of selection pressure. By this strategy, the adapted strains retained their ability to grow on either cellobiose or crystalline cellulose, and displayed a higher growth rate and biomass yield than the wild-type strain in the absence of ethanol. Another strain, selected with only increasing doses of ethanol, was more tolerant to ethanol but grew poorly. Several systems biology studies elucidated key metabolites, genes and proteins that form the foundation of its distinctive physiology and the multifaceted response to ethanol stress for the *C. thermocellum* wild-type strain and several ethanol tolerant mutant strains. The genomes of three ethanol tolerant mutant strains and a wild-type strain were resequenced, which revealed a mutated bifunctional acetaldehyde-CoA/alcohol dehydrogenase gene (*adhE*) in each of the mutants. We hypothesized based on structural analysis that cofactor specificity may be impacted, and confirmed this hypothesis using enzyme assays. Biochemical assays confirm a complete loss of NADH-dependent activity with concomitant acquisition of NADPH-dependent activity, which likely affects electron flow in the mutant strain. The simplicity of the genetic basis for the ethanol-tolerant phenotype observed here informs rational engineering of mutant microbial strains for cellulosic 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.

# 23

## Understanding Cellulosomal Enzymes: Experiment and Modeling

Yannick J. Bomble[1,2]* (yannick.bomble@nrel.gov), Markus Alahuhta,[1,2] Vladimir V. Lunin,[1,2] Michael G. Resch,[1,2] Qi Xu,[1,2] Michael F. Crowley,[1,2] Mark R. Nimlos,[1,2] **Michael E. Himmel,[1,2]** and **Paul Gilna[2]** (BESC PI)

[1]National Renewable Energy Laboratory, Golden, Colo.; and [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

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 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 X-ray structure of *Clostridium thermocellum* family 4 carbohydrate binding module from cellobiohydrolase A (CbhA) was solved recently. The new structure of CBM4 exhibits a binding site peptide loop with a tryptophan (Trp118) residing midway in the loop. Molecular dynamics simulations and experimental binding studies with the Trp118Ala mutant suggest that Trp118 contributes to the binding and possibly to the orientation of the module to soluble cellodextrin chains. Furthermore, the binding cleft aromatic residues, Trp68 and Tyr110, play a crucial role in binding to bacterial microcrystalline cellulose (BMCC), amorphous cellulose, and soluble oligodextrins. However, CBM4 binding to BMCC is in disagreement with the structural features of the binding pocket. We therefore propose that *Clostridial* CBM4 modules have the ability to bind the free chain ends of crystalline cellulose in addition to their ability to bind soluble cellodextrins. Additionally, the two X1 modules from CbhA have also been crystallized and we hypothesize that they could serve as tethers between several modules in CbhA to fine-tune the positioning on this enzyme during cellulose deconstruction. We conducted steered molecular dymanics simulations (directly comparable to AFM pulling experiments) to evaluate the energy and the profile for unfolding these modules. We found that the energy required to partially unfold this domain is accessible in the cellulosome and the unfolding pathways are similar in the multiple simulations conducted.

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.

# 24

## Progress in Quantitative Assessment and Interpretation of Cellulosome Activity

Qi Xu[1]* (qi.xu@nrel.gov), John Baker,[1] Michael Resch,[1] Roman Brunecky,[1] John Yarbrough,[1] William S. Adney,[1] Shi-you Ding,[1] **Michael E. Himmel,[1]** and **Paul Gilna[2]** (BESC PI)

[1]BioEnergy Science Center, National Renewable Energy Laboratory, Golden, Colo.; and [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

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 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.**

"Cellulosomes" produced by certain strains of bacteria are very large and complex supramolecular "machines" that are extremely efficient in catalysis of depolymerization of cellulose to soluble sugars. Our experimental results emphasize the extent to which cellulosome activity is affected by a wide variety of factors that can be roughly grouped into (1) those factors that alter the composition of the cellulosome complex itself, and (2) those assay conditions affecting activity and survivability of cellulosome components. Cellulosome compositions are dependent on the growth conditions under which the organism produces the complex and upon the procedure used to separate the "cellulosomal" fraction from other proteins. Enzyme activity and survivability are strongly affected by assay conditions such as substrate, pH, temperature, redox potential, enzyme loading, assay duration, substrate conversion, and accumulation of inhibitory products such as cellobiose. Our laboratory is investigating the effects of these factors on assay results using two different categories of "cellulosomal" preparations, one category being "native" cellulosomal fractions produced and purified under a variety of conditions, the other category being genetically-engineered "designer minicellulosomes" containing selected, defined arrays of recombinant catalytic, substrate-binding, scaffoldin and linker domains. Findings from this large experimental array are presented and discussed in terms of both assay-development and implications for fundamental understanding of cellulosome action.

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.

# 25
## Cellulosome Self Assembly

Yannick J. Bomble [1,2]* (Yannick.Bomble@nrel.gov), Michael F. Crowley,[1,2] Michael G. Resch,[1,2] Richard J. Giannone,[2] Qi Xu,[1,2] Daniel G. Olson,[2,3] Javier A. Izquierdo,[2,3] **Lee R. Lynd**,[2,3] **Michael E. Himmel**,[1,2] and **Paul Gilna**[2] (BESC PI)

[1]National Renewable Energy Laboratory, Golden, Colo.; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [3]Dartmouth College, Hanover, N.H.

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). Addressing the roadblock of biomass recalcitrance will require a greater understanding of plant cell walls from synthesis to deconstruction. This understanding would generate models, theories and finally processes that will be used to understand and overcome biomass recalcitrance. This grand challenge calls for an integrated research approach, which is illustrated by the project described below. BESC research is multidisciplinary by design and multi-institutional in composition.**

Most bacteria and fungi use free enzymes to degrade plant cell walls. However, some bacteria have developed a fundamentally different approach, where enzymes are tethered to a large protein scaffold forming a complex known as the cellulosome. The study of these large protein assemblies is an ongoing research topic that has already yielded numerous breakthroughs. However, the mechanism of assembly of the enzyme subunits onto the natural scaffoldin; as well as the modes of action of the cellulosome and its enzymatic components, are not currently well understood.

In this study, we focused on the cellulosome-integrating protein (CipA) of *C. thermocellum*; as well as cellulosomal enzymes from families 5, 9 and 48. These three enzymes are representative of the diversity of enzymes secreted by *C. thermocellum*. This work not only aims at understanding the mechanisms involved in the sequential binding of the cellulosomal enzymes to the CipA scaffold of *C. thermocellum*, but also the binding of CipA to secondary scaffolds. We focused on the physical properties of the binding of component enzymes to the scaffoldin. The modularity of the enzymes was found to be one of the main influences on the cellulosome assembly process.

Additionally, we have grown *C. thermocellum* on four different carbon sources, which resulted in different cellulase and hemicellulase enzymes associated with the cellulosome. To confirm this observation, we identified and quantified the protein composition of purified *C. thermocellum* cellulosomes by mass spectrometry. The differences in enzyme composition were ultimately found to affect cellulosome activity. However, these different cellulosome compositions cannot be directly compared to our computational model until the local enzyme concentration in the vicinity of the scaffold is known. We are at the moment developing methodologies to access this crucial parameter. A new CipA delete strain of *C. thermocellum* will be used to test various hypotheses regarding engineering the scaffold.

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.

# 26
## Comparative Genomic Analysis and Plant Biomass-Degrading Mechanisms of the Extremely Thermophilic Genus *Caldicellulosiruptor*

Ira Kataeva[1]* (kataeva@uga.edu), Sara Blumer-Schuette,[4]* Inci Ozdemir,[4] Dhaval Mistry,[4] Jeff Zurawski,[4] Marcus Foston,[5] Sivakumar Pattathil,[3] Art Ragauskas,[5] Michael Hahn,[3] Phuongan Dam,[2] Ying Xu,[2] **Robert M. Kelly,[4] Michael W.W. Adams,[1]** and **Paul Gilna**[6] (BESC PI)

[1]Department of Biochemistry and Molecular Biology, University of Georgia-Athens; [2]Institute of Bioinformatics, University of Georgia-Athens; [3]Complex Carbohydrate Research Center, University of Georgia-Athens; [4]North Carolina State University, Raleigh; [5]School of Chemistry and Biochemistry, Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta; and [6]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

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 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.**

Phylogenetic, microbiological and comparative genomic analysis was used to examine diversity among members of the genus *Caldicellulosiruptor* with an eye towards the capacity of these extremely thermophilic bacteria for degrading the complex carbohydrate content of plant biomass (PB). Although 16S ribosomal RNA sequences across the genus

are at least 94% similar, differences in growth physiology across the genus have been noted. In order to further understand the observed physiological differences, whole genome sequencing projects for five members of the genus were initiated, in addition to three already completed genomes. Complete genome sequences indicate that biogeography influences the genomic composition of related species. Additionally, comparative carbohydrate active enzyme (CAZy) inventories indicated that the absence of a single glycoside hydrolase family and carbohydrate binding motif family appear to be responsible for some *Caldicellulosiruptor* species' diminished cellulolytic capabilities. Overall, the genus *Caldicellulosiruptor* appears to contain more genomic and physiological diversity than previously reported, and is well suited for biomass deconstruction applications. One of this group, *C. bescii*, is the most thermophilic cellulose-degrading organism known, with an optimal growth temperature near 80˚C. It is also capable of degrading untreated PB, even from plants with a high lignin content such as switchgrass (SWG). The mechanism by which SWG is degraded by *C. bescii* was investigated. Growth of the organism on SWG that had been previously washed for 18 hours at 78˚C with water (wSWG) in three consecutive cultures resulted in 85% solubilization, compared to 18% solubilization if the organism was not present. Analysis of the 15% insoluble material remaining after the three treatments with *C. bescii* revealed that its glucose:xylose:lignin ratio was not significantly changed by treatments. Similarly, the physical and spectroscopic properties of cellulose and lignin in the residual biomass were largely unaffected. SWG profiling with 150 monoclonal antibodies (AB), raised against PB epitopes excluding lignin and cellulose, showed that binding of majority of ABs to SWG was not significantly changed by the microbial action. The results suggest that thermal treatment and *C. bescii* work in a concert, and that conversion of SWG follows an "onion peeling" mechanism whereby the residual material resembles the untreated biomass.

This research was funded by the BioEnergy Science Center, which is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

## Great Lakes Bioenergy Research Center (GLBRC)

# 27

## Post-Glacial Evolution of Switchgrass: Centers of Diversity, Gene Pools, and Gene Flow

**Michael D. Casler**[1]* (michael.casler@ars.usda.gov), Yunwei Zhang,[2] Juan Zalapa,[3] Andrew R. Jakubowski,[4] David L. Price,[5] Ananta Acharya,[6] Yanling Wei,[7] E. Charles Brummer,[8] and Shawn M. Kaeppler[9]

Co-authors YZ, JZ, and ARJ contributed equally to the work described in this abstract. This research represents a formal collaboration between the Great Lakes Bioenergy Research Center (GLBRC) and the BioEnergy Science Center (BESC).

[1]USDA-ARS, U.S. Dairy Forage Research Center, Madison, Wis. and DOE Great Lakes Bioenergy Research Center; [2]Grassland Institute, China Agricultural University, Beijing, China; [3]USDA-ARS, Vegetable Crops Research Unit, Department of Horticulture, University of Wisconsin, Madison (formerly DOE Great Lakes Bioenergy Research Center); [4]Department of Agronomy, University of Wisconsin, Madison; [5]Department of Agronomy, University of Wisconsin, Madison; [6]Center for Applied Genetic Technologies, Crop and Soil Science Department, University of Georgia, Athens and DOE BioEnergy Science Center; [7]The Samuel Roberts Noble Foundation, Ardmore, Okla.; [8]The Samuel Roberts Noble Foundation, Ardmore, Okla. and DOE BioEnergy Science Center; and [9]Department of Agronomy, University of Wisconsin, Madison, and DOE Great Lakes Bioenergy Research Center

http://www.glbrc.org/

**Project Goals:**

1. **Identify primary and secondary centers of diversity for switchgrass,**
2. **Identify patterns of genetic diversity among and within the three major taxa of switchgrass (lowland 4x, upland 4x, and upland 8x),**
3. **Determine the level of historic gene flow between upland and lowland ecotypes,**
4. **Identify accessions that may be sources of hybrids between upland and lowland ecotypes.**

A total of 480 switchgrass genotypes, belonging to 71 accessions, were evaluated for variation in 19 SSR markers and cpDNA sequences at five introns or intergenic regions. Approximately 75% of the individuals were unequivocally classified as upland or lowland individuals, based on phenotype, genotype (SSR markers), and cytotype (cpDNA sequences). The remaining 25% of individuals represented a wide range of unusual individuals with various combinations of phenotype, genotype, and cytotype. Each of these individuals represents some level of ancient gene flow; 46 of these individuals represent the strongest evidence for

hybrid origin and ancient gene flow (Figure 1). The various combinations of phenotype, genotype, and cytotype reveal bidirectional gene flow between upland and lowland ecotypes, between tetraploid and octoploid chromosome levels, and different levels of gene flow indicative of ancient hybridizations of differing ages, occurring over an extremely long period of time.

The primary center of diversity for switchgrass is represented by remnant patches of prairie and savanna in the eastern Gulf Coast region, extending northward along the Atlantic Seaboard (Figure 2). Many of the southeastern sites are likely ancient remnants that served as prairie and savanna refugia during major ice age events. Much of the variation present in the eastern Gulf Coast region was preserved along the Atlantic Seaboard, largely due to the relatively mild climate change along this latitudinal gradient. Molecular clock computations suggest that the earliest upland-lowland divergence and the earliest transitions from tetraploids to octoploids occurred approximately 1.5 to 1 M ybp. Repeated ice age cycles, leading to cyclic migrations between northern prairie and savanna sites and southern refugia, have preserved massive amounts of genetic variability within individual polyploid genotypes, within local habitats, and across a broad geographic landscape. Much of this variability is available for use in improving switchgrass as a dedicated bioenergy crop.

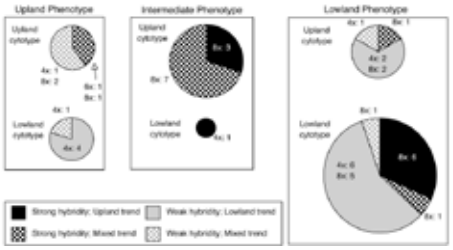



Figure 1. Pie diagrams showing the relationships among phenotype, cytotype (cpDNA sequence), and nuclear SSR marker profiles for 46 switchgrass plants that were identified as having marginal to high probabilities of hybrid descent (pH > 0.05) and presented in Tables 3 and 4. Each pie chart refers to one group defined by cytotype and phenotype and the size of each pie represents group size, with n = 1 for the smallest and n = 19 for the largest group. The strength of hybrid support is indicated by black (strong; pH > 0.50) or grey (moderate; 0.05 < pH < 0.50) and the predominant ecotypic trend of the hybrid-origin genotypes by the pattern (solid for dominant upland or lowland SSR pattern or trend; checkered for mixed SSR pattern or trend). Number of tetraploids (4x) and octoploids (8x) are identified for each slice of pie.

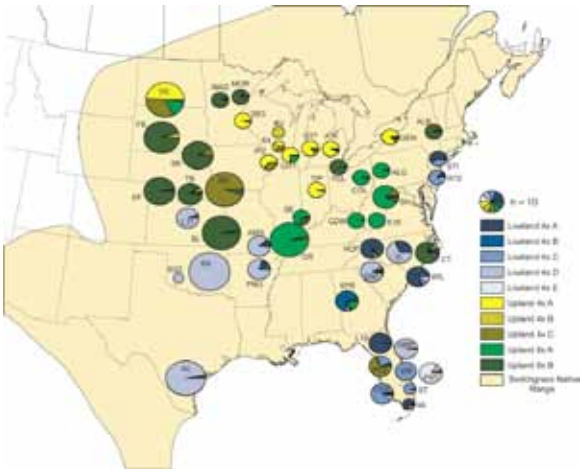


Figure 2. Partial map of the USA, showing the approximate location of each switchgrass cultivar or accession and the probabilities of accession membership in one of 10 groups, identified by STRUCTURE analysis. The size of each circle represents the sample size for each accession, with n = 10 shown in the legend. Each cultivar or accession is identified by a two- or three-character code from Table 1. Each "slice of pie" represents the probability of membership in one of the 10 STRUCTURE groups shown in the legend.

# 28

## Dynamic Carbon Utilization Patterns in *E. coli* Grown in AFEX-Pretreated Corn Stover Hydrolysate

M.S. Schwalbach* (mschwalbach@glbrc.wisc.edu), M. Tremaine, J. Grass, A. Higbee, D. Keating, P. Kiley, and R**. Landick**

Great Lakes Bioenergy Research Center, University of Wisconsin

http://www.glbrc.org/

**Project Goals: This research project is part of a larger effort aimed at improving microbial conversion of lignocellulose material to biofuels. Specifically, in this work we utilized multiomic technologies to examine substrate utilization patterns of *E. coli* cultivated in complex plant hydrolysate media.**

A core mission at the Great Lakes Bioenergy Research Center is the identification and mitigation of bottlenecks in the production of biofuels from lignocellulose. Previous data have suggested that a key bottleneck involves the inhibition of microbial fermentation by compounds liberated during pre-treatment and enzymatic degradation of plant cell walls. To investigate the effect of plant-derived inhibitors on microbial conversion, we carried out a multiomic analysis of wild type and ethanologenic *E. coli* strains cultivated in hydrolysate prepared from ammonium fiber expansion (AFEX)-pretreated corn stover (CS). Clarification and pH adjustment of AFEX-CS hydrolysate yielded growth media

containing up to 60 g glucose/L and 30 g xylose/L. Surprisingly, although there were high concentrations of glucose in all hydrolysate media, results from fermentation experiments indicated that additional carbohydrates were apparently being co-utilized at the onset of fermentation, including both pentoses (arabinose) and hexoses (mannose, galactose, fructose). Conversely, xylose utilization occurred at a much slower rate than glucose, and did not occur until nearly all other carbohydrates had been exhausted, identifying this process as a target for future strain improvement. Although carbohydrates were the primary carbon source utilized in the fermentation, we observed low expression levels of genes encoding amino acid (AA) biosynthetic pathways at the onset of the fermentation, suggesting cells were utilizing free AA to meet protein biosynthesis demands. At later time points transcription levels of genes encoding some AA biosynthesis pathways increased, suggesting that specific amino acids may have been depleted from the hydrolysate medium. Interestingly, we also observed increased transcription of genes encoding AA degradation pathways and the glycine cleavage system, suggesting free AA were also utilized to meet cellular nitrogen demands and to augment intracellular C1 pools during the experiment. Thus, our data suggests that AA significantly impacted *E.coli* growth in AFEX-CS hydrolysate to the extent that AA warrant inclusion in future carbon flux models. We will use these results to identify physiological and fermentative bottlenecks in the production of ethanol, which will be alleviated via a combined approach of rational engineering and directed evolution.

# 29

## Systems Biology of *Cellvibrio japonicus* Lignocellulose Degradation

David H. Keating* (dkeating@glbrc.wisc.edu) and Jeffrey G. Gardner

DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison

http://www.glbrc.org/

**Project Goals: We aim to identify the core set of glycosylhydrolases necessary for lignocellulose breakdown in the Gram-negative bacterium *Cellvibrio japonicus*. Our approach involves the use of transcriptomic analysis of *C. japonicus* cells cultured in the presence of monosaccharides, artificial cellulosic substrates, and authentic lignocellulose. We will use the expression data to identify candidate glycosylhydrolases predicted to be critical for lignocellulose degradation, which will be tested by gene disruption and phenotypic evaluation.**

Lignocellulosic biofuels possess significant potential as replacements for petroleum. However, the development of economically viable lignocellulosic biofuels requires overcoming multiple challenges, most importantly the recalcitrance of plant cell walls to enzymatic deconstruction. One approach to improving the economic viability of biofuels is the use of consolidated bioprocessing (CBP), which involves the engineering into a single organism the ability to deconstruct biomass, and the subsequent conversion of the resulting sugars to biofuel. Research within the bioconversion area of the Great Lakes Bioenergy Research Center (GLBRC) involves the identification of Gram-negative bacteria capable of lignocellulose degradation, and the introduction of their lignocellulose-degrading genes into ethanologenic bacteria. Our studies have shown that the Gram-negative bacterium *Cellvibrio japonicus* can degrade the plant cell walls of bioenergy-relevant substrates such as corn stover and switchgrass. Furthermore, a large number of lignocellulose degrading enzymes from this bacterium have previously been cloned, expressed in *E. coli*, and biochemically characterized. Examination of the genome of *C. japonicus* has identified 154 candidate genes potentially involved in lignocellulose degradation. However, it is not known which of these genes are essential for efficient degradation of plant cell walls.

To identify the core set of lignocellulases necessary for cell wall deconstruction, we have developed methodology for measurement of global transcription in *C. japoncius*, during growth in the presence of lignocellulose and its purified polysaccharide components. Initial results suggest that the expression of lignocellulases is hierarchical in nature, and may involve both transcriptional and post-transcriptional forms of regulation. These transcriptional studies are being used to construct a prioritized list of candidate genes predicted to be critical for lignocellulose degradation. To determine which of these candidate genes are essential for degradation of plant cell walls, we have developed methodology for construction of targeted gene disruptions in *C. japonicus*. Initial results from gene disruption studies show that mutations in candidate glycosylhydrolases display unique and unexpected degradative phenotypes. Collectively, these results demonstrate the feasibility of our approach to identify high priority lignocellulase genes that can be used for engineering of CBP organisms, and to contribute to our knowledge of microbial cell wall degradation.

# 30

## High Throughput Production and Characterization of Cellulytic and Hemicellulytic Enzymes

**David Mead**[1,2*] (dmead@lucigen.com), Colleen Drinkwater,[1] Julie Boyum,[1] Jan Deneke,[1] Krishne Gowda,[1] Ronald Godiska,[1] Eric Steinmetz,[1] and **Phil Brumm**[2*]

[1]Lucigen Corp. and DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison; and [2]C56Technologies Inc. and DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison

http://www.glbrc.org/

**Project Goals: Produce new cellulytic and hemicellulytic enzymes for GLBRC partners.**

The efficient hydrolysis of biomass to five and six carbon sugars is limited by the lack of affordable, high specific activity biomass-degrading enzymes. Random shotgun screening of genomic and metagenomic libraries for genes encoding these biomass-degrading enzymes has had very limited success. A large and growing database of sequenced bacterial genomes encoding thousands of putative carbohydrase active enzymes (CAZymes) presents a rich resource for enzyme discovery. A functional survey of the CAZyme activities encoded in a single cellulolytic genome is daunting when performed one gene at a time. A simple high throughput expression cloning system was developed in conjunction with a multiplex assay for *endo* and *exo*-cellulases and hemicellulases in a microplate format. The simultaneous detection of multiple polysaccharide-degrading enzyme clones permits efficient whole genome cloning, expression and characterization. Using this system we have expressed, purified and characterized over a hundred unique CAZymes from the thermophilic, mesophilic and alkaliphilic microbes *Dictyoglomus turgidum*, *Fibrobacter succinogenes*, and *Bacillus cellulosilyticus*, respectively.

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

# 31

## Improved Enzyme Cocktails for Biomass Conversion

Goutami Banerjee[1]* (gbanerjee11@gmail.com), Suzana Car,[1] John S. Scott-Craig,[1] Melissa Borrusch,[1] and **Jonathan D. Walton**[1,2]

[1]Department of Energy (DOE) Great Lakes Bioenergy Research Center and [2]DOE Plant Research Laboratory, Michigan State University, East Lansing, Mich.

http://www.glbrc.org/

**Project Goals: To create defined enzyme mixtures of high specific activity in order to understand the activities and proportions for optimal release of sugars from diverse feedstock/ pretreatment combinations.**

A major bottleneck to producing lignocellulosic ethanol is the high cost of enzymes for converting biomass to fermentable sugars. Currently available commercial enzyme preparations are complex and poorly defined mixtures of enzymes and other proteins and have generally been optimized only for acid-pretreated grass stovers. We are working toward the creation of more efficient, and hence less expensive, enzyme cocktails by rationally designing defined mixtures. Mixtures of highly pure proteins optimized using a platform called GENPLAT, which uses statistical design of experiment, robotic pipetting of stover slurries and enzymes, and automated Glc and Xyl analysis. We have used GENPLAT to create optimized enzyme cocktails containing up to 18 components for release of sugars from multiple combinations of pretreatments (e.g., AFEX, alkaline peroxide, dilute base) and feedstocks (e.g., corn stover, *Miscanthus*, switchgrass, DDG, and poplar) (Fig. 1). The results obtained with GENPLAT indicate which enzymes are important and their optimal proportions. For example, GENPLAT results indicate that Cel61A of *Trichoderma reesei* is one of the most important enzymes for Glc release from corn stover (optimal proportion >20%), even though its precise enzymatic activity is not well-understood (Fig. 1). Also, GENPLAT indicates that a combination of two xylanases (of families GH10 and GH11) is superior to either xylanase alone for release of Glc or Xyl (Fig. 1). Just as importantly, GENPLAT results indicate which enzymes do not contribute to Glc or Xyl release from a particular pretreatment/feedstock combination (i.e., 0% optimal proportion), and are therefore unnecessary in industrial enzyme cocktails.

GENPLAT has several additional uses. One is as a platform to guide "bioprospecting" for novel accessory enzymes. Any new enzyme or protein, from any source, can be tested in defined mixtures on GENPLAT to ascertain and quantitate its possible utility in biomass deconstruction. GENPLAT can also be used to find more efficient examples of known enzymes, by substitution for the enzymes already in our defined mixtures. For example, alternate cellobiohydrolase 1 (CBH1) enzymes could be tested with GENPLAT in the context of a realistic enzyme cocktail and a realistic biomass substrate.

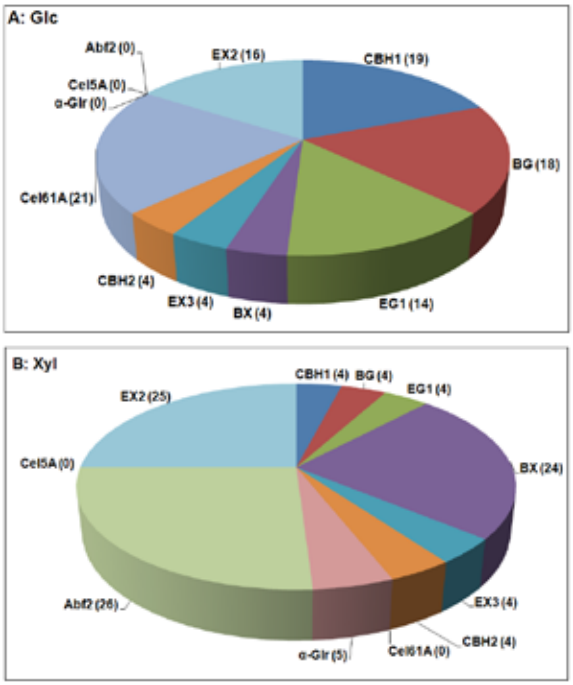



Figure 1. (A): Optimal proportions of 11 enzymes for release of Glc from alkaline peroxide-pretreated corn stover determined using GENPLAT. (B): Optimal proportions of the same enzymes for release of Xyl. Yields of Glc and Xyl were 85% and 65%, respectively, at an enzyme loading of 15 mg/g glucan and 48 h digestion (unpublished results).

**References**

1. Banerjee G, Car S, Scott-Craig JS, Borrusch MS, Aslam N, Walton JD (2010a) Synthetic enzyme mixtures for biomass deconstruction: production and optimization of a core set. Biotechnol Bioengineer 106:707-720.
2. Banerjee G, Car S, Scott-Craig JS, Borrusch MS, Bongers M, Walton JD (2010b) Synthetic multi-component enzyme mixtures for deconstruction of lignocellulosic biomass. Bioresour Technol 101:9097-9105.
3. Banerjee G, Car S, Scott-Craig JS, Borrusch MS, Walton JD (2010c) Rapid optimization of enzyme mixtures for deconstruction of diverse pretreatment/biomass feedstock combinations. Biotechnol Biofuels 3:22. 31.

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

# 32

## Discovery and Characterization of Cellulolytic Enzymes from Insect-Associated Microbial Communities

Adam Book* (ajbook@wisc.edu), Taichi Takasuka, **Brian Fox**, and **Cameron Currie**

DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison

http://www.glbrc.org/

**Project Goals: The U.S. Department of Energy has issued the challenge of discovering and designing enzymes with novel biomass-degrading capabilities. Our research in the Great Lakes Bioenergy Research Center has focused on the discovery of better enzymes.**

One of the major hurdles for producing cost effective biofuels is the enzymatic conversion of lignocellulosic biomass into sugars that subsequently can be fermented into ethanol. The U.S. Department of Energy has issued the challenge of discovering and designing enzymes with novel biomass-degrading capabilities. Our research in the Great Lakes Bioenergy Research Center has focused on the discovery of better enzymes. To achieve this goal, we have targeted our bioprospecting efforts on microbial communities that already thrive in lignocellulose-rich niches. The organisms in these environments have likely evolved over millions of years, and thus have optimized the dominant enzymes needed to support growth of the community. Throughout nature there are numerous examples of insects and symbiotic microbes that flourish in cellulose-rich environments. These co-evolved systems represent a diverse source for microbes that may have specialized in biomass utilization. This poster describes our initial work on microbes associated with the wood-boring wasps from the genus *Sirex*. We have used a systems biology approach to identify the dominant proteins and enzymes present. Initial results indicate that microbes from this niche express a large variety of complementary enzymes that are capable of the deconstruction of many constituent parts of plant biomass.

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

# 33

## Exploitation of Endogenous Variation for the Identification of Genes and Pathways Associated With Enhanced Biofuel Production in Maize

**Natalia de Leon**[1]* (ndeleongatti@wisc.edu), **Shawn Kaeppler**[1]* (smkaeppl@wisc.edu), Rajandeep Sekhon[1] (rsekhon@glbrc.wisc.edu), Candice Hansey[2] (hansey@msu.edu), Robin Buell[2] (buell@msu.edu), Kevin Childs[2] (kchilds@plantbiology.msu.edu), and Haining Lin[2] (linha@msu.edu)

[1]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison; and [2]DOE Great Lakes Bioenergy Research Center, Michigan State University

http://www.glbrc.org/

**Project Goals: Identify genes and pathways that will enhance ethanol potential per unit of land by exploiting available endogenous variation in maize.**

Biomass yield and composition are the primary determinants of ethanol potential per unit of land. Substantial phenotypic variation for biofuel related traits exists in available populations and germplasm collections of maize. The objective of our project is to exploit such variation for the identification of genes and pathways that enhance biomass production as well as carbohydrate concentration and digestibility of the cell wall in maize. In addition to enhancing the potential of maize as a source of biomass for biofuel production, information generated in this project will facilitate the improvement of other dedicated biofuel grass species such as switchgrass and *miscanthus*.

To genetically dissect endogenous variation for biofuel traits, biomass yield (stover yield/unit of land), quality (fiber digestibility and carbohydrate concentration), and component traits are being evaluated on genetic mapping populations and diverse inbred lines. Biomass yield is measured as the total dry weight of diverse genotypes in standard field settings. Quality is being evaluated using forage analytical tools such as neutral and acid detergent fiber, lignin concentration and composition as well as digestibility based both, on rumen bacterial assays as well as enzymatic digestion. Among component traits, plant height, internode length and leaf number are closely related to biomass yield. Populations being evaluated include the Nested Association Mapping (NAM) population, the Wisconsin Diverse (WIDIV) maize association panel as well as other recombinant inbred line populations. Substantial phenotypic variation is observed for biofuel related traits in these populations (Table 1). DNA and RNA from these materials are being genotyped to generate anonymous SNP markers. Numerous

quantitative trait loci (QTL) have been identified for the different traits (Table 2).

An initial trait being further characterized involves transition from juvenile to adult tissue, a developmental trait potentially related to biomass quality and quantity. The three most significant QTL for this trait are located on chromosomes 2, 3 and 9. The chromosome 9 region contains the gene *Glossy15*, involved in expression of juvenile leaf traits; however, this candidate gene has not been confirmed to be the QTL in this region. No candidate gene related to transition is known in the QTL regions of chromosomes 2 or 3 (Figure 1). Organ- and paralog-specific expression patterns of genes involved in pathways of interest are being examined in a collection of 60 different tissues of inbred line B73, the reference maize genome. This information has been used to further validate likely gene candidates.

| Trait | Number of Individuals | Population | Minimum | Maximum |
|---|---|---|---|---|
| | | | - unit = cm - | |
| Ear height | 1708 | NAM | 38.0 | 183.6 |
| Ear height | 904 | WIDIV | 29.0 | 144.7 |
| Internode length | 1708 | NAM | 3.8 | 17.2 |
| Internode length | 903 | WIDIV | 5.5 | 12.8 |
| Leaf width | 4099 | NAM | 7.0 | 15.8 |
| Plant height | 1708 | NAM | 88.0 | 276.0 |
| Plant height | 904 | WIDIV | 88.8 | 239.5 |
| Stalk diameter | 528 | WIDIV | 1.3 | 3.3 |

Table 1**.** Range of phenotypic variation for biomass yield component traits in the Nested Association Mapping (NAM) and Wisconsin Diverse (WIDIV) populations evaluated in Southern Wisconsin in 2008 and 2009.

| Trait | Year evaluated | Number of QTL deteceted | LOD | Chromosome location | QTL effect |
|---|---|---|---|---|---|
| | | | | | - unit = cm - |
| Plant height | 2009 | 23 | 72.1 | 9 | -8.7 |
| Ear height | 2009 | 18 | 85.4 | 9 | -7.7 |
| Internode length | 2009 | 17 | 31.2 | 9 | -0.45 |
| Leaf width | 2008 | 22 | 98.6 | 2 | -0.25 |

Table 2**.** Number of quantitative trait loci (QTL) detected and level of significance and effect size for the largest QTL identified for biomass yield component traits in the Nested Association Mapping (NAM) population evaluated in 2008 or 2009 at the Arlington, WI Agricultural Research station.

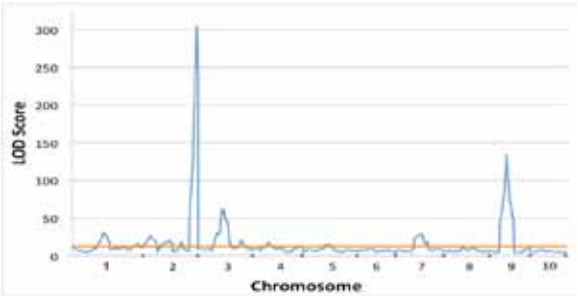



Figure 1. Joint quantitative trait loci (QTL) analysis of transition from juvenile to adult tissue on the Nested Association Mapping (NAM) population evaluated in 2008 and 2009 at the Arlington, WI Agricultural Research station.

# 34

## Cloning and Characterization of an Acyltransferase that Synthesizes Monolignol Ferulate Conjugates for Generation of 'Zipped' Lignins

Saunia Withers,[1] Fachuang Lu,[2,3] **John Ralph**,[2,3,4] and **Curtis Wilkerson***[1,5,6] (wilker13@msu.edu)

[1]DOE Great Lakes Bioenergy Research Center, East Lansing, Mich.; [2]DOE Great Lakes Bioenergy Research Center, Madison, Wis.; [3]Department of Biochemistry, University of Wisconsin, Madison; [4]Department of Biological Systems Engineering, University of Wisconsin, Madison; [5]Department of Plant Biology, Michigan State University, East Lansing; and [6]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing

http://www.glbrc.org/

**Project Goals: We are attempting to create transgenic plants with alterations in lignin that would allow for lower cost processing and higher yields of fermentable sugars.**

Lignin is an important cell wall component that provides structural support to plants and is necessary for vascular tissue function. This essential component is also responsible for much of the recalcitrance of the cell wall toward processing, impacting the task of producing paper and biofuels from plant cell walls. The chemical structure of lignin makes this polymer difficult to degrade by either chemical or enzymatic means. Grabber et al. (Biomacromolecules 2008, pp. 2510-6) have shown that the inclusion of ester bonds into the backbone of the lignin produces a polymer that is easier to chemically cleave. In order to engineer plants with such lignins it is necessary to introduce an enzyme into the plant that would produce a compound such as coniferyl ferulate to be utilized as a lignin 'monomer'. No enzyme has been cloned to date that produces such a product. We will describe the successful cloning and characterization of an acyltransferase that produces coniferyl ferulate. The properties of the cloned enzyme indicate that it is a good candidate for introducing into plants to, hopefully, produce lignins that allow improved access to the polysaccharides for subsequent utilization. We are currently generating transgenic *Arabidopsis* and poplar plants to evaluate the utility of this enzyme.

# 35

## Restructuring Crystalline Cellulose Hydrogen Bond Network Enhances its Depolymerization Rate

Shishir P.S. Chundawat[1,2]* (chundawa@msu.edu), Giovanni Bellesia,[3] Nirmal Uppugundla,[2] Leonardo da Costa Sousa,[1] Dahai Gao,[1] Albert Cheh,[4] Umesh P. Agarwal,[5] Christopher M. Bianchetti,[2,6] George N. Phillips Jr.,[2,6] Paul Langan,[7] Venkatesh Balan,[1,2] S. Gnanakaran,[8] and **Bruce E. Dale**[1,2]

[1]Biomass Conversion Research Laboratory, Department of Chemical Engineering and Material Science, Michigan State University, Lansing; [2]DOE Great Lakes Bioenergy Research Center; [3]T6 and CNLS, Los Alamos National Laboratory, Los Alamos, N.M.; [4]Department of Environmental Science and Department of Chemistry, American University, Washington D.C. [5]USDA Forest Service, Forest Product Laboratory, Madison, Wis.; [6]Department of Biochemistry, University of Wisconsin, Madison; [7]Bioscience Division, Los Alamos National Laboratory, Los Alamos, N.M.; and [8]T6, Los Alamos National Laboratory, Los Alamos, N.M.

http://www.glbrc.org
http://www.everythingbiomass.org

**Project Goals: In this study we investigate an alternative approach to pretreatment that doesn't involve cellulose decrystallization (to amorphous cellulose) using expensive chemicals that are difficult to recycle to enhance cellulose depolymerization kinetics, but rather a subtle structural conversion between crystalline forms catalyzed by ammonia. This ammonia–based pretreatment produces cellulose $III_I$ (without any relevant loss of crystallinity) with enzymatic hydrolysis rates comparable to amorphous cellulose. We study the impact of this structural modification of cellulose on both cellulase binding and their synergistic activity using enzymes isolated from a well known cellulose–degrading fungus, *Trichoderma reesei*. We also complement our experimental study with extensive molecular dynamics (MD) simulations on model fibrils of cellulose $I_\beta$ and cellulose $III_I$. The fundamental insights gained from this combined experimental–theoretical approach will be critical to guide the development of improved ammonia–pretreatment processes and novel engineered cellulases that are optimized for rapid and efficient hydrolysis of ammonia treated lignocellulosic biomass.**

Deconstruction of lignocellulose for biofuel applications is rather inefficient due to the deleterious impact of cellulose crystallinity on enzymatic saccharification rates. We demonstrate how the synergistic activity of cellulases was enhanced by altering the hydrogen–bonding network within crystalline cellulose fibrils. We provide a molecular–scale explanation of this phenomenon through molecular dynamic (MD) simulations and enzymatic assays. Ammonia was used to convert the naturally occurring crystalline allomorph Iβ to IIII that caused a decrease in the number of cellulose intrasheet hydrogen bonds and increase in the intersheet hydrogen bonds. This rearrangement of the hydrogen–bond network within cellulose IIII, which increased the number of solvent–exposed glucan chains hydrogen–bonds with water by 50%, resulted in enhancing saccharification rates by upto five fold (comparable to amorphous cellulose) while interestingly reducing the maximum surface bound cellulase capacity by 60-70%. The enhancement in cellulase activity was attributed to the amorphous–like nature of the cellulose IIII fibril surface that facilitated easier glucan chain extraction. Unrestricted substrate accessibility to active–site clefts of certain endocellulase families further accelerated deconstruction of cellulose IIII. Structural and dynamical features of cellulose IIII, revealed by MD simulations, gave additional insights into the role of cellulose crystal structure on surface hydration that influenced interfacial enzyme binding. Subtle alterations within cellulose hydrogen–bonding network provides an attractive solution to enhancing its deconstruction and offers a unique insight into the nature of cellulose recalcitrance that can lead to unconventional pathways for development of novel pretreatments and cellulases for cost-effective biofuel production.

Student Oral Presentation–Monday

# 36

## Functional Genomics Study of Fatty Acid Overproduction in *Escherichia coli*

**Brian F. Pfleger**,[1] Rebecca M. Lennen[1]* (lennen@wisc.edu), Max A. Kruziki,[1] Kritika Kumar,[1] Ruwan Ranatunga,[2] Tyler M. Wittkopp,[2] and Wesley D. Marner II[2]

[1]University of Wisconsin-Madison, Department of Chemical and Biological Engineering; and [2]Department of Energy Great Lakes Bioenergy Research Center

http://www.glbrc.org/

**Project Goals: A baseline strain of *E. coli* has been metabolically engineered to overproduce medium-chain length free fatty acids, however yields in our strain and other published strains that overproduce free fatty acids are still well below the maximum theoretical yield. As one part of a strategy to improve production levels, a functional genomics study was undertaken to determine differences in transcript, protein, and selected metabolite levels between a fatty acid overproducing strain and a control strain. Insights gained from this study on induced stress responses and metabolic/physiologic perturbations are being used to guide the next stage of strain modifications.**

Free fatty acids have the potential to be a useful intermediate for chemical or biological conversion to high energy density liquid fuels. We previously engineered a fatty acid overproducing strain with a deletion of *fadD* to eliminate β-oxidation; and heterologous expression of the *Umbellularia californica* acyl-acyl carrier protein (ACP) thioesterase, which hydrolyzes free fatty acids from acyl-ACP intermediates. This strain exhibits an approximately eight-fold

increase in fatty acid production over a strain only deficient in β-oxidation. Numerous rationally-guided strain modifications have proven unable to improve production levels over the initially engineered strain. As part of an effort to further approach the maximum theoretical yield of fatty acids from lignocellulosic or other carbon sources, a functional genomics study was undertaken to identify differences in gene and protein expression profiles between a fatty acid overproducing strain and a non-overproducing strain. By comparing transcriptomic and proteomic data collected from overproducing strains grown under two different sets of conditions, a few key stress responses and metabolic perturbations have been identified that could serve as targets for a second iteration of strain engineering. Identified stress responses include strong induction of the phage shock system, which is induced by conditions that depolarize the inner membrane, and induction of the MarA/Rob/SoxS regulon, which is involved in counteracting oxidative stress and stress induced by exposure to solvents, detergents, and antibiotics. Some metabolic perturbations identified include strong down-regulation of genes involved in unsaturated fatty acid biosynthesis and a marked increase in long-chain unsaturated fatty acid content, increased levels of several genes and proteins involved in β-oxidation, and a decrease in acetate levels coupled with increased protein levels of acetyl-CoA synthetase.

# 37

## Enhancing Energy Density of Biofuel Crops by Engineering Oil in Vegetative Tissues

Sanjaya*[1,2] (sanjaya@msu.edu), **Christoph Benning**,[1,2] John Ohlrogge,[2,3] Yair Shachar-Hill,[2,3] Timothy Durrett,[2,3] Aruna Kilaru,[2,3] Que Kong,[1,2] and Mike Pollard[2,3]

[1]Dept. of Biochemistry and Molecular Biology, Michigan State University; [2]Great Lakes Bioenergy Research Center; and [3]Dept. of Plant Biology, Michigan State University

http://www.glbrc.org/

**Project Goals:The GLRBC plant oil project aims at increasing the content of high energy compounds in vegetative tissues of plants. The concept of enhancement of energy density in plant biomass is synergistic with other GLBRC efforts to develop lignocellulosic feedstocks for biofuel. Enhancing the energy yield of plant biomass can be achieved by accumulating energy-dense compounds such as triacylglycerol (TAGs). After extracting oil, the remaining lignocellulosic feedstock can be used for processing and fermentation. Ideally, next generation biofuel crops will accommodate multiple strategies by providing feed stocks for efficiently producing biodiesel, ethanol, and/or other fuels. The plant oils group collaboratively addresses these aims.**

An initial focus is on altering carbon partitioning from starch to oil in the storage root of rutabaga. Rutabaga like canola or *Arabidopsis* has an active pathway of storage oil biosynthesis in the embryo that will have to be activated and optimized in the developing root storage organ. A multifaceted approach is employed that involves the expression of multiple trans-genes in the rutabaga root to inhibit starch accumulation, increase the conversion of sugars into fatty acids, and to create a sink by enhancing oil synthesis. To rapidly test the underlying strategies, we engineered *Arabidopsis thaliana* to ectopically overproduce the transcription factor WRINKLED1 (WRI1) involved in the regulation of seed oil biosynthesis. Furthermore, we reduced the expression of APS1 encoding a major catalytic isoform of the small subunit of ADP-glucose pyrophosphorylase involved in starch biosynthesis using an RNAi approach. The resulting AGPRNAi-WRI1 lines accumulated less starch and more hexoses. In addition, these lines produced 5.8-fold more oil in vegetative tissues than plants with WRI1 or AGPRNAi alone. Numerous oil droplets were visible in vegetative tissues. TAG molecular species contained long-chain fatty acids, similar to those found in seed oils. The relative contribution of TAG compared to starch to the overall energy density increased 9.5-fold in one AGPRNAi-WRI1 transgenic line consistent with altered carbon partitioning from starch to oil. In addition, the transgenic *Arabidopsis* lines resulted in 10% per DW oil (TAG) on medium supplemented with 3% sugar. Transgenic rutabaga lines with the above constructs were generated and molecular and biochemical analysis is in progress. Overexpression of other B3 domain transcription factors like FUS3 and LEC1 has also shown considerable accumulation of TAG in the vegetative tissues. Currently, we are working on microarray experiments to reveal new oil regulatory mechanisms in vegetative tissues.

Moreover, we are using biodiversity and deep transcriptional profiling techniques for the discovery a novel oil regulatory genes. For example, we discovered the novel EaDAcT (*Euonymys alatus d*iacyl*g*lycerol *a*cetyl*t*ransferase) from burning bush that synthesizes acetyl-glycerols. Expression of EaDAcT under the control of a strong, seed-specific promoter in *Arabidopsis* resulted in the accumulation of acTAGs, up to 40 mol % of total TAG in the seed oil. These novel oils are low viscosity and therefore can be used directly in some diesel engines. The transcriptional profiling of oil palm and date palm revealed several interesting key regulatory factors involved in oil biosynthesis. Functional analysis of these genes in model system is in progress. The development of novel strategies to address compartmentalization of oil metabolism by metabolic flux analysis is in progress.

# 38

## Development of a Microarray Platform for Profiling Gene Diversity in Agricultural and Grassland Soil Microbial Communities

David Duncan* (dsduncan@wisc.edu) and **Randall Jackson** (rdjackson@wisc.edu)

Agronomy Department, University of Wisconsin-Madison and DOE Great Lakes Bioenergy Research Center

http://agronomy.wisc.edu/jackson/

**Project goals:**

1. **Develop a microarray platform for profiling functional genetic diversity of microbial communities in agricultural and grassland soils.**
2. **Explore the effects of management intensity on gene diversity.**
3. **Evaluate the capacity of microbial gene diversity patterns to predict or explain variation in greenhouse gas emissions and nutrient cycling in agroecosystems.**

The use of low-input, spatially extensive bioenergy cropping systems is predicated on managing vast areas for biomass harvests, creating a need for high throughput methods for diagnosing agroecosystem health and function. One potential avenue for such diagnostics involves monitoring the soil microbial community under managed agroecosystems. Soil microbes serve as both integrators of biologically-relevant ecosystem dynamics, and as drivers of many processes, such as nutrient cycling and greenhouse gas emission, that aggregate to become relevant at the management scale. Through this project we are developing a tool for monitoring the gene diversity of soil microbial communities and evaluating the capacity of this tool to predict greenhouse gas emission and nutrient cycling properties of model agroecosystems under consideration as biomass sources.

While other functional gene array platforms already exist, this project will focus exclusively on sequences derived from agricultural and grassland soil samples, which will permit the inclusion of a broader range of sequence variations while limiting the number of sequences not present in the soil. In addition, we will be taking a data-driven approach of including gene families whose diversity and abundance differs among sites sampled, rather than restricting the gene families to those genes that have been typically used in the literature as indicators of function. To do this, we are analyzing metagenomic datasets generated from bulk soils and rhizospheres in model cropping systems in Wisconsin and Minnesota. Analysis of these arrays will emphasize patterns of diversity within individual gene families, to determine whether certain conditions cause selection for a set of specific gene variants, or lead to a more even distribution across variants.

The predictive power of these arrays will be evaluating using the biogeochemical cycling information being generated through the GLBRC's bioenergy cropping systems trials. Through these trials, frequent measurements of greenhouse gas emissions and nutrient pools are being taken for a range of potential cropping systems including conventionally grown corn, switchgrass monocultures, and diverse assemblages of native prairie species. By comparing the patterns of diversity observed in the microbial community in these systems to their greenhouse gas emission and nutrient cycling properties, we will determine to what extent changes in the gene diversity of the soil microbial community can predict manage-relevant function.

Funding source: DOE Great Lakes Bioenergy Research Center; DOE BER Office of Science DE-FC02-07ER64494; DOE OBP Office of Energy Efficiency and Renewable Energy DE-AC05-76RL01830

# 39

## Improving the Sustainability of Bioenergy Crops Through Arbuscular Mycorrhization

Arijit Mukherjee* (amukherjee@wisc.edu), Shawn Kaeppler, and **Jean-Michel Ané**

Department of Agronomy, University of Wisconsin, Madison; and Great Lakes Bioenergy Research Center, University of Wisconsin, Madison

http://www.glbrc.org/

**Project Goals: Arbuscular mycorrhization (AM) is the most beneficial symbiosis between microbes and bioenergy crops. The goal of our project is to characterize the signals and the genes controlling the establishment of AM in bioenergy crops. The specific objectives are:**

1. **Identify plant genes controlling AM in monocots and analyze their regulation by biotic and abiotic stresses.**
2. **Test the plant responses to AM signals (Myc factors) and the influence of stresses on these responses.**

Availability of water and nutrients is a major constraint for crop productivity and sustainable agriculture. Over the last decades, there has been an excessive dependence on chemical fertilizers with major economic, ecological and heath consequences. Taking better advantage of plant-microbe symbioses like arbuscular mycorrhization (AM) seems a reasonable alternative to improve crop yields and the sustainability of our agricultural systems. All the major bioenergy crops can form associations with AM fungi that improve their acquisition of water and nutrients (especially phosphorus). AM is the most efficient symbiosis between soil microbes and bioenergy crops. On a global level, AM symbiosis contributes significantly to phosphate, nitrogen and carbon cycling. Therefore, improving the efficiency and the development of AM associations especially under sub-optimal conditions has a tremendous potential for improving the sustainability of biofuel production. Our goal is to characterize the signals and the genes controlling the establishment of AM in energy crops. We developed a high-throughput screening of maize mutants affected in AM symbiosis. We screened more than 3850 lines of mutagenized population of B73 already and identified five

mutants which are unable to establish AM. The absence of AM in these mutants was confirmed by microscopy. These mutants have been self-fertilized to produce M3 progenies for further phenotypic characterization and crossed to a polymorphic parent for positional cloning. The screening of additional maize lines is ongoing. In order to characterize the signals produced by AM fungi, we developed an easy procedure to collect such diffusible signals in germinating spore exudates; these stimulate plant growth in monocots (maize, rice) and eudicots (alfalfa). They also induce expression of AM-specific genes through several genetic pathways and this induction is negatively regulated by the stress hormone, ethylene. In order to analyze Myc factors-induced early responses in maize at the transcript level and their regulation by ethylene, microarray experiments were performed. Preliminary data analysis indicates that Myc factors induce gene expression in maize and ethylene inhibits this gene expression. We have selected genes of interest to validate microarray results using qRT-PCR.

# 40

## A Lattice Monte Carlo Model of Substrate Channeling on Catalytic Scaffolds

He Zhang* (hzhang32@wisc.edu) and **George N. Phillips Jr.**

Department of Biochemistry and Great Lakes Bioenergy Research Center, University of Wisconsin-Madison

http://www.glbrc.org/

Substrate channeling is a process by which two or more sequential enzymes in the same pathway interact to transfer an intermediate from one enzyme active site to another without allowing free diffusion of the intermediates into bulk solvent. It has been indicated that substrate channeling is of fundamental importance in regulating metabolic and signaling pathways in cell. A growing number of studies have been performed by engineering the spatial organization of enzymes to mimic nature's synergy. These artificial systems have important implications for the overall efficiency, specificity and complex regulation of metabolic and signaling pathways. We have developed a highly modular and coarse-grained lattice model by combining the Bond Fluctuation Lattice Model and Hard-sphere Particle Model. The hybrid method employs the Metropolis Monte Carlo algorithm for both reaction and diffusion processes. With the new tool, we generated a model of a synthetic catalytic scaffold system. We tested roles of different scaffold architectures in determining substrate channeling probabilities. The prediction suggests branched scaffolds have higher efficiency than linear scaffolds, which is the only topology tested by the experiments. Rearranging catalytic domains can further increase the channeling probability of both single-step and multi-step reactions. Simulations were applied to both diffusion-controlled and reaction-controlled systems. This new model can help to understand and optimize similar modular systems like synthetic scaffolds, signaling scaffolds and biomass degradation systems, which have important value in medical and industrial applications.

# 41

## Metaproteomics: Development of Hybrid Database-Spectral Library Searches and Optimization of Proteome-Spectra Matches as an Alternative to De Novo Sequencing

**William R. Cannon**,[1]* Alys Hugo,[1] Mitchell M. Rawlins,[1] Douglas J. Baxter,[1] Gaurav Kulkarni,[1,2] Ananth Kalyanaraman,[2] Kristin Burnum,[1] Frank O. Aylward, [3,4] Jarrod J. Scott,[3,4] Garret Suen,[3,4] Gabriel J. Starrett,[3,4] Cameron R. Currie,[3,4] Mary S. Lipton,[1,4] Stephen J. Callister,[1] and Donald A. Bryant[5]

[1]Pacific Northwest National Laboratory, Richland, Wash; [2]Washington State University, Pullman; [3]University of Wisconsin-Madison; [4]Great Lakes Bioenergy Research Center, Madison Wis.; and [5]The Pennsylvania State University, University Park

http://www.glbrc.org/

**Project Goals: This project is to develop novel methods for global proteomics and analysis of environmental samples. First, we present a hybrid spectral library-database search for identifying peptides from CID MS/MS spectra that increases the number of spectra that are identified by 60%-147%. The large increase in identifications is a result of several factors. First, we employ a probabilistic method for incorporating intensities into the identification process that sufficiently allows for variability in intensities between spectra. This enables the use of model spectra that have a wide range of fidelity to the experimentally observed spectrum. Second, the nature of the statistical distributions of scores obtained using the different model spectra allows us to analyze the error rates using a single method regardless of whether the model spectrum was obtained from a spectral library or a database search.**

**Second, is the identification of proteins from unknown (environmental) samples from MS/MS spectra using a novel database search strategy. The method provides an effective way to control the false discovery rate for environmental samples and provides an alternative to de novo peptide sequencing. Furthermore, the method can obviate the need to use DNA-based identification methods to find appropriate genomes when proteomic characterization is the primary goal and sub-species identification based on ribosomal phylogeny is not needed.**

**Results**

*Hybrid Search to Identify Peptides and Proteins.* A hybrid spectral library-database search of MS/MS peptide spectra was made possible by observing that both multinomial data analysis and statistical thermodynamics use the same statistical likelihood function, which not only provides a physically and chemically principled way to incorporate

abundances into the interpretation of spectra, but also allows for an integrated estimation of the FDR.

In exhaustive testing, many more spectra were matched to peptides using a hybrid spectral library-database search than with a database search alone. When using a small spectral library we observed an approximately 60% improvement in the number of identified peptides at a 5% False Discovery Rate, and a 147% improvement while using a larger library. In multiple studies, the improvement in identification rate clearly increased with the size of the spectral library. Additional room for performance gains was also implied by statistical tests evaluating the quality of the model spectra.

We applied this method to a global proteomics study of the marine bacterium *Synechococcus* sp. PCC 7002, a marine, unicellular cyanobacterium that performs oxygenic photosynthesis and is involved in carbon sequestration and cycling. For *Synechococcus*, the hybrid search increased the number of identifiable spectra by 125% at a 5% FDR. The use of spectral libraries resulted in additional peptide/protein identifications that are involved in cellular responses to important physiological processes, such as photosynthesis and $CO_2$ fixation.

The number of peptides identified for photosystem I increased 160%, while the number of peptides associated with the pentose phosphate pathway increased 250%. Furthermore, increases in identifications of 50-60% were observed for proteins in pathways related to photosynthesis and $CO_2$ metabolism—light harvesting for photosystem II, chlorophyll biosynthesis, $CO_2$ fixation (Calvin Benson Cycle), $CO_2$ uptake, and photorespiration. Notably, there was a 30 000% increase in the number of peptides identified for the $CO_2$ transporter of the ICT family.

*Optimization Approach to Metaproteomics*. Global proteomics of environmental samples is challenging because the usual database search approach used to identify peptides and proteins is difficult to apply because the identity of the microorganisms has not been made and, therefore, even if the organism has been sequenced it is not known which genome to search. Alternatively, applying a standard database search using large protein databases, such as NR, results in a large number of matches by chance. As an alternative to de novo sequencing, we developed an optimization method for identifying peptides and proteins from fully sequenced microbial genomes. The method uses high performance computing to optimize proteome-spectra matches and iteratively eliminates microbes that are not likely to be in the sample.

The method has been tested using samples containing blind mixtures of spectra from known microbes and samples containing unknown mixtures of microbes. In the case of five blind mixtures of varying complexity, the method has been able to identify the correct microbes reliably. In addition, the spectra identified with each microbe has a high overlap with spectra identified at a 5% FDR when searching only the known organisms protein sequences.

The optimization method was applied to samples from the leaf-cutter ant fungal garden, which has a microbiome with a high plant biomass degrading capability. These ants avoid the ingestion of toxic plant chemicals by having a fungus perform digestion of the cut leaves to less complex sugars. However, the fungus is susceptible to pathogenic molds of the genus Escovopsis. In response, the ants cultivate the presence of antibiotic-producing bacteria in the fungal garden. The optimization-based proteomic analysis revealed the presence of species related to the genomes of multiple antibiotic-producing microbes, in addition to the presence of cellulose degrading species.

# 42

## A Rapid-Development Environment to Produce Custom Genomic Data Management and Analysis Applications

Nicholas Thrower[1], Drew Bomhoff[1], Matthew Larson[1], Steve Slater[2], and **Curtis Wilkerson***[1,3] (wilker13@msu.edu)

[1]DOE Great Lakes Bioenergy Research Center, East Lansing, Mich.; [2]DOE Great Lakes Bioenergy Research Center, Madison, Wis.; [3]Department of Plant Biology, Michigan State University, East Lansing

http://www.glbrc.org/

**Project Goals: We are using modern web 2.0 tools to produce software for the analysis and management of large-scale datasets.**

We believe that highest productivity is achieved when software tools are adapted to the idiosyncratic needs of individual research groups. The GLBRC has developed a genomic data management suite to facilitate the use of high-throughput sequencing data. Our investigators are able to view, manage and interact with experimental results through a secure web accessible interface. The application was developed using the popular Ruby on Rails (RoR) web framework. Our design goal was to create an easily customizable system that could be altered to suit individual research group's needs. Relying on RoR conventions allowed for rapid development of such a system. Framework plugins were used for tasks such as authentication and auditing. In RoR the Model-View-Controller pattern underlies the core architecture and enforces strict design patterns tailored for agile development. Models are a proxy to the underlying BioSQL schema and indexed binary experiment files. This schema was chosen for its community support and ability to represent our data types. Sequence and annotations can be directly accessed through bioSQL bindings within bio-ruby, bio-perl and bio-python. Binary files were chosen to store large datasets after investigating options to store the data in relational form. Insert time, index maintenance and resulting query performance were all considered in this decision. Using community supported formats, such as BigWIG and BigBED from UCSC, data can be uploaded in a fraction of the time it takes to parse and insert results into a relational schema. This application has multiple complex views includ-

ing a genome sequence display, experimental data views and individual gene reports. The sequence view is based on the Anno-J javascript interface, which relies heavily on the Ext3 javascript library for graphical user interface elements and design. Experimental data views are rendered using a combination of html and Google visualizations allowing data to be assessed without the need to load genomic context. Gene views are organized around globally unique locus_tag identifiers and grant detailed access to genomic annotations. Annotations can be viewed, modified, created and exported in different formats for use in other applications.

Several design patterns were followed within the application besides the overarching Model-View-Controller organization. The strategy pattern is prevalent throughout the system. The strategy pattern allows for a variety of re-usable painting algorithms, interchangeable tracks, and experiment types. An observer pattern is used to manage interaction with the sequence view. Javascript events such as mousedown, and keydown are broadcast and registered listeners respond accordingly. Experiments follow the composite pattern allowing for the combination of datasets to create new synthetic experiments. Operators such as mean, max, sum, difference and quotient can be used while compositing objects. The sequence view is built with a mediator pattern. The mediator object links components together and de-couples the track browsing tree, track navigation area and visual track displays. Many other patterns are followed within the various software packages used including adapters, prototypes, decorators, flyweights and templates.

Careful consideration was given to the performance of the application and our observations led us to use the indexed binary data format. The C routine wigToBigWig was used to convert the datasets to an indexed file in ~1 minute. The newly created binary files can then be queried with another executable, bigWigSummary. The performance of bigWigSummary is excellent returning results in milliseconds. Many optimizations were implemented to extend the usability of the sequence view interface. Data is delivered in incremental pieces large enough to allow for smooth scrolling and small enough to load in less than 1 second. Query results from gene models, features, and experiments are also cached for immediate display of previously viewed data. When dealing with large datasets, the instantiation of class objects is bypassed. Instead, JSON strings are created directly from queries and delivered to the javascript client for display. The number of datapoints returned is limited by the view and varying levels of detail are sent as users change the viewing scale.

# 43

## Biomass Trait Analyses of Diverse Wild Type *Brachypodium* Accessions

Cynthia L. Cass[1,2*] (ccass@ilstu.edu), Megan Galvin,[1,2] Jackie Brown,[1,2] Mike Krzyskowski,[1,2] Laura Severs,[1,2] Samantha Sakol,[1,2] Nick Santoro,[3] David F. Garvin,[4] and **John C. Sedbrook**[1,2]

[1]School of Biological Sciences, Illinois State University, Normal; [2]Great Lakes Bioenergy Research Center, Madison, Wis.; [3]Great Lakes Bioenergy Research Center, Michigan State University; and [4]USDA-ARS Plant Science Research Unit, St. Paul, Minn.

**Project Goals: Quantify biomass trait differences between a diverse set of wild type *Brachypodium* accessions and employ genetic and biochemical approaches to identify novel genes and gene variants affecting those traits.**

Large perennial grasses such as switchgrass and *Miscanthus* hold much promise as next generation bioenergy crops owing to their high biomass yields, hardiness, and relatively low fertilizer requirements. The vast majority of that biomass is composed of secondary cell wall material, a complex structure comprised of crystalline cellulose cross-linked with hemicellulose and lignin polymers. While progress is being made in breeding varieties of switchgrass for improved biomass traits including higher yields and a cell wall composition and structure more amenable to deconstruction, that work is hampered by the polyploid nature of these grasses and their long generation times. Hence, the diminutive wild grass *Brachypodium distachyon* is being utilized as a tractable model system in which to rapidly identify novel biomass trait genes and gene variants. Subsequently, this knowledge can be translated into improving bioenergy crop grasses.

In exploring the utility of *Brachypodium* as a bioenergy crop model, we are phenotypically and genotypically characterizing the relevant biomass traits of a diverse set of wild type *Brachypodium* accessions. Comparing plants from a core set of seven diverse inbred accessions, we have identified statistically significant and reproducible accession differences in plant height, aboveground mass, cell wall composition (hemicellulose and lignin), and enzymatic digestibility of ground stem tissue that were processed by a variety of pretreatment conditions. Interestingly, we have found that plant height correlates with biomass digestibility, with the biomass of taller accessions exhibiting significantly higher digestibility. In order to determine how many loci influence this observed accession variation, we are analyzing recombinant inbred lines generated from a cross between two accessions with the largest difference in biomass digestibility. These results will be presented along with a discussion of the relevance of these findings.

# 272‡

## Microbial Community Analysis of Soil and Rhizosphere of Biofuel Crops Done With Next Generation Sequencing

Adina C. Howe,[1] Erick Cardenas,[1] Ederson da C. Jesus,[1] John Quensen,[1] Patrick Chain,[2,3] Jordan Fish,[1] Qiong Wang,[1] James Cole,[1] Susannah Tringe,[2] and **James M. Tiedje**[1]*

[1]Michigan State University, East Lansing; [2]DOE Joint Genome Institute, Walnut Creek, Calif.; and [3]Los Alamos National Laboratory, Los Alamos, N.M.

In order to understand the plant-microbe-soil relationships of biofuel crops, we studied the taxonomic and functional composition of microbial communities under switchgrass and Miscanthus in Michigan and Wisconsin (Figure 1a). Both rhizosphere and bulk soil microbial community metagenomes from switchgrass and *Miscanthus* stands have been sequenced, assembled, and annotated in the IMG data management system. A total of approximately 3.2 billion reads were obtained from the 454 and Illumina GAII sequencing platforms (Figure 1b). Assemblies combining 454 and Illumina reads were evaluated based on contig GC content, size, and coverage.

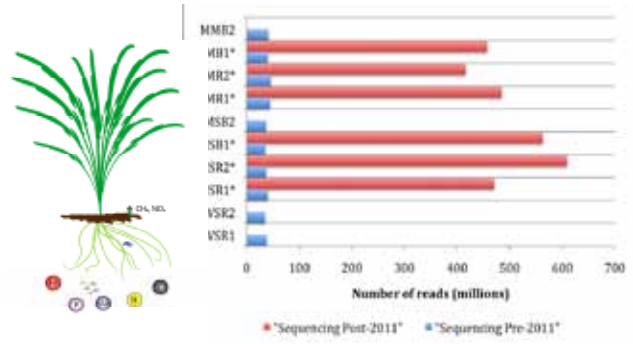



Figure 1a (left). Conceptual model of plant-microbe interactions related to bioenergy sustainability efforts. We target genes mostly related to phosphorus, nitrogen, and iron cycling, as well as plant growth promoting genes. Figure 1b (right). Number of metagenomic short reads sequenced from switchgrass and Miscanthus soils. (M/W=Michigan/Wisconsin, M/S= Miscanthus/Switchgrass, B/R=Bulk/Rhizosphere, 1/2=Replicate ID).

Community profiles based on distribution of cluster of orthologous genes (COGs) of 454 short reads suggest that soil communities shared similar profiles regardless of location (Michigan or Wisconsin), ecosystem (bulk or rhizosphere), or replicate. Composition of COGs present in all rhizosphere and bulk soil samples showed that both habitats shared most of their conserved genes, however rhizosphere and bulk soils showed prevalence of genes involved in the nitrogen cycle and phosphorus cycle, respectively (Figure 3).

Additionally, we compared amplicon sequencing of the *nifH* gene (using primers designed by Poly et al, 2001) to *nifH* sequences recovered from Illumina shotgun sequencing of soil bacterial communities. For Illumina shotgun sequencing of related Midwest soils, we estimated one detectable *nifH* per 370 bacteria. The diversity of *nifH* genes from shotgun sequencing was greater than that of amplicon sequencing, suggesting that Poly primers were not comprehensive but also suggesting where these primers could be improved (Figure 4). These improvements would lead to sequences similar to those of the primers designed Zehr et al. (1989), which are more degenerate than the Poly primers but target the same region.

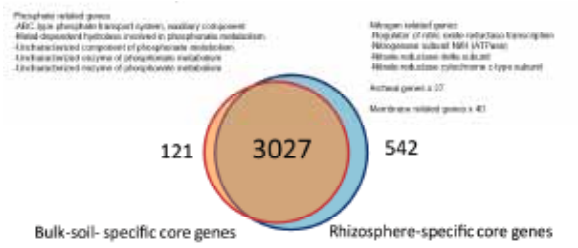



Figure 3. Comparison of COGs identified from 454 short reads from bulk and rhizosphere soil communities.

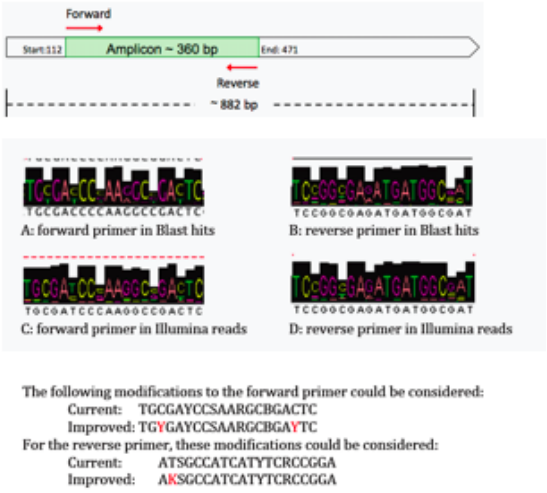



Figure 4. Primers used in 454 Titanium sequencing of *nifH* gene, using *Trichodesmium thiebautii* as an example. Consensus sequences of the primer region from top Blast NifH hits and Illumina reads suggest improvements to current primers.

This work will allow us to further explore genes involved in plant growth promotion, carbon, nitrogen, and phosphorus cycling contained in our metagenomic sequences, which combined with site-specific environmental metadata, can be used to explore the effects of gene suite and habitat on plant-microbe-soil relationships.

**Reference:**

1. Poly, F., L. Monrozier, et al. (2001). Improvement in the RFLP procedure for studying the diversity of *nifH* genes in communities of nitrogen fixers in soil. Res. Microbiol. 152:

95-103; Zehr, J. P. and L. A. McReynolds (1989). Use of degenerate oligonucleotides for amplification of the nifH gene from the marine cyanobacterium *Trichodesmium thiebautii*. Appl. Environ. Microbiol. 55: 2522.

For more information: Please visit http://www.glbrc.org.

## Joint BioEnergy Institute (JBEI)

# 44

### Development of a High Throughput Pre-Treatment and Saccharification Protocol to Screen Plant Biomass

Miguel E. Vega-Sánchez[1]* (mevega-sanchez@lbl.gov), Mawsheng Chern,[1] Brad Holmes,[2] Wing Hoi Sze-To,[1] Masood Hadi,[3] **Blake Simmons**,[2] and **Pamela Ronald** [1]

[1]Feedstocks, [2]Deconstruction and [3]Technology Divisions, Joint BioEnergy Institute, Emeryville, Calif.

http://www.jbei.org/

**Project Goals: We seek to identify genes that control grass cell wall recalcitrance. For this purpose, we have developed a high-throughput pre-treatment and saccharification screen to isolate mutants affected in the release of fermentable sugars.**

Understanding plant cell wall biosynthesis is crucial for the development of the next generation of biofuels derived from lignocellulosic material. Current limitations in the harvest of fermentable sugars from cellulose derive from the inherent recalcitrance of plant cell walls. Basic knowledge of how the structure and composition of the cell wall can be modified to obtain biomass suitable for efficient and economically viable biofuel production is needed. We are using a forward genetics approach to identify genes responsible for cell wall characteristics affecting the deconstruction of the plant cell wall. By means of fast neutron mutagenesis, we have generated a rice mutant population consisting of 6,500 M0 lines and harvested more than 100,000 M1 seed from approximately 4,000 M0 plants. Stems from these lines have been collected and we are in the process of screening them for alterations in saccharification efficiency. To screen for changes in fermentable sugar release from rice stems, we have optimized a protocol using either hot water or dilute acid pre-treatment followed by enzymatic saccharification for adaptation into a 96 well format. Once cell wall mutants are confirmed, we will extract DNA from wild type and highly prioritized mutant candidates and then carry out whole genome comparative hybridization on rice tiling arrays. This approach will allow us to identify genes in deleted region responsible for the mutant phenotypes. Mutants will be complemented with candidate genes using transgenic analysis and assayed for restoration of the cell wall phenotypes.

# 45

### Improving Biofuel Production by Reducing *O*-Acetylation of Cell Wall Polysaccharides

Yuzuki Manabe[1]* (ymanabe@lbl.gov), Yves Verhertbruggen,[1] Soe Myat Htwe,[1] Emilie A. Rennie,[1,2] Mario Ouellet,[1] Sascha Gille,[2] Markus Pauly,[2] Aindrila Mukhopadhyay,[1] Bradley Holmes,[1] Seema Singh,[1] Blake Simmons,[1] Dominique Loque,[1] and **Henrik Vibe Scheller**[1,2]

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.; and [2]Department of Plant and Microbial Biology, University of California, Berkeley

http://www.jbei.org/

**Project Goals: *O*-acetylation of polysaccharides in lignocellulosic biomass inhibits enzymatic degradation of the polysaccharide and its fermentation into fuels. Our goals are to better understand the mechanism of acetylation and use this knowledge to generate genetically modified feedstocks with lower acetylation.**

We have recently identified Reduced Wall Acetylation2 (RWA2), the first putative polysaccharide *O*-acetyltransferase in plants. *Arabidopsis* rwa2 loss-of-function mutants are indistinguishable from wildtype but display 20% reduction in acetylation of overall cell wall, pectin and xyloglucan in leaves. The relatively small change in acetylation and lack of phenotype is apparently due to functional redundancy among the four RWA genes. RWA double mutants display no or minimal morphological phenotype while the acetylation level is reduced up to 25%. In contrast, triple and quadruple rwa mutants have severe growth defects. These severe growth phenotypes imply the importance of *O*-acetylation in planta. The mechanism by which RWA proteins mediate acetylation is not understood, and we are investigating this by expressing the proteins in tobacco and in yeast. We are currently working towards assessing the effect of reduced acetate in rwa double mutants on enzymatic saccharification and subsequent biofuel fermentation by *Saccharomyces cerevisiae.*

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.

# 46

## Identification and Characterization of Monolignol Transporters in *Arabidopsis*

F. Yang,[1]* K. Zheng,[1] P. Pradhan-Mitra,[1] J.S. Kim,[1] L. Sun, A. Eudes,[1] P. Varanasi,[2] P. Benke,[3] A. Mukhopadhyay,[3] B. Holmes,[2] S. Singh,[2] B. Simmons,[2] J.K. Keasling,[3] and **D. Loqué**[1]

[1]Feedstocks Division, [2]Deconstruction Division, and [3]Fuels Synthesis Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.

http://www.jbei.org

Currently, biofuels such as ethanol are produced largely from starch contained in grains. But, this represents only a small proportion of sugar polymer availability on Earth. Large quantities of sugar from polysaccharides that are not utilized thus far are cellulose and hemicellulose, which are the main constituents of plant cell walls. The third main constituent of plant cell walls is lignin, a strong aromatic polymer recalcitrant to degradation. Lignin inhibits efficient extraction and hydrolysis of cell wall polysaccharides and prevents cost-effective lignocellulosic-biofuel production. Unfortunately, lignin cannot simply be genetically removed without incurring deleterious consequences on plant productivity. The lignin polymer provides structural support to the plant and protects the plant against biotic and abiotic stresses. Therefore, it is important to develop strategies to control lignin deposition and composition to reduce its recalcitrance without disturbing some of its key functions, in order to maintain plant yield and to increase the effectiveness of sugar recovery from the plant cell wall.

Several strategies are currently under development, in order to modify lignin composition or deposition. Some focus on the biosynthesis of "novel monolignols" and their incorporation into the lignin structure to perturb its recalcitrance. In contrast to monolignol biosynthesis, little is known about the mechanism mediating monolignol export into the apoplast. Two routes could be utilized: a direct export at the plasma membrane or a vesicular mediated export mechanism. In both cases, transporters either mediating vesicular loading or plasma membrane export are required. Therefore, we developed a strategy using yeast complementation to identify protein mediating monolignol export and screened *Arabidopsis* cDNA libraries and a large transporter library mainly composed of MDR transporters. Several candidate genes were identified from the screen, and we are currently analyzing their biological relevance in the lignin deposition in *Arabidopsis*. Screening approach and preliminary data on the characterization will be presented.

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.

# 47

## Isolation and Proteomic Characterization of the *Arabidopsis* Golgi Apparatus

Harriet T. Parsons, Bernhard Knierim, Tanveer S. Batth, Christopher J. Petzold, Katy Christiansen, Jun Ito, Manfred Auer, Aindrila Mukhopadhyay, Henrik V. Scheller, and **Joshua L. Heazlewood*** (jlheazlewood@lbl.gov)

Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.

http://www.jbei.org

**Project Goals: Isolation and proteomic characterization of the *Arabidopsis* Golgi apparatus.**

The plant cell wall is comprised of complex sugar polymers including cellulose, hemicellulose and pectin. The Golgi apparatus within the plant cell synthesizes a significant proportion of these matrix polysaccharides prior to their incorporation into the cell wall. We have been isolating this compartment using density centrifugation and charge based separation on a Free Flow Electrophoresis system. Analysis of Golgi purified fractions from *Arabidopsis* cell culture by mass spectrometry after FFE separation indicates the method is suitable for isolation of this organelle from plants. We have identified around 450 proteins from these fractions and identified over 50 glycosyl transferases (from multiple families) whose major functions are likely involvement in matrix polysaccharide biosynthesis. These glycosyl transferases likely represent the core set of enzymes required for the biosynthesis of hemicellulose and pectin bound for the cell wall. Overall around half of the proteins in identified are of known or likely Golgi in origin; while about a third are unknown or are derived from the endosomal system and 10 – 20% appear to be contaminants from other organelles and membrane systems. This technique will enable us to commence in-depth comparative cell wall proteomics focusing on protein function and changes in the plant Golgi apparatus.

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.

# 48

## Feedstock-Adapted Anaerobic Consortia Derived from Tropical Forest Soils

Kristen M. DeAngelis,[1,2] Julian Fortney,[1] Sharon Borglin,[1] Whendee Silver,[1,3] and **Terry C. Hazen**[1,2*] (tchazen@lbl.gov)

[1]Ecology Department, Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]Joint BioEnergy Institute, Deconstruction Division, Microbial Communities Group, Lawrence Berkeley National Laboratory, Emeryville, Calif.; [3]Department of Ecosystem Science, Policy and Management, University of California, Berkeley; and *Joint BioEnergy Institute, Deconstruction Division, Microbial Communities Group, Lawrence Berkeley National Laboratory, Emeryville, Calif.

http://www.jbei.org

Tropical soils in Puerto Rican rain forests likely are capable of deconstructing biofuel plant materials to basic components, and frequent episodes of anoxic conditions make it likely that these decomposing consortia are primarily bacteria, not fungi as are usually observed in temperate systems. We cultivated feedstock-adapted anaerobic consortia (FACs) derived from Puerto Rico forest soils and added the terminal electron acceptors nitrate, sulfate, or iron to examine the effect on switchgrass deconstruction. Soils from two forest types were used as inoculum; short cloud forest (SCF) soils are perennially soaked, while Bisley Ridge soils (BisR) are more iron-rich and experience fluctuating redox. Soil communities were anaerobically passed through a succession of transfers in minimal media with switchgrass as the sole carbon source, and the FAC was established after the fourth transfer. Based on methane and carbon dioxide production rates, nitrate and iron caused the highest C mineralization in BisR-FACs, while switchgrass alone had the highest C mineralization in SCF-FACs. Specific enzyme activity rates were higher overall in SCF-FACs compared to BisR-FACs. Microbial community profiling was performed using PLFA and pyrotag sequencing of the small subunit ribosomal RNA gene, revealing Actinobacteria and Gammaproteobacteria as dominant organisms. Metagenomic analysis was performed on BisR-FACs from switchgrass only and iron-amended microcosms, revealing that the iron-amended FAC contained 324 distinct taxa compared to 81 taxa in the unamended. The diversity of anaerobic degraders found in these soils reiterates the importance of anaerobic decomposition in these environments and highlights the potential for discovery.

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.

# 49

## Characterization of CAZY-Like Enzymes from HT Sequencing of Microbial Communities

Michael J. Dougherty,[1] Patrik D'haeseleer,[2] Blake A. Simmons,[2] Paul Adams,[1] and **Masood Hadi**[1*] (mzhadi@sandia.gov, mzhadi@lbl.gov)

[1]Joint BioEnergy Institute, Technology Division, Lawrence Berkeley National Laboratory, Emeryville, Calif; [2]Joint BioEnergy Institute, Deconstruction Division, Lawrence Berkeley National Laboratory, Emeryville, Calif.

http://www.jbei.org/

Microorganisms from natural environments are a rich source of new biocatalysts, and recent advances in DNA sequencing technology have made the recovery of large numbers of gene sequences from the environment feasible. However, a major bottleneck still exists—the expression and characterization of the genes obtained from these environments is challenging due to effort required to identify appropriate expression hosts and/or conditions for protein purification and characterization.

Enzymatic hydrolysis of lignocellulose is currently one of the most expensive steps in processes for biofuel production. Identifying and/or engineering glycoside hydrolases (GHs) with improved enzymatic properties is a major research challenge in this effort. We have been using a metagenomics approach interrogating a switchgrass-adapted compost microbial community, we identified genes in this community that putatively encode enzymes with diverse activities, including endoxylanase, b-xylosidase, and a-arabinofuranosidase.

In order to validate the metagenomic approach for finding new biocatalysts, these ORFs have been cloned, expressed, and assayed for various hemicellulase activities which can then be prioritized for diverse biofuel process conditions. We have begun characterizing the genes that are well-behaved in *E. coli*, focusing on properties important for the process of biomass hydrolysis, such as thermostability, pH dependence, and ionic liquid tolerance.

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.

# 50

## Metagenomics, Proteomics, and Metabolic Reconstruction of a Thermophilic Feedstock-Adapted Bacterial Community

**P. D'Haeseleer**[1,2*] (pdhaeseleer@lbl.gov), J.M. Gladden,[1,2] A.M. Redding-Johanson,[3] C.J. Petzold,[1] P.I. Benke,[1] M. Allgaier,[4] D.C. Chivian,[1] J.S. Vandergheynst,[1,5] T.C. Hazen,[1,6] B.A. Simmons,[1,7] and S.W. Singer[1,6]

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.; [2]Lawrence Livermore National Laboratory, Livermore, Calif.; [3]Sapphire Energy, San Diego, Calif; [4]Leibniz-Inst. of Freshwater Ecology and Inland Fisheries (IGB), Leibniz, Germany; [5]University of California, Davis; [6]Lawrence Berkeley National Laboratory, Berkeley, Calif.; and [7]Sandia National Laboratory, Livermore, Calif.

http://www.jbei.org/

**Topic:** O10 Biofuels and Biobased Materials ; N11 Molecular Microbial Ecology—Communities

**Keyword:** Metagenomics; lignocellulosic biomass; biofuels

**Background:** Biomass degrading enzymes are a critical component of biofuel production from lignocellulosic feedstocks. Commercially available enzyme cocktails are poorly adapted to next-generation feedstocks, pretreatments and processing conditions. In previous work, we derived a range of novel enzymes from a switchgrass degrading compost community. Here, we focus on a thermophilic feedstock-adapted enrichment culture, which is better adapted to our desired processing conditions, yields a greater number of full-length high quality enzyme sequences, and is more amenable to community metabolic modeling.

**Methods:** An enrichment culture with high cellulase and xylanase activity was selected for 454 and Illumina metagenomic sequencing, annotated using the Joint Genome Institute's IMG/M system, and binned into phylogenetic groups using ClaMS. The supernatant was analyzed using zymograms and MS proteomics to identify biomass degrading enzymes. Draft metabolic reconstructions for the individual phylogenetic bins were generated using Pathway Tools, allowing us to assign metabolic roles to the different members of the bacterial community.

**Results:** Metagenome sequencing resulted in almost 65,000 gene calls, including hundreds of full-length biomass degrading enzymes of interest. Phylogenetic binning identified eleven thermophilic species, including *Thermus*, *Rhodothermus*, a novel member of the poorly sampled Gemmatimonadates phylum, *Paenibacillus*, *Conexibacter*, and *Thermobaculum*. Proteomics identified some of the abundant extracellular hemicellulases and cellulases, and suggested oxidative lignin degradation and sugar isomerase activity. Metabolic reconstruction highlighted the key roles of the *Rhodothermus*, *Paenibacillus* and *Gemmatimonadetes* species in degrading biomass components.

**Conclusions:** Focusing on a thermophilic feedstock-adatped enrichment community yields an order of magnitude more useful enzyme sequences, and the resulting enzymes are more likely to be well suited to our targeted feedstock, pretreatment, and processing conditions. Combining enzymatic assays, metagenomics, zymography, MS proteomics, and metabolic modeling provides a multidimensional view of the internal functioning of this highly active biomass degrading bacterial community.

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.

# 51

## Strategies for Improving Production and Resistance Phenotypes in Engineered Microbes

Suzanne Ma,[1,2*] Adrienne McKee,[1,2] Helcio Burd,[1,2] Alex Juminaga,[3] Alyssa Johanson,[4] Chris Petzold,[1,2] Wendy Schackwitz,[5] Greg Friedland,[1,2] Adam Arkin,[1,2,3] Dylan Chivian,[1,2] Taek Soon Lee,[1,2] Jay Keasling,[1,2,3] and **Swapnil Chhabra**[1,2*] (srchhabra@lbl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.; [3]University of California, Berkeley, Calif., [4]Sapphire Energy, San Diego, Calif.; and [5]Joint Genome Institute, Walnut Creek, Calif.

http://www.jbei.org/

**Project Goals: This poster describes research efforts in the Host Engineering Department of the Fuels Synthesis Division of JBEI. Overall research goals of this department as related to the model platforms, *Escherichia coli* and *Saccharomyces cerevisiae*, fall into four categories: (i) enhancement of production yields of strains engineered for advanced biofuel synthesis; (ii) improvement in resistance towards toxic pretreatment growth inhibitors; (iii) use of systems level understanding to combine aforesaid phenotypic improvements in a single strain; and, (iv) extension of knowledgebase generated from model platforms to non-model systems with desirable traits. To this end, we employ the tools of mathematical modeling and bibliomic information for local and global optimization of production and resistance phenotypes in engineered microbes. Specifically, we highlight our efforts in pathway component variation, multi-gene regulation and chemostat driven evolution to achieve the aforesaid objectives in *E. coli* engineered for isoprenoid-derived biofuels.**

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.

# 52

## Development of Fatty Acid-Based Fuels at JBEI: Alkenes and Fatty Acid Ethyl Esters

E.B. Goh, E.J. Steen, N.J. Hillson, J.W. Thorne, **H.R. Beller*** (HRBeller@lbl.gov), and J.D. Keasling

Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.

http://www.jbei.org/

**Project Goals: The Joint BioEnergy Institute (JBEI) aims to produce a chemically diverse suite of biofuels from lignocellulosic biomass. Many of the target fuels at JBEI rely on well-characterized metabolic pathways (such as the straight-chain fatty acid biosynthetic pathway) to provide precursors for synthesis of biofuel molecules. To date, modified fatty acid biosynthetic pathways have been used at JBEI to synthesize alkenes and fatty acid ethyl esters (FAEE) in engineered *E. coli* strains. Aliphatic hydrocarbons (alkanes and alkenes) are appealing targets for advanced biofuels, as they are predominant components of petroleum-based gasoline and diesel fuels; long-chain alkenes are useful as feedstocks that can be refined (cracked) into diesel- or gasoline-range hydrocarbons. FAEE, which are chemically similar to the fatty acid methyl esters that compose biodiesel, are favorable as petroleum diesel substitutes. A goal of the work described here is to improve the production of these biofuels in *E. coli*.**

We have discovered a gene cluster (oleABCD) in the actinobacterium Micrococcus luteus that, when heterologously expressed in a fatty acid-overproducing *E. coli* strain, produces long-chain (predominantly C27 and C29) alkenes, as well as unsaturated, aliphatic ketones of the same carbon number as the alkenes (Beller et al. 2010). We have proposed that the key enzyme in this pathway, OleA, catalyzes a head-to-head decarboxylative Claisen condensation of fatty acid derivatives (b-ketoacyl-CoAs) and have carried out in vivo and in vitro studies to determine if this is the case. One of our findings supporting this hypothesis is that titer of the alkenes can be improved if the host strain overproduces b-ketoacyl-CoAs rather than fatty acids. We have also shown in vitro that OleA can convert b-ketoacyl-CoAs into the same unsaturated, aliphatic ketones that were observed during in vivo studies with oleA. Furthermore, we have mutated three conserved residues in OleA (Cys132, His272, and Asn302) that also occur in the homologous, well-characterized protein FabH and are known to be essential to decarboxylative Claisen condensation catalyzed by that enzyme. Mutations to these three catalytic residues completely inactivated OleA, suggesting that these residues are essential in OleA as they are in FabH and may serve a similar function. We are continuing to use information about the OleA reaction to enhance alkene production.

Steen et al. (2010) engineered an *E. coli* strain to produce FAEE from glucose by introducing multiple modifications: overproduction of fatty acyl-CoAs (by overexpression of the thioesterase 'tesA and the acyl-CoA synthetase fadD, and deletion of the b-oxidation gene fadE), introduction of a pathway for producing ethanol from pyruvate (pdc, adhB), and addition of an acyltransferase that synthesizes FAEE from acyl-CoAs and ethanol. These genes were carried on three plasmids, which made the host susceptible to instability due to plasmid loss. Recent efforts at enhancing the FAEE-producing strain have included the successful incorporation of multiple copies of the five FAEE genes into the chromosome of the host strain; this was accomplished by Chemically Inducible Chromosomal Evolution (CIChE; Tyo et al. 2009).

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.

# 53

## Microbial Production of a Terpene Based Advanced Biodiesel

Pamela Peralta-Yahya,[1,2] Mario Ouellet,[2,3] Rossana Chan,[2,3] Helcio Burd,[2,3] Aindrila Mukhopadhyay,[2,3] Jay D. Keasling,[1,2,3] and **Taek Soon Lee**[2,3*] (tslee@lbl.gov)

[1]Chemical Engineering Department, University of California, Berkeley; [2]Joint BioEnergy Institute, Fuels Synthesis Division, Lawrence Berkeley National Laboratory, Emeryville, Calif.; and [3]Physical Bioscience Division, Lawrence Berkeley National Laboratory, Emeryville, Calif.

http://www.jbei.org/

**Project Goals: Sesquiterpenes (C15) are potential Diesel or Jet fuel alternatives. They have branched and cyclic structure, which improves their cold weather performance and adds more advantages to use sesquiterpenes as Diesel or Jet fuel alternatives. In nature, sesquiterpenes are mostly produced from plants, and the engineering of microorganisms that can produce sesquiterpenes has been an attractive topic to more and more researchers in the field of medicine, perfumery, and recently in biofuel.**

In this study, we designed and identified sesquiterpene compounds that have positive properties to be used as a fuel. We have tested the appropriate fuel properties of the target compounds, and engineered the heterologous biosynthetic pathway into two model hosts, *E. coli* and *S. cerevisiae*, to produce this sesquiterpene compound with a relatively high yield.

To achieve higher production titer of this potential sesquiterpene fuel molecule, we have optimized the pathway to accumulate the precursors, and also engineered the host strain to make fermentation process more efficient.

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.

# 54

## Re-Engineering Secondary Cell Wall Deposition in *Arabidopsis*

L. Zhang,[1] F. Yang,[1] P. Pradhan-Mitra,[1] L. Prak,[2] JS. Kim,[1]A. Eudes,[1]M. Auer,[2] and **D. Loqué**[1]*

[1]Joint BioEnergy Institute, Feedstocks Division, Lawrence Berkeley National Laboratory, Emeryville, Calif.; and [2]Joint BioEnergy Institute, Technology Division, Lawrence Berkeley National Laboratory, Emeryville, Calif.

http://www.jbei.org

**Project Goals: The plant cell wall represents a large source of polysaccharides that could be used to substitute for sugar derived from starchy grains currently used to feed and to produce biofuels. This lignocellulosic biomass, largely under-utilized, is mainly composed of sugar polymers (cellulose and hemicellulose) embedded in a strong aromatic polymer called lignin. Recalcitrant to degradation, lignin inhibits efficient extraction and hydrolysis of the cell wall polysaccharide and prevents cost-effective lignocellulosic-biofuel production. Unfortunately, lignin cannot simply be genetically removed without incurring deleterious consequences on plant productivity. The cost effectiveness of the conversion of the lignocellulosic biomass into sugars is still one of the major components to produce cheap biofuels. Therefore, strategies that can be used to reduce the lignin recalcitrance and that can increase polysaccharide deposition into the cell wall without altering plant growth should be developed.**

We used synthetic biology to re-engineer cell wall biosynthesis and deposition without affecting plant growth. We generated strategies to manipulate the spacio-temporal deposition of lignin as well as to modify its composition to tackle lignin recalcitrance. We also developed tools to manipulate the control of cell wall biosynthesis and used them to enhance cell wall polysaccharides deposition. All these modifications could be translated into an improvement in saccharification efficiency. Developed approaches and preliminary data of these approaches will be presented.

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.

# 55

## $^{13}$C Metabolic Flux Analysis-Aided Exploration of the High-Glucose Role of the Sip1 b-Subunit of the Snf1 Kinase Complex in *Saccharomyces cerevisiae*

Christopher Shymansky* (CMShymansky@lbl.gov), Dan Weaver, **Hector Garcia Martin,** and **Jay Keasling**

Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.

http://www.jbei.org

**Project Goals: Prototrophic base and sip1Δ mutant strains of *S. cerevisiae* capable of galactose-tunable bisabolene production were engineered in order to characterize the effect of SIP1 knockout on sesquiterpene production and flux profiles. The GAL1 promoter system was complemented by an already existing knockout of GAL1 in the original parent strain, so as to avoid direct contribution of carbon flux by the inducer. Recombinant strains were characterized by growth, six-day bisabolene production, and $^{13}$C metabolic flux analysis ($^{13}$C MFA) experiments. A global transcriptional analysis of a sip1Δ mutant was reported in 2010[1]. However, this is the first study to examine the high-glucose role of Sip1 from a fluxomic perspective and the first to examine production of a biofuel candidate upon knockout of SIP1.**

An organism's metabolic fluxes or fluxome are the final output of reaction thermo-dynamics and cellular regulation[2,3]. Intracellular networks are so interconnected that genetic changes often produce unexpected results. $^{13}$C metabolic flux analysis is a method of obtaining intracellular flux profiles, which can then be used to troubleshoot, make further strain engineering decisions, and/or improve predictive models. The Snf1 kinase complex plays a central role in glucose repression. Snf1 associates with Snf4 and then with one of three b-subunits upon depletion of glucose. That associated with Sip1 is sequestered in the vacuole, that with Sip2 remains in the cytosol, and that containing Gal83 activates genes involved in catabolism of other fermentable (sucrose, etc.) and non-fermentable (ethanol, acetate, etc) carbon sources[1,4]. This heterotrimer is completely dissociated and all of its components remain in the cytosol in the presence of glucose. Little is known about the role of Sip1 under these conditions due to a lack of phenotypic difference between wildtype and sip1Δ mutants[1,4,5]. Sip1 has been shown to be a negative regulator of genes involved in galactose catabolism[6] and has been found to play a role in adherent growth[7].

Knockout of Sip1 had no effect on bisabolene production and neither mevalonate nor biasabolene were excreted during exponential phase during glucose repressing conditions. Deletion of SIP1 was found to increase specific growth rate by 15%. Comparison of base and sip1Δ mutant flux profiles indicated that knockout of SIP1 corresponded to a decrease in overall TCA cycle flux by 58.5% and an increase in ethanol excretion by about 25%, overall flux to biomass flux by

35% and an activation of the relatively inactive section of the metabolic network responsible for serine, glycine, threonine, tryptophan, methionine, and aspartate production. Also, an inverse relationship between TCA cycle flux and specific growth rate was confirmed. To what extent these trends are due to derepression of GAL genes upon SIP1 knockout or the lapse of some other cytosolic role(s) of Sip1 is unclear. Regardless, these results are consistent with previous growth, global transcriptional, and $^{13}C$ MFA studies.

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.

**References**

1. Zhang J., Olsson L., Nielsen J. The b-subunits of the Snf1 kinase in *Saccharomyces cerevisiae*, Gal83 and Sip2, but not Sip1, are redundant in glucose derepression and regulation of sterol biosynthesis. *Molecular Biology*. 77[2]: 371-383. 2010.
2. Nielsen J., It Is All About Metabolic Fluxes. *J. Bacteriol.* 185: 7031-7035. 2003.
3. Kelleher J., Stephanopoulos G. How to make a superior cell. *Science.* 292: 2024-2025. 2001.
4. Zaman S., Lippman S.I., Zhao X., and Broach J.R. How Saccharomyces Responds to Nutrients. *Anu. Rev. Genet*. 42: 27-81. 2008.
5. Breslow DK, et al. A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat Methods* 5[8]: 711-8. 2008.
6. Mylin L.M, Bushman V.L., Long R.M., Yu X., Lebo C.M., Blankt T.E., Hopper J.E. SZPl Is a Catabolite Repression-Specific Negative Regulator of GAL Gene Expression. *Genetics.* 137: 689-700. 1994.
7. Vyas V.K., Kuchin S., Berkey C.D., Carlson M., Snf1 Kinases with Different b-Subunit Isoforms Play Distinct Roles in Regulatin Haploid Invasive Growth. *Molecular and Cellular Biology*. 23[4]; 1341-1347. 2003.

# 56

## Thermophilic Cellulase Cocktail for the Saccharification of Ionic Liquid Pretreated Biomass

Joshua I. Park* (jipark@lbl.gov), John M. Gladden, Eric J. Steen, Blake A. Simmons, and **Kenneth L. Sale**

Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.

http://www.jbei.org

To determine the feasibility of developing a thermophilic bacterial cellulase cocktail to hydrolyze biomass, we supplemented the secretome of a thermophilic bacterial community cultivated on microcrystalline cellulose with two recombinant enzymes. An enzyme activity screen of the secretome detected high endoglucanase and xylanases activities, but relatively low cellobiohydrolase and β-glucosidase activities, which limited release of glucose from cellulose and ionic liquid(IL)-pretreated switchgrass. We show that supplementing the secretome with a recombinant CBM3-GH5 from *Caldicellulosiruptor saccharolyticus* (cellobiohydrolase) and a GH3 from *Thermotoga maritima* (β-glucosidase) facilitated the release of glucose from IL-pretreated switchgrass at 70˚C and 80˚C. Furthermore, the hydrolysate from the enzymatic hydrolysis was converted to biofuel by *E. coli* engineered to produce fatty acid ethyl esters (FAEE). Analysis of the metagenome derived from a thermophilic community closely related to the one described above identified 37 full-length genes annotated as cellulases. By employing rapid cell-free protein expression and activity screening of these 37 genes, we identified 18 to be active endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91), or β-glucosidases (EC 3.2.1.21). The fact that the aforementioned hybrid secretome/recombinant cocktail efficiently saccharified IL-pretreated switchgrass suggests that these newly discovered cellulase enzymes are good candidates from which to assemble a purely recombinant thermophilic cellulase cocktail that can efficiently saccharify IL-pretreated biomass at temperatures greater than those possible with current commercial cocktails.

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.

# 57

## Discovering and Engineering Ionic Liquid Tolerant Cellulases

Supratim Datta,[1] Hanbin Liu,[1] Ling Qin,[2] Blake A. Simmons,[1] and **Kenneth L. Sale**[1]* (KLSale@lbl.gov)

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.; and [2]Sandia National Laboratory, Livermore, Calif.

http://www.jbei.org

One of the main barriers to the enzymatic hydrolysis of cellulose results from its highly crystalline structure. Pretreating biomass with ionic liquids (IL) increases enzyme accessibility, and the cellulose can be recovered through precipitation with an anti-solvent. For an industrially feasible pretreatment and saccharification process, it is necessary to develop cellulases that are stable and active in the presence of ILs that are either coprecipitated with recovered cellulose or that are present in a simultaneous pretreatment and saccharification process. However, a significant decrease in cellulase activity in the presence of trace amounts of ILs has been reported in the literature, necessitating extensive processing to remove residual ILs from the regenerated cellulose. Towards that end, we have investigated the stability of extremophilic enzymes in the presence of the IL, 1-ethyl-3-methylmidazolium acetate [C2mim][OAc], and compared it to the industrial benchmark *Trichoderma viride* (*T. viride*) cellulase. Under their optimum conditions, the thermophilic enzymes showed significantly higher

[C2mim][OAc] tolerance than *T. viride* cellulase and were active on IL pretreated substrates with little loss in activity after exposure to 15% [C2mim][OAc] for 15 hours. Since these results demonstrate the potential of using IL-tolerant extremophilic cellulases for hydrolysis of IL-pretreated lignocellulosic biomass for biofuel production, our goals are to engineer enhanced IL tolerance and IL tolerant cellulose cocktails at different pH regimes. Towards that end, we have chosen a thermophilic cellobiohydrolase, Cel9A from *Alicyclobacillus acidocaldarius* and rationally designed mutations to enhance tolerance towards [C2mim][OAc]. The effects of these mutations on enzyme efficiency will be discussed. We will also present our initial results towards a cellulose cocktail functional under acidic and alkaline conditions.

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.

# 58

## Lignocellulosic Biomass Degradation via Targeted Glycoside Hydrolases

Richard Heins[1,2,3]* (raheins@lbl.gov), Vimalier Reyes-Ortiz,[1,5] Edward Kim,[4] Rajat Sapra,[1,3] Kenneth Sale,[1,3] Blake Simmons,[1,3] **Danielle Tullman-Ercek**,[1,3,4,5] and **Masood Hadi**[2,3]

[1]Joint BioEnergy Institute, Deconstruction Division, Lawrence Berkeley National Laboratory, Emeryville, Calif., www.jbei.org; [2]Joint BioEnergy Institute, Technology Division, Lawrence Berkeley National Laboratory, Emeryville, Calif.; [3]Sandia National Laboratory; [4]Department of Chemical Engineering University of California, Berkeley; and [5]Joint Department of Bioengineering University of California, Berkeley and San Francisco

http://www.jbei.org/

Second-generation biofuels produced from renewable lignocellulosic feedstocks present an attractive alternative to traditional fossil fuels. Developing an energy-efficient and cost effective process to deconstruct and convert plant derived cellulose and hemi-cellulose into glucose presents a significant challenge due to the recalcitrant nature of the biomass.

Following pretreatment, cellulases are used to break down cellulosic fibers but their efficiency is low and commercial cocktails are prohibitively expensive. Furthermore, these cocktails lose their activity at high temperatures and salt concentrations typically employed in next generation pretreatment conditions and carried over into the enzyme conversion process. We have employed a modular approach to rapidly prototype and engineer highly robust cellulases by fusing thermophillic carbohydrate-binding modules (CBMs) to robust catalytic domains.

Here we demonstrate that the addition of CBMs enhances enzymatic activity compared to the catalytic domain alone at high temperatures when assayed on insoluble crystalline substrate Avicel and on a potential energy crop, switchgrass. We discuss some of our initial results and outline a strategy directed towards improving the activity of these chimeric cellulases under high salt conditions.

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.

# 59

## Effect of Cellulose Binding Module on the Activity of Thermophilic Endo/Exocellulase Cel9A from *Alicyclobacillus acidocaldarius*

Rajiv Bharadwaj[1,2]* (rbharadwaj@lbl.gov), Vimalier Reyes-Ortiz,[1,4] Supratim Datta,[1,3] April Wong,[1,2] Kenneth L. Sale,[1,3] Blake A. Simmons,[1,3] Danielle Ercek,[1,4] Paul D. Adams,[1,5] and **Anup K. Singh** [1,2]

[1]Joint BioEnergy Institute, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Emeryville, Calif.: [2]Biotechnology and Bioengineering Department, Sandia National Laboratories, Livermore, Calif.; [3]Biomass Science and Conversion Technology Department, Sandia National Laboratories, Livermore, Calif.; [4]Department of Chemical Engineering, University of California, Berkeley; [5]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.

http://www.jbei.org/

The hydrolysis of biomass to fermentable sugars using glycosyl hydrolases such as cellulases is a limiting and costly step in the conversion of biomass to biofuels. Identification and characterization of novel enzymes with high specific activities under industrially relevant conditions (T > 65˚C, pH ~5) is necessary. Cel9A, a thermophilic and ionic liquid (IL)-tolerant endoglucanase from *Alicyclobacillus acidocaldarius*, is an attractive candidate for the hydrolysis of IL pretreated biomass. We are performing microfluidic kinetic characterization of the enzymatic activity of Cel9A with and without thermophilic cellulose binding modules (CBM). The kinetic analyses suggest that the presence of CBMs significantly enhances the activity of the enzyme on IL pretreated substrates. In addition, the analysis shows that the primary hydrolysis products are cellobiose and glucose. This suggests that Cel9A has dual endo- and exo-activity. We have leveraged the dual activity of Cel9A to develop a minimal two-component thermophilic cellulase "cocktail" by including b-glucosidase in the saccharification reaction. Currently, we are characterizing the performance of the two-component cellulase cocktail for the hydrolysis of IL and AFEX pretreated biomass substrates.

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.

# 60

## Neutron Reflectometry and QCM-D Study of the Interaction of Endoglucanase Enzymes with Films of Amorphous Cellulose

Gang Cheng[1,2]* (gcheng@lbl.gov), Zelin Liu,[3] Supratim Datta,[1] Jaclyn K. Murton,[2] Michael Jablin,[4] Jaroslaw Majewski,[4] Candice Halbert,[5] James Browning,[5] John Ankner,[5] Bulent Akgun,[6] Alan Esker,[3] Blake A. Simmons,[1,2] and **Michael S. Kent**[1,2]

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.; [2]Sandia National Laboratories, Livermore, Calif. and Albuquerque, N.M.; [3]Department of Chemistry, Virginia Polytechnic Institute, Blacksburg; [4]Lujan Neutron Science Center, Los Alamos National Laboratory, Los Alamos, N.M.; [5]Spallation Neutron Source, Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [6]National Institute of Standards and Technology, Gaithersburg, Md.

http://www.jbei.org

Improving the efficiency of enzymatic hydrolysis of cellulose is one of the key technological hurdles to reduce the cost of producing ethanol and other transportation fuels from lignocellulosic material. A better understanding of how soluble enzymes interact with insoluble cellulose will aid in the design of more efficient enzyme systems. We report a study involving neutron reflectometry (NR) and quartz crystal microbalance with dissipation (QCM-D) of the interaction of a commercial fungal enzyme extract (*T. viride*) and a series of endoglucanases with amorphous cellulose films. The endoglucanases studied include two from thermophilic bacteria (Cel9A from *A. acidocaldarius* and Cel5A from *T. maritima*), a processive endoglucanase from a marine bacterium (Cel5H from *S. degradans*) and two mesophilic fungal endoglucanases (Cel45A from *H. insolens* and a GH12 endoglucanase from *A. niger*). The use of amorphous cellulose is motivated by the promise of ionic liquid pretreatment as a second generation technology that disrupts the native crystalline structure of cellulose. NR reveals the profile of water through the film at nm resolution, while QCM-D provides changes in mass and film stiffness. The measurements were made in the absence of flow or agitation. At 20°C and 0.3 mg/ml, the *T. viride* cocktail rapidly digested the entire film, beginning from the surface followed by activity throughout the bulk of the film. The endoglucanases, while all showing less activity than the commercial *T. viride* cocktail, displayed a wide range of behavior. A consistent finding for all the data is that the endoglucanases lacking a cellulose binding domain digested to a very limited extent at the surface of the film, whereas the endoglucases possessing CBMs penetrated and digested within the bulk of the films.

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.

# 61

## Targeted Proteomics for Metabolic Pathway Optimization: Application to Terpene Production

**Christopher J. Petzold*** (CJPetzold@lbl.gov), Alyssa M. Redding-Johanson, Tanveer S. Batth, Rossana Chan, Rachel Krupa, Heather L. Szmidt, Paul D. Adams, Jay D. Keasling, Taek Soon Lee, and Aindrila Mukhopadhyay

Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.

http://www.jbei.org

**Project Goals: Successful metabolic engineering relies on methodologies that aid assembly and optimization of novel pathways in microbes. Many different factors may contribute to pathway performance, and problems due to mRNA abundance, protein abundance, or enzymatic activity may not be evident by monitoring product titers. To this end, synthetic biologists and metabolic engineers utilize a variety of analytical methods to identify the parts of the pathway that limit production. In this study, targeted proteomics, via selected-reaction monitoring (SRM) mass spectrometry, was used to measure protein levels in *E. coli* strains engineered to produce isoprenoids from the mevalonate pathway. From this analysis, two mevalonate pathway proteins, mevalonate kinase (MK) and phosphomevalonate kinase (PMK) from S. cerevisiae, were identified as potential bottlenecks. Codon-optimization of the genes encoding MK and PMK and expression from a stronger promoter led to significantly improved MK and PMK protein levels and over three-fold improved final amorpha-4,11-diene titer (> 500 mg/L).**

Many metabolically engineering efforts are aimed at reducing the amount of time required to assemble and optimize novel heterologous pathways in host organisms. Generally, methods to confirm successful cloning of the biosynthetic pathway and subsequent metabolite production are well established and routinely used. However, methods that reveal potential bottlenecks arising from problems involving protein production are still limited. Historically, monitoring protein levels has been accomplished by using Western blot analysis, which can be difficult, time consuming, and cost prohibitive especially during exploratory stages of pathway design when many different proteins must be monitored simultaneously. Alternatively, a targeted proteomics approach, via selected-reaction monitoring (SRM) mass spectrometry, can be used to quantify many specific proteins in a sample. SRM methods provide high selectivity, sensitivity and low cost to enable rapid quantification of multiple proteins from a sample. We validated this method for pathway engineering projects by using standardized plas-

mids expressing RFP. These systems were used to quantify relative changes in protein production by changing plasmid copy number using different origins of replication or by expression from multiple promoters. Secondly, SRM mass spectrometry was applied to various constructs of *E. coli* engineered with the mevalonate pathway from *S. cerevisiae*. *E. coli* containing a high-flux mevalonate pathway has the potential to provide a vast range of isoprenoid-based bulk and high value compounds that are typically obtained from petrochemical or plant sources. In these experiments, monoterpene and sesquiterpene end products were measured to study the impact of vector modification on pathway bottlenecks and to refine hypotheses going forward. In particular, these studies revealed that levels of two pathway proteins, mevalonate kinase (MK) and phosphomevalonate kinase (PMK), were particularly low in several constructs. To overcome these bottlenecks, several different strategies (i.e., codon optimization, expression via stronger promoter) were employed to increase MK and PMK and balance overall protein levels. These results demonstrate that measuring protein levels constitutes an important metric to facilitate pathway optimization of metabolically engineered organisms and enable characterization of parts for use in synthetic biology.

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.

# 62

## A Systematic Pipeline for Biomass Characterization Using Aligned Mechanical Stress Analysis, Polarized Raman Microspectroscopy and Scanning Electron Microscopy

Patanjali Varanasi,[1,2‡*] (pvaranasi@lbl.gov), Lan Sun,[1,2‡] Jacob Katsnelson,[3] Rita Sharma,[1,4] Manoj Sharma,[1,4] Miguel Vega-Sanchez,[1,4] Marcin Zemla,[5] David Larson,[5] Pamela Ronald,[1,4] Blake Simmons,[1,2] Paul Adams,[1] Seema Singh,[1,2] and **Manfred Auer**[1,5] (mauer@lbl.gov)

[1]Joint BioEnergy Institute, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Emeryville, Calif.; [2]Biomass Science and Conversion Technology Department, Sandia National Laboratories, Livermore, Calif.; [3]Department of Chemistry, University of California, Berkeley; [4]Department of Plant Pathology, University of California, Davis; [5]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.
‡Equal authorship

http://www.jbei.org/

Cellulose and hemi-cellulose present in lignocellulosic biomass can be converted to simple sugars through enzymatic hydrolysis and hence to advanced biofuels. However, lignocellulosic biomass is not yet economically viable due to the saccharification barrier. Genetic modification of lignocellulosic biomass may enhance saccharification yields, but may weaken the plant's strength and recalcitrance to biochemical attack in nature. Any successful rational engineering approach requires an in-depth structural and chemical understanding of the consequences of biomass genetic engineering.

Employing a suite of biophysical tools, we build a pipeline for plant mutant characterization, including mechanical strength measurement via tensile stress testing, determination of fracture patterns by scanning electron microscopy (SEM) and evaluation of chemical compositions and fiber orientation using polarized Raman microspectroscopy.

To demonstrate the concept of this screening pipeline, we have compared a known rice mutant, Brittle Culm, with the wild type plant. Lower mechanical strength with a brittle fracture of the leaves was observed for the mutant by the tensile test and SEM. We further found lower cellulose content and increased disorder of cellulose microfiber orientation by polarized Raman microspectroscopy, demonstrating proof-of-concept for this integrated biophysical approach.

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.

# 63

## A Rapid, Non-Destructive Method to Screen for Cell Wall Mutants using Fourier Transform-Near Infrared Spectroscopy

Andreia Michelle Smith[1*] (asmith@lbl.gov), Jeemeng Lao,[1] Mawshen Chern,[2] Joshua Heazlewood,[1] Pamela Ronald,[1,2] and **Miguel Vega-Sanchez**[1,2]

[1]Joint BioEnergy Institute, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Emeryville, Calif.; and [2]Department of Plant Pathology, University of California, Davis

http://www.jbei.org/

Selection of mutant plants with altered cell wall composition or structure can prove useful in the discovery of novel genes involved in the biosynthesis pathway of plant cell walls. We are using a forward genetics approach to identify genes affecting cell wall composition. In this manner, a well-constructed mutant screen that has a balance between high-throughput and robustness is essential. Sugar composition determination, done by High-Performance Anion-Exchange Chromatography (HPAEC), can be a conclusive identifier of altered plant cell walls however it is extremely labor intensive and is not feasible on a large scale. Recently, Near Infrared Spectroscopy has shown promise in characterizing plant material. The non-destructive, rapid and possible quantitative nature of this technique makes it very attractive to use for mutant screen. However, it is limited

without a calibration set that includes biological variability associated with batches of plants grown at separate times. Here, we introduce a method that can be implemented in a blind cell wall by using fast scanning of intact plant leaves by NIR -NIR spectroscopy without the need for sample pre-processing. The feasibility of the approach was first validated using known cell wall mutants in *Arabidopsis* and then applied to a rice mutant collection consisting of thousands of unknown samples. By using monosaccharide composition analysis of selected NIR outliers after multivariate analysis, a calibration set was used to develop a model that allows prediction of cell wall sugar composition from the NIR spectra alone.

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.

# 64

## Encoding Substrates with Mass Tags to Resolve Stereospecific Reactions

Kai Deng[1,2*] (kdeng@lbl.gov) Wolfgang Reindl,[1,3] Paul D. Adams,[1] **Anup K. Singh**,[1,2] and **Trent N. Northen**[1,3]

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.; [2]Biotechnology and Bioengineering Department, Sandia National Laboratories, Livermore, Calif.; and [3]Department of Bioenergy/GTL and Structural Biology, Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.

http://www.jbei.org/

Resolving multiple pathways to identical products is an important but often difficult process typically requiring isotopically labeled reactants. Glycans in particular have tremendous structural diversity making it difficult to resolve reaction pathways. Here fluorous phase synthetic methods are combined with a "bar-coding" strategy to encode reactants based on the mass of the fluorous tag. This allows differentiation of reaction pathway based on nanostructure-initiator mass spectrometry (NIMS) based mass readout. We demonstrate analysis of three stereoisomers (maltose, lactose and cellobiose) and show that this approach resolves stereospecific reaction pathways resulting in the same product (glucose). The method is generally applicable but particularly compatible with existing fluorous phase synthetic strategies.

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.

# 65

## Microfluidic Platform for Synthetic Biology Applications

Chieh Chang[1,2,3*] (cchang@lbl.gov), Rajiv Bharadwaj[1,2] Aarthi Chandrasekaran,[1,2] Nathan J. Hillson,[3] Paul Adams,[1] and **Anup K. Singh**[1,2]

[1]Joint BioEnergy Institute, Technology Division Lawrence Berkeley National Laboratory, Emeryville, Calif.; [2]Bioengineering and Biotechnology Department, Sandia National Laboratories; and [3]Joint BioEnergy Institute, Fuels Synthesis Division Lawrence Berkeley National Laboratory, Emeryville, Calif

http://www.jbei.org

**Project Goals: Synthetic biology applications require assembly of several biological parts (e.g., genes) in a plasmid. This leads to a large combinatorial problem where several thousands of combinations of parts have to be assembled. The scale of the problem is a significant challenge in terms of cost and time required to generate all the possible combinations. Currently, a combination of 96/384 well plates and manual pipetting or very expensive robotics instrumentation is used to perform parts assembly and screening.**

In this project, our droplet-based microfluidic platform enables a more efficient way to perform combinatorial parts assembly and screening. Specific advantages of our method include:

- Cost of reagents reduced by a factor of 100-fold
- 2 to 10-fold reduction in time taken for reaction
- Significant reduction in the equipment cost
- Significant reduction in device footprint compared to that of liquid handling robotic systems

Combination of the above advantages will allow practitioners to undertake a large number of experiments; a number that will not be feasible with conventional approaches.

The microfluidic chip permits generation and manipulation of droplets to carry out molecular biology steps in the following sequence (Figure 1):

- Encapsulation of different genes in individual droplets and programmable merging (Figure 2).
- Droplets can carry reagents necessary to enable various cloning, expression, and transformation steps.
- Droplets maintain their discrete nature even when flow is stopped, allowing one to carry out timed incubations and other steps (Figure 2).
- Plasmids are mixed with cells for transformation. Alternately, the plasmids can be mixed with a cell-free expression reagent to express protein in-vitro.
- Droplets can carry cell culture media permitting growth and division of cells.
- Microfluidic chip can be interfaced with an optical detection or imaging instrument (such as an optical

microscope) to image cells, count them, or monitor expression of a fluorescent protein.

Acknowledgement: 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.

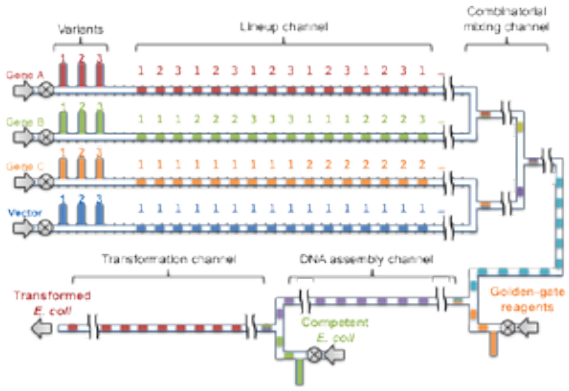



Figure 1. Schematic of microfluidic platform for synthetic biology applications.

Figure 2. Photos showing droplet generation, mixing, and queuing.

# 66

## The Joint BioEnergy Institute (JBEI) Computational Biology Core

Dylan Chivian[1,2]* (DCChivian@lbl.gov), John T. Bates,[1,2] Greg D. Friedland,[1,3] Vinay Satish Kumar,[1,2] and **Adam P. Arkin**[1,2,4] (APArkin@lbl.gov)

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.; [2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; [3]Biomass Science and Conversion Technology Department, Sandia National Laboratories, Livermore, Calif.; [4]Department of Bioengineering, University of California, Berkeley

http://www.jbei.org/

The Computational Biology Core Group in the Technology Division of the Joint BioEnergy Institute (JBEI) is responsible for data integration and comparative, evolutionary, and functional genomic analysis for the purpose of engineering microbes for biofuel production. Leveraging the MicrobesOnline web resource (www.microbesonline.org) for comparative and functional genomics, we are extending the capabilities to permit analysis of microbial systems at various scales. Biological engineering requires an understanding of systems from atoms to communities. The efforts of the Computational Biology Core are therefore geared towards creating tools to facilitate our research at the level of components, systems, cells, and communities. For example, biological degradation of plant cells walls is accomplished by enzymes containing multiple domains that in combination confer specificity and activity. We are studying the combinations nature has employed to allow for making our own combinations. At the atomic scale, we are computationally designing the structures of proteins for increased stability under industrial conditions. At the system level, we are working to discover metabolic pathways for biofuel production as well as the genetic factors involved in tolerance toxic molecules found in pretreatment conditions and resistance to biofuel toxicity. To aid efforts in synthetic pathway engineering, we are building a framework for analyzing functional data in a metabolic network context and tools for discovery of genes for retrosynthetic pathways. At the cellular level, we are using evolutionary studies with phenotype data and genetic analysis to engineer cell lines that are better suited to industrial conditions or have superior yields of the desired biofuel. Finally, at the community level, discovery of genes from environmental samples will expand the repertoire of enzymes we can engineer for biomass degradation under varying conditions. Our approach is to analyze the environmental genomic data in a phylogenetic context to allow for a higher-resolution annotation of the role of each enzyme, as well as to obtain an evolutionary picture of the key functional genes and organisms in each ecosystem.

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.

# 67

## Electronic Laboratory Notebook Usage at JBEI

David Pletcher[1]* (dcpletcher@lbl.gov), John Gladden,[2] Brad Holmes,[2] Jay Keasling,[3] Steve Lane,[1] William Morrell,[1] Arthur Herrera Panganiban,[4,5] Julie Petersen,[5] Francis Wayne Townsend-Merino,[1] and **Paul Adams**[1]

[1]Technology Division, [2]Deconstruction Division, [3]Fuels Synthesis Division, [4]Technology Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.; and [5]Lawrence Berkeley National Laboratory, Information Technology, Berkeley, Calif.

http://www.jbei.org/

Electronic laboratory notebooks (ELNs) have great potential to improve laboratory organization and performance. A successful implementation will allow users to embed digital data and files with experimental narratives, foster collaborative work, improve communication, provide live search capabilities, assure durable and secure record storage,

and assure the traceability of intellectual property. JBEI has developed an integrated ELN system based on commercially successful, off-the-shelf products. The organization of this system will be presented. Our initial deployment to a pilot user community has yielded encouraging results, which will be discussed. We are continuing with improvements to the system and development of support for additional user interface formats.

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.

# 68

## DNA Assembly Design with j5 and DeviceEditor Biocad Tools

J. Chen, R.D. Rosengarten, D. Densmore, T.S. Ham, J.W. Thorne, J.D. Keasling, and **N.J. Hillson*** (njhillson@lbl.gov)

Joint BioEnergy Institute, Fuels Synthesis Division, Lawrence Berkeley National Laboratory, Emeryville, Calif.

http://www.jbei.org,
http://jbei-exwebapp.lbl.gov/j5

The production of clean renewable biofuels from cellulosic starting material requires concerted feedstock engineering, deconstruction of plant matter into simple sugars, and microbial fermentation of the sugars into biofuel. These three efforts share significant molecular biological challenges, including the construction of large enzymatic libraries (e.g. vast collections of glycosyl transferases, cellulases, and efflux pumps), the generation of combinatorial libraries (e.g. multi-functional enzyme domain fusions; variations in copy number, promoter and ribosomal binding site strength), and the concurrent assembly of multiple biological parts (e.g. the incorporation of an entire metabolic pathway into a single target vector). With these challenges in mind, we have developed two on-line software tools, j5 and DeviceEditor, that automate the design of sequence agnostic, scar-less, multi-part assembly methodologies and translates them to robotics-driven protocols. Given a target library to construct, the software provides automated oligo, direct synthesis, and cost-optimal assembly process design, and integrates with liquid-handling robotic platforms to set up the PCR and multi-part assembly reactions. This work reduces the time, effort and cost of large-scale cloning and assembly tasks, as well as enables research scales otherwise unfeasible without the assistance of computer-aided design tools and robotics.

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.

# Biofuels: Analytical and Imaging Technologies, Engineering, and Production

# 69

## Engineering Bacterial Surfaces to Display Cellulosomes for Biofuel Production

**Robert Clubb*** (rclubb@mbi.ucla.edu)

University of California, Los Angeles

**Project Goals: Engineer surface of *Bacillus subtilis* to display multi-enzyme cellulolytic protein complexes that can efficiently degrade biomass into fermentable sugars.**

To cost-efficiently produce biofuels, improved methods are needed to convert lignocellulosic biomass into fermentable sugars. One promising approach is to degrade biomass using cellulosomes, surface displayed multi-cellulase containing complexes present in cellulolytic *Clostridium* and *Ruminococcus* species. In this study we created genetically modified strains of *B. subtilis* that display on their surface heterologous proteins and protein-protein complexes. Proteins containing the appropriate cell wall sorting signal are covalently anchored to the peptidylglycan by co-expressing them with the *B. anthracis* sortase A (SrtA) transpeptidase (SrtA). Greater than 300,000 heterologous proteins per cell are displayed in strains in which the WprA cell wall protease has been deleted. A two-component minicellulosome was constructed that consists of a cell wall attached scaffoldin protein that non-covalently binds to the CelA endoglucanase from *C. thermocellum*. Unlike the wild-type organism, *B. subtilis* displaying the minicellulosome robustly grow on acid-treated cellulose by degrading it into its component sugars. Importantly, the cells exhibit greater cellulolytic activity than several previously reported in vitro and yeast displayed minicellulosomes. *B. subtilis* has a robust genetic system and is currently used in a wide range of industrial processes. Thus, grafting more elaborate multi-enzyme containing minicellulosomes onto the surface *B. subtilis* may yield more potent cellulolytic bacteria that can be used to degrade biomass.

# 70

## Employing Cyanobacteria for Biofuel Synthesis and Carbon Capture and Storage

**Christer Jansson**[1]* (PI) (CGJansson@lbl.gov), **Hoi-Ying Holman**[1,2] (Co-PI), **Ping Hu**[1] (Co-PI), **Nina Kamennaya**,[1] **Laura Mahoney**,[1] and **Elizabeth Holman**[1,3]

[1]Earth Sciences Division (ESD), Lawrence Berkeley National Laboratory, Berkeley, Calif., [2]Advanced Light Source (ALS), and [3]School of Chemsitry, University of California, Berkeley

**Project Goals: In our two inter-linked research projects, CyanoFuels and CyanoCarbon, we explore the potential for employing cyanobacteria in biofuel synthesis, and biological carbon capture and storage (CCS), respectively. We exploit the two major modes of $CO_2$ uptake exhibited by cyanobacteria, the photosynthetic conversion of $CO_2$ to biomass and metabolites, as well as the biomineralization of $CO_2$ and Ca2+ to calcium carbonate ($CaCO_3$) (Figure 1). We utilize synchrotron radiation Fourier transform infrared (SR-FTIR) spectromicroscopy as a non-invasive technique for metabolic fingerprinting of individual live cells in real time.**

**Background:** Cyanobacteria are photoautotrophic Gram-negative bacteria that carry out oxygenic photosynthesis. Many are diazotrophic and thus can assimilate not only $CO_2$ nut also $N_2$ from the atmosphere. Cyanobacteria occupy a wide array of terrestrial, marine, and freshwater habitats, which include extreme environments such as hot springs, deserts, bare rocks, and permafrost zones. In their natural environments, some cyanobacteria are often exposed to the highest rates of UV irradiance known on our globe. Cyanobacteria generally thrive in high $CO_2$ levels and are considered attractive systems for $CO_2$ capture from flue gas. Since many cyanobacteria are halophilic and, strains for biofuel production or carbon capture and storage (CCS) can be cultured in either marine waters, saline drainage water, brine from petroleum refining industry, or $CO_2$ injection sites, thereby sparing freshwater supplies. A large number of strains are thermophilic and thus tolerate high temperatures characteristic of flue gas. Also, being bacteria, cyanobacteria are amenable to homologous recombination, which allows rapid site-directed mutagenesis, gene insertions, replacements and deletions in a precise targeted and predictable manner.

In our two inter-linked research projects, CyanoFuels and CyanoCarbon, we explore the potential for employing cyanobacteria in biofuel synthesis, and biological carbon capture and storage (CCS), respectively. We exploit the two major modes of $CO_2$ uptake exhibited by cyanobacteria, the photosynthetic conversion of $CO_2$ to biomass and metabolites, as well as the biomineralization of $CO_2$ and $Ca^{2+}$ to calcium carbonate ($CaCO_3$) (Figure 1). We utilize synchrotron radiation Fourier transform infrared (SR-FTIR) spectromicroscopy as a non-invasive technique for metabolic fingerprinting of individual live cells in real time.

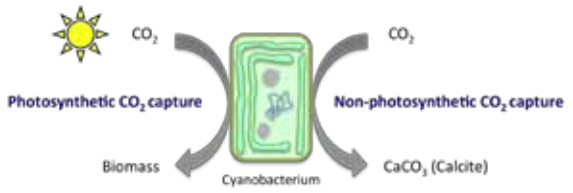



Figure 1. Two modes of $CO_2$ uptake in cyanobacteria.

**CyanoFuels:** We are using the cyanobacterium *Synechocystis* PCC6803 (*S.* 6803) as a model to study phototrophic biosynthesis of free fatty acids (FAs) and liquid biofuels such as FA alkylesters (biodiesel), alkanes, and isoprenoids. Cyanobacteria synthesize a wide array of hydrocarbons, including terpenes such as carotenoids and hopanoids, as well as linear and branched alkanes and alkenes, like branched methyl- and ethylalkanes which are unique to cyanobacteria. Biosynthesis of free FAs is of interest since they can be used for downstream chemical processing to biofuels and also because cyanobacterial biosynthesis of alkanes with specified chain lengths may require free FAs as an intermediate metabolite. Using SR-FTIR with multivariate analysis of *S.* 6803 strains engineered for free FA accumulation, we were able to demonstrate that the metabolic state of a strain could be fingerprinted at the single cell level. We conclude that SR-FTIR spectromicroscopy offers the opportunity to be employed as a high-throughput diagnostic tool for characterization and screening of genetically engineered cells.

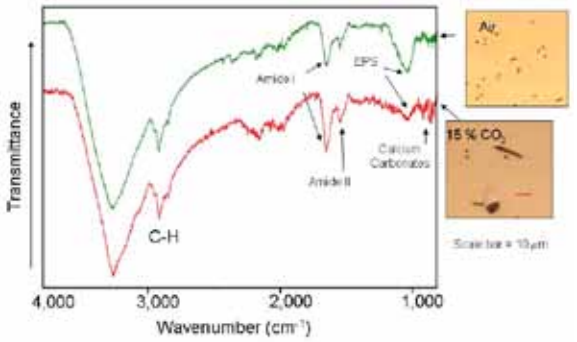



Figure 2. SR-FTIR spectra of calcification in the cyanobacterium *Synechococcus* 8802.

**CyanoCarbon:** Biomineralization of atmospheric or dissolved $CO_2$ by calcium carbonate ($CaCO_3$) precipitation (calcification) is a common phenomenon in marine, freshwater, and terrestrial ecosystems and is a fundamental process in the global carbon cycle. Formation and deposition of $CaCO_3$ occur at the outer cell surface, either at the proteineous surface layer (S layer) or in the exopolysaccharide substances (EPS). The cyanobacterial $CO_2$ concentrating mechanism (CCM) is often a critical part of the calcification process. Biological or biomimetic strategies for point-source CCS of flue gas based on cyanobacterial calcification offer potential strategies for reducing anthropogenic $CO_2$ emissions. However, much research is urgently needed to further our understanding of the biochemical and physical

processes in cyanobacteria that promote calcification. We are studying calcification in several lacustrine and marine cyanobacteria under different $CO_2$ levels, e.g 15% $CO_2$ typical of flue gas. We use SR-FTIR to monitor $CaCO_3$ crystallization and morphology, and to follow the metabolic shifts in cells grown under increasing $CO_2$ concentrations (Figure 2). We apply SIMS to study the cell surface properties and their influence in the calcification process.

# 71

## *Synechococcus* Sp. Pcc 7002: A Robust Cyanobacterial Platform For Systems Biology and Biofuels Engineering

D.A. Bryant[1]* (dab14@psu.edu), A.E. Konopka,[2] G.E. Pinchuk,[2] S.M. Stolyar,[2] M.C. Posewitz,[3] and **A.S. Beliaev**[2]

[1]Dept. of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park; [2]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash.; and [3]Colorado School of Mines, Dept. of Chemistry and Geochemistry, Golden

**Project Goals: The PNNL Biofuels Scientific Focus Area (BSFA) will conduct fundamental research on cyanobacteria with specific emphasis on pathways of carbon, nitrogen, and redox metabolism that consume reductant and conserve energy produced by photosynthetic light reactions. Consistent with DOE BER Genomic Sciences Program goals, the long-term objective of the BSFA is to develop a predictive understanding of metabolic subsystems and regulatory networks involved in solar energy conversion to biofuel precursors or products. Toward this goal, we will interrogate fluxes through metabolic subsystems that include (i) photosynthetic energy conservation and reductant generation, (ii) $CO_2$ accumulation, fixation, and reduction, (iii) biosynthesis of metabolic intermediates and monomers, and (iv) macromolecular synthesis. Our studies will exploit two complementary model cyanobacterial systems, *Cyanothece* sp. strain ATCC 51142 and *Synechococcus* sp. strain PCC 7002, and address discrete scientific hypotheses.**

The PNNL Biofuels Scientific Focus Area (BSFA) conducts fundamental research on cyanobacteria with specific emphasis on pathways of carbon and redox metabolism that consume reductant produced by photosynthetic light reactions. The long-term objective is to develop a predictive understanding of metabolic subsystems and regulatory networks involved in solar energy conversion to biofuel precursors or products. This requires a systems-level understanding of metabolic modules (see Figure 1) at levels beyond the biophysics of photosynthesis and the primary production of fixed $CO_2$. Optimizing biofuel production by cyanobacteria also requires an improved understanding of the physiological constraints and the regulatory controls affecting "downstream" carbon partitioning between biomass, metabolite storage pools, and desired product synthesis. Therefore, understanding the factors constraining maximum rates of carbon processing, including production of autocatalytic macromolecules (i.e., RNA and protein) is necessary to identify regulatory mechanisms and devise strategies to overcome and manipulate these constraints. A systems approach to these issues is being undertaken, exploiting the fast-growing cyanobacterium, *Synechococcus* sp. PCC 7002, which has the added benefit of a well-developed genetic system. Our experimental strategy dissects growth-limiting factors by introducing external (i.e., light and $CO_2$ availability) and internal perturbations (i.e., gene knock-outs and knock-ins) that affect the internal state of the cell and, then investigating system performance in response to variable growth rates imposed by chemostat culture. The effect of these perturbations is analyzed by assaying the state and activity of key markers within the metabolic subsystems and allows calculation of flux distributions through the system that, in combination with activity measurements are used for estimating the performance of metabolic modules.

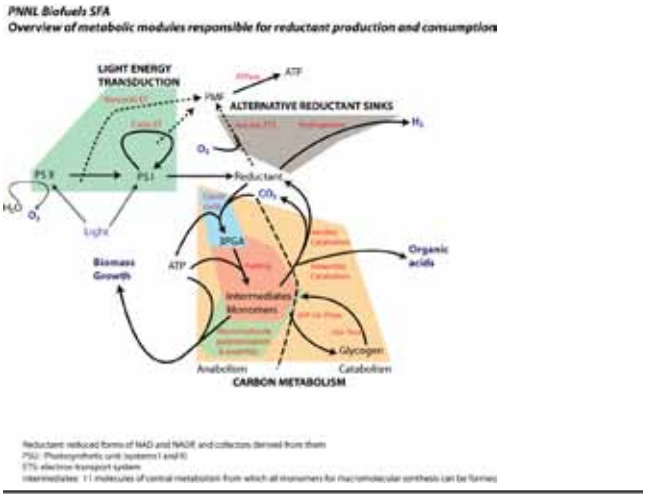



Figure 1. Metabolic modules of cyanobacteria. A central focus is the formation of reductant and conservation of energy from the coupled operation of Photosystems I and II (light energy transduction module), consumption of reductant via $CO_2$ reduction in the Calvin cycle or via other reductant sinks, and energy consumption (primarily the Calvin cycle or macromolecule polymerization) or storage (as glycogen). Anoxic or oxic catabolism of stored glycogen can generate ATP or ATP + reductant, respectively. "Carbon metabolism" is comprised of three principal submodules essential for biomass synthesis; $CO_2$ fixation; fueling reactions that produce the 11 precursor metabolites and ~75-100 building blocks from which all cellular molecules are formed; and macromolecule polymerization; in addition, the carbon metabolism entails glycogen synthesis and catabolism.

An experimentally tractable system is important for rapid progress and *Synechococcus* 7002 presents an attractive platform for functional genomics and biotechnological applications through metabolic engineering. This cyanobacterium has several useful and remarkable properties: it grows much faster than other well-characterized cyanobacteria; it tolerates extremely high light intensities; it is highly resistant to reactive oxygen and nitrogen species; it grows over a very wide range of salinity; it is easily manipulated genetically by natural transformation; and it can be grown photomixotrophically, photoheterotrophically, or heterotrophically on glycerol. The relatively small 3.4-Mb genome has been

completely sequenced and comprises a 3.0 Mb chromosome and 6 plasmids, which vary in size from 4.8 to 186 kb and in copy number (up to ~10 times the chromosome number for pAQ1). Using neutral-site integration platforms within these plasmids, a flexible and effective gene expression system has been developed. Various promoters, including $P_{cpcBA}$, $P_{psbA1}$, $P_{ntcABCD}$, and $P_{cpeC}$, allow gene expression at differing levels. Two of these systems allow regulatable expression using either N-source or light wavelength.

These advances have established *Synechococcus* 7002 as an exceptional platform for developing a more informed understanding of how metabolic networks are regulated and integrated in cyanobacteria. A thorough understanding of the fundamental mechanisms regulating metabolic pathway flux and photosynthetic activity will inform iterative metabolic engineering approaches designed to improve the synthesis/accumulation of products relevant to DOE bioenergy missions. We have over-expressed a variety of genes, including those encoding soluble and membrane proteins, complex multi-subunit proteins, and various tagged and fusion proteins. Based on transcript levels determined by cDNA sequencing, the *ldhA* gene, encoding D-lactate dehydrogenase, was overproduced by ~3000-fold relative to transcript levels in wild-type cells when expression was driven by the *cpcBA* promoter. The resulting strain excreted D-lactate and acetate continuously for up to 58 days; this led to the accumulation of up to 20 mM D-lactate and nearly 80 mM acetate in the external medium. This system has also been used in complementation analyses, in the production of novel carotenoids, in the expression of a heterologous, oxygen-tolerant [NiFe]-hydrogenase from *Ralstonia eutropha* (13 genes encoded by 14 kb) and in producing YFP-protein fusions for protein localization studies. Mutants lacking various fermentative enzymes can increase hydrogen production or redirect metabolites during dark anoxic conditions. Mutants lacking the ability to synthesize glycogen can be osmotically manipulated to store and then rapidly excrete up to 200 μg/ml glucosylglycerol and substantial amounts of sucrose.

We are also in the process of generating four distinct pAQ1-based constructs for expressing enzymes that will allow the synthesis of either fungible fuel feedstocks or biopolymer monomers in *Synechococcus* sp. PCC 7002. These include medium chain (C10-14) thioesterases, an alkane biosynthesis operon, a β-caryophyllene synthase (sesquiterpene), and an oleoyl-12-hydroxylase (biopolymer monomer). These enzymes have been shown to generate the desired product when transgenically expressed in a variety of other organisms, however, product yields are typically low. Regulatory and metabolic constraints are also likely to emerge in *Synechococcus* sp. PCC 7002 that limit the overall conversion efficiencies and produce unintended effects on photosynthetic efficiencies. Systematic efforts are therefore required to examine in detail the metabolic and regulatory consequences of introducing exogenous metabolic pathways into *Synechococcus* sp. PCC 7002. These efforts are anticipated to result in the production of the desired product, but are also likely to reveal limitations that must be addressed by subsequent research to improve yields.

# 72

## Constraint-Based Modeling of Cyanobacterial Metabolism

J.L. Reed[1]* (reed@engr.wisc.edu), T. Vu,[1] S.M. Stolyar,[2] G.E. Pinchuk,[2] O.V. Geydebrekht,[2] E.A. Hill,[2] L. Kucek,[2] J. De Ingeniis,[3] A. Osterman,[3] J. Hamilton,[1] A.E. Konopka,[2] D.A. Bryant,[4] and **A.S. Beliaev**[2]

[1]University of Wisconsin-Madison; [2]Pacific Northwest National Laboratory, Richland, Wash.; [3]Burnham Institute for Medical Research, La Jolla, Calif.; and [4]The Pennsylvania State University, University Park

**Project Goals: same as abstract above**

Oxygenic photosynthetic microorganisms possess a unique ability to convert light energy into chemical energy using water as an electron donor for the reduction of electron carriers (ferredoxin and NADP+). Despite the wealth of information on the mechanistic details of bacterial photosynthesis and supramolecular complexes involved in photosynthesis and $CO_2$ fixation, a system-level understanding of photosynthetic metabolism is yet to be achieved. Phototrophic microorganisms display varying levels of light conversion efficiencies, which ultimately translate into different rates of electron transfer, ATP/NAD(P)H production, and growth. Understanding the origin of these properties will provide new fundamental insights that can be widely applied to the development of photosynthetic systems for biofuel development. To increase biofuel production using photosynthetic strains we first need to improve our understanding of energy and reductant partitioning in photoautotrophs. This requires connecting energy-and reductant-generating reactions with reactions involved in the biosynthesis of biomass precursors and storage compounds. Within the scope of the PNNL Biofuels Scientific Focus Area (BSFA), we are exploring the mechanisms of energy conservation and carbon and reductant partitioning in cyanobacteria. An important outcome of the project will be the development of predictive tools, i.e., genome-scale metabolic and regulatory models, which will provide a platform for integrating all knowledge and experimental data generated within the project. The models will also have the ability to serve as an in silico tool for intelligently manipulating photosynthetic microorganisms to act as catalysts for solar energy conversion and will potentially allow development of a highly efficient biofuel production process.

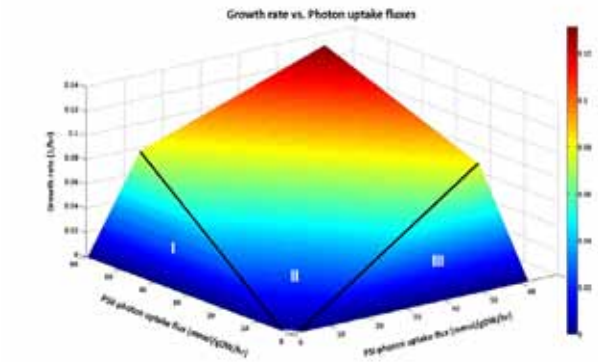



Figure. *In silico* prediction of the effect of photon uptake fluxes on growth rate. The 3D plot shows the solution space of growth rate vs. photon uptake fluxes. Region I and III are single-substrate limiting regions, while region II is dual-substrate limiting region.

We have built a genome-scale metabolic network for *Cyanothece* sp. strain ATCC 51142, a unicellular diazotrophic cyanobacterium that can temporally separate the process of light-dependent autotrophic growth and glycogen accumulation from $N_2$ fixation. The resulting model currently includes 806 genes, 585 metabolites, and 664 reactions accounting for common pathways such as central metabolism, respiration, nucleotide and amino acid biosynthesis, as well as those that are more unique to cyanobacteria such as photosynthesis, carbon fixation, and cyanophycin production. Photosynthesis was modeled as a set of three sequential reactions that involve photosystem II, the cytochrome $b_6f$ complex, and photosystem I. This was done in order to study how different wavelengths of light and separate photosystem activities affect electron flow through photosynthesis and respiration, nitrogen source utilization, cellular growth, and hydrogen production (Figure 1). Using batch data collected with different qualities of light, we have used the model to estimate energy requirements for *Cyanothece* 51142. We have also used the model to investigate how fluxes through the different photosystems affects cellular growth and to estimate the fluxes through linear and cyclic photosynthesis pathways, cytochrome oxidases, NADH dehydrogenases, Mehler reactions, and ferredoxin-NADP reductase. Using a custom-built photobioreactor, which allows for the control and monitoring of incident and transmitted light, we have also studied the physiological response of *Cyanothece* 51142 to nitrogen limitation and light limitation imposed on photosystems I and II (Table 1). Biomass composition, gene expression, and protein expression measurements were also carried out for these two chemostat conditions. These experimental datasets have been incorporated into the model to improve flux predictions for both the nitrogen and light-limited chemostat conditions. For these simulations, fluxes were favored through reactions whose proteins were detected, while fluxes through reactions catalyzed by poorly expressed genes were avoided.

Table 1. Model Predicted Growth Rates *vs.* Experimentally Determined Growth Rates Under Variable Light Quality and Intensity

| Batch # | Photon flux through PSII (mmol/gDW/$h^{-1}$) | Photon flux through PSI (mmol/gDW/$h^{-1}$) | Experimental growth rate ($h^{-1}$) | Predicted growth rate ($h^{-1}$) | Error, % |
|---|---|---|---|---|---|
| 3-A | 35.0 | 34.6 | 0.079 | 0.074 | 6.3 |
| 4-A | 26.4 | 53.6 | 0.080 | 0.085 | 6.3 |
| 6-A | 15.5 | 19.0 | 0.035 | 0.035 | 0 |
| 9-A | 26.0 | 15.6 | 0.041 | 0.043 | 4.9 |
| 11-A | 13.6 | 33.4 | 0.051 | 0.049 | 3.9 |

We have additionally developed a draft metabolic model for *Synechococcus* sp. strain PCC 7002, a fast-growing, non-nitrogen-fixing cyanobacterium which is remarkably tolerant to high light intensities. Understanding the origin of these properties could provide fundamental new insights that could be widely applied to the development of other biological systems for biofuel development. The *Synechococcus* 7002 model contains 614 genes and 552 reactions. Initial comparisons between the reconstructed metabolic networks of *Cyanothece* 51142 and *Synechococcus* 7002 suggested that both networks share a significant number of pathways. Some differences between the two models involve reactions associated with acetate metabolism, adenosyl-homocysteine recycling, and glycoaldehyde transport. Once the model for *Synechococcus* 7002 is complete, we will apply metabolic engineering algorithms to identify genetic strategies for improving chemical and biofuel production.

Early Career Program
Speaking Wednesday 11:10 a.m.

# 73

## Applying the Biology of Brown Rot Fungi to Consolidated Bioprocessing

**Jonathan S. Schilling*** (schilling@umn.edu)

University of Minnesota

**Project Goals: Resolve the reaction partitioning mechanisms of brown rot fungi so that they can be applied to bioconversion of lignocellulosic feedstocks.**

**Hypothesis:** During brown rot, oxidative pretreatments occur ahead of enzymatic saccharification, spatially, and the fungus partitions these reactions using gradients in pH, lignin reactivity, and plant cell wall porosity. These can be recreated without the fungus present for faster bioconversion, integrating otherwise incompatible steps.

Consolidated bioprocessing (CBP) of lignocellulose combines enzymatic sugar release (saccharification) with fermentation, but pretreatments remain separate and costly. In nature, lignocellulose-degrading brown rot fungi

consolidate pretreatment and saccharification, likely using spatial gradients to partition these incompatible reactions. To characterize this relevant biological system, my objectives are to (1) physically sample wood degraded by the brown rot fungus *Postia placenta*, spatially map the coincident locations of pretreatment and saccharification reactions, and correlate with pH and lignin chemistry; (2) to 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 cellulase involved in saccharification. These are spatially focused goals. Therefore, my respective approaches involve either small-scale, spatially resolved characterization (Objective 1) or appropriately resolved microscopy (Objectives 2 and 3). Small-scale physical sample analysis includes traditional wet chemical characterization, coupled both with spin-trap adduct recovery of hydroxyl radicals produced by the fungus and with $C^{13}$-labeled tetramethyl ammonium hydroxide thermochemolysis for specific brown rot lignin modifications. For microscopy, I am using fluorescence lifetime imaging (FLIM) with confocal detection for pH measurements, cryo-transmission electron microscopy (TEM) with electron tomography for porosity measures, a complementary scanning transmission X-ray approach for porosity, and traditional TEM with immunolabeling to track cellulase ingress. I am also planning to co-localize chitinous fungal biomass, imaged using a traditional WGA-FITC dye, with fluorescence in-situ hybridization to measure mRNA transcribed from iron reductase and endoglucanase DNA sequences, made possible by the recent DOE JGI-funded annotation of the *P. placenta* sequence. Overall, this research will help resolve how brown rot fungi consolidate oxidative pretreatments with enzyme-based saccharification, so that we might better understand and exploit natural synergies between bioconversion steps currently approached as separate, distinct steps.

Student Oral Presentation–Tuesday

# 74

## Biomimicry of the Fungal Consolidation of Biomass Pretreatment with Saccharification

Justin T. Kaffenberger* (kaffe002@umn.edu) and **Jonathan S. Schilling**

Department of Bioproducts and Biosystems Engineering, University of Minnesota, St. Paul

http://faculty.bbe.umn.edu/schilling

**Project Goals: The goal of this project is to incorporate systems biology to elucidate the brown-rot fungal mechanism for biomass degradation. As a proven evolved biological means of consolidating chemical pretreatment and saccharification of biomass, this mechanism can guide process design in the consolidation of commercial biomass conversion.**

Efficient breakdown of ligno-cellulosic material prior to fermentation for bio-fuel production is a key step in optimizing bio-fuel production. Current industrial practice involves chemical pretreatment and acid or enzymatic saccharification to prepare ligno-cellulosic biomass for fermentation. The commercial viability of bio-fuels hinges in large part on production cost. One approach to reducing production cost is process consolidation. Consolidation not only eliminates the capital costs associated with the additional processing steps, but can also trim operational costs by reducing total reaction time and down time between steps.

In nature, the brown-rot fungus *Postia placenta* successfully performs both chemical pretreatment and enzymatic saccharification of biomass in a consolidated manner. As evidenced by its recently sequenced genome, *P. placenta* produces a limited suite of cellulases, lacking genes for exoglucanases. Despite this limitation, it is still capable of rapidly degrading wood. Brown-rot fungi quickly reduce the degree of polymerization of wood with little weight loss through a reduction-oxidation pathway that yields highly reactive hydroxyl radicals through a chelator-mediated fenton reaction. While hydroxyl radicals would readily damage its cellulases, the fungus manages to conduct both reaction types, enzymatic and oxidative, in the same general location. Examination of the spatial and temporal relationship of these two reaction systems as brown-rot fungal degradation of wood progresses will be presented.

# 75

## Integrated Nondestructive Spatial and Chemical Analysis of Lignocellulosic Materials During Pretreatment and Bioconversion to Ethanol

**Gary Peter*** (gfpeter@ufl.edu), Kyle Lundsford, Sharon Portnoy, Choong Heon Lee, Steve Blackband, Lonnie Ingram, and Richard Yost

University of Florida, Gainesville

**Project Goals: Our *long-term goal* is to develop a quantitative structural model for changes that occur in the organization and chemical composition of plant biomass during pretreatment, enzymatic degradation and bioconversion to ethanol or other products. The *objectives* of this proposed work are to 1) use advanced high resolution magnetic resonance microscopy (MRM), micro and nano x-ray computed tomography (x-ray CT), electron microscopy and imaging mass spectrometry (IMS) to quantify changes in the architecture, porosity, permeability, surface area, pore size, interconnectivity, chemical organization and composition of bagasse and particularly *Populus* and pine wood chips during pretreatment and enzymatic degradation, and 2) integrate the results from these different quantitative imaging methods into a model for disassembly of the plant cell wall during pretreatment and bioconversion. We are developing methods for imaging biomass with MRM, x-ray CT and IMS.**

IMS with a MALDI linear ion trap + MS: The full-scan MS of biomass shows intense ions at every mass-to-charge ratio (*m/z*), making the analysis very complex. Because so many ions are present, we analyzed standard compounds that are normally present in wood. Full-scan and $MS^n$ spectra were obtained for β 1,4-glucan, 4-O-methylglucouronxylan, β-glucan, starch and microcrystalline cellulose. $MS^2$ spectra from intact wood and holocellulose show ions characteristic of cellulose and 4-O-methylglucouronxylan. The MS image of m/z 1319 a $[glu]^6$ oligomer characteristic of cellulose shows a uniform signal in young *Populus* stems; however, $MS^2$ demonstrates two isobaric ions at m/z 1319, one corresponding to cellulose (m/z 995) localized to the secondary xylem and phloem bundles and one additional ion (m/z 1275) localized to the bark (Figure 1). Thus, the use of the linear ion trap (LIT) and $MS^n$ fragmentation are required to increase the selectivity and interpret the complex spectra for mapping the distribution of cellulose, lignin and hemicelluloses within the biomass during pretreatment and hydrolysis.

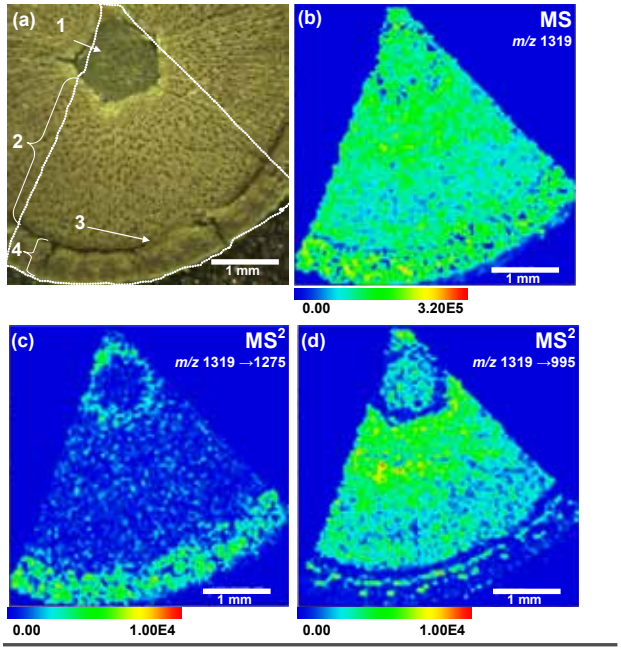



Figure 1. a) optical image, b) MS of m/z 1319, c) MS2 of m/z 1319-1275, d) MS2 of m/z 1319 to 995

Magnetic Resonance Microscopy: Excellent image quality is obtained from *Populus* wood and bagasse samples using T2 and diffusion weighted modes. New images were collected with state of the art rf microcoils at 8 uM resolution. Figure 2 shows the improved signal to noise ratios provide much better images.

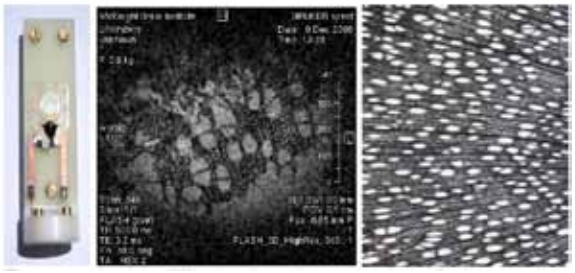


Figure 2. Using the new rf micrcoils (left) this figure shows an MR microimage (8 microns, 4hrs 33 min) of a wood chip sample (middle) with example optical histology 5X objective.

For quantitative diffusion imaging, a series of scans are acquired with diffusion encoding in several different directions. From this data, the calculated diffusion tensor reflects the diffusion anisotropy within materials. Within the wood chip, diffusion occurs preferentially along, rather than across, the xylem/phloem vessels. Figure 3 shows the resultant fractional isotropy (FA) map that is quantitatively generated. Blue represents no anisotropy (as seen in the water surrounding the wood chip), while within the chip there is considerable anisotropy (red and yellow speckling). Note the blue regions in the wood chip, which represent the isotropic diffusion of the relatively unrestricted water within the xylem vessels.

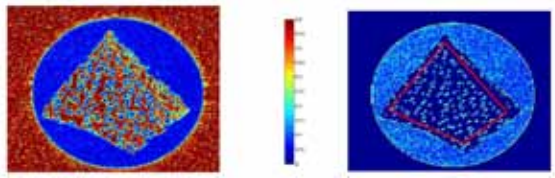
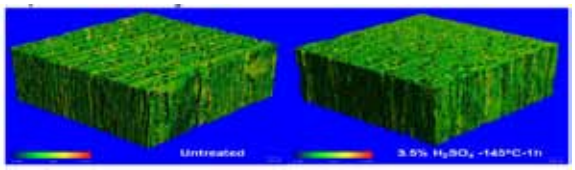


Figure 3. Wood chip FA map with the vessels are clearly defined (blue holes). TR/TE=2000/35.6, FOV=0.6 cm, in-plane resolution=31 µm, slice thickness = 1.0 mm, 19 averages, diffusion time (Δ)=25 msec, b-value =900 s/mm2, temp=40°C. A similar FA map (right), with all the fiber and ray cells segmented out. Increased anisotropy can be seen at the edges of the xylem.

X-ray micro CT: Excellent images have been obtained at high resolution from *Populus*, pine, and bagasse samples. In addition to the basic density, images are readily segmented and the material and airspace sizes can be quantified. Small changes in surface area and surface area to volume ratios were observed after dilute acid pretreatment. Micro CT imaging of the recalcitrant material shows thinner cell walls and some degree of degradation on the periphery compared with the internal regions of the bundles.



# 76

## Towards Real-Time and High Throughput Molecular Characterization of Microbial Deconstruction of Biomass

Zhao Hao* (ZHao@lbl.gov), Ping Hu, Darcie Long, Andrew Tauscher, and **Hoi-Ying N. Holman**

Lawrence Berkeley National Laboratory, University of California, Berkeley

**Project Goals: Our project aims to (1) develop a high throughput methodology to monitor and characterize key chemical and biological processes during microbial deconstruction of plant biomass and conversion to biofuels directly, and (2) apply the technology to investigate enzyme-microbe synergy during biomass deconstruction.**

Synchrotron radiation-based Fourier transform infrared (SR-FTIR) spectromicroscopy is a label-free, non-invasive molecular technique that couples the high brightness of synchrotron radiation with the high throughput and vast analytical capabilities of FTIR spectrometers. One key challenge to our effort to apply SR-FTIR spectromicroscopy to bioenergy research has been the richness of infrared spectral features arising from the complex systems of lignocellulose biomass, the biomass degraders, and their interactions. More specifically, we are facing an extremely large amount of data from the multidimensional hyperspectral imaging, where the number of data points is on the order of billions. A suite of highly efficient methods is needed to systematically identify the evolving patterns and features of different chemical components. Furthermore, the dynamics of the infrared signatures of many important biochemical molecules during the deconstruction and conversion of plant biomass to biofuels have yet to be characterized fully. To improve confidence in interpretation of spectral data, we are conducting well-controlled experiments using model compounds and organisms. In parallel, we utilize quantum chemical simulation and modeling to guide our analysis of infrared measurements. During the first phase of our effort, we investigated model systems important to biofuel production, such as *Clostridium* actions on bioenergy crops. We also compared the biomass deconstruction processes between the *Clostridium* cellulosomes and the living *Clostridium* cells. A comparison between the bacteria/plant biomass and cellulosomes/plant biomass systems reveals direct molecular evidence of microbe-enzyme synergy. We found that the high throughput SR-FTIR technology under development at the Lawrence Berkeley National Laboratory can potentially aid in the optimization of microbes for an accelerated bioprocessing of plant biomass into biofuels in the future.

# 77

## Proteomic Measurements of Redox Sensing and Post-Translational Modifications in Cyanobacteria

T.C. Squier[1]* (thomas.squier@pnl.gov), D.J. Bigelow,[1] R.N. Brown,[1] D.A. Bryant,[2] B. Chen,[1] S. Datta,[1] N. Fu,[1] A.E. Konopka,[1] M.S. Lipton,[1] W. Qian,[1] S. Stolyar,[1] D. Su,[1] A.T. Wright,[1] and **A.S. Beliaev**[1]

[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash.; and [2]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park

**Project Goals: The PNNL Biofuels Scientific Focus Area (BSFA) will conduct fundamental research on cyanobacteria with specific emphasis on pathways of carbon, nitrogen, and redox metabolism that consume reductant and conserve energy produced by photosynthetic light reactions. Consistent with DOE BER Genomic Sciences Program goals, the long-term objective of the BSFA is to develop a predictive understanding of metabolic subsystems and regulatory networks involved in solar energy conversion to biofuel precursors or products. Toward this goal, we will interrogate fluxes through metabolic subsystems that include (i) photosynthetic energy conservation and reductant generation, (ii) $CO_2$ accumulation, fixation, and reduction, (iii) biosynthesis of metabolic intermediates and monomers, and (iv) macromolecular synthesis. Our studies will exploit two complementary model cyanobacterial systems, *Cyanothece* sp. strain ATCC 51142 and *Synechococcus* sp. strain PCC 7002, and address discrete scientific hypotheses.**

As part of PNNL's Scientific Focus Area "Genome-Enabled Studies of Photosynthetic Microorganisms for Bioenergy Applications" this project is applying and developing approaches for live cell imaging that in conjunction with advanced separations, mass spectrometry and traditional biochemical techniques are being used to identify redox sensors of environmental change and post-translational protein modifications associated with intracellular signaling. Specifically, we seek to identify sequence-specific motifs within individual protein sensors in *Synechococcus* PCC 7002 that are targets for post-translational modification (PTM). We are investigating the central hypothesis that site-specific PTMs modulate cellular pathways in response to environmental change, and that their identification is critical to an ability to construct predictive models of metabolism needed for bioenergy applications. Investigated PTMs include glycosylation, phosphorylation, and redox-dependent alterations involving cysteine (e.g., reversible dithiol formation, glutathione adduct formation, and oxidation to form sulfenic acid). Using existing and newly developed affinity-based methods, we have fractionated cells and using advanced separations and mass spectrometry identified the following targeted PTMs in cellular proteins. All modified peptides are analyzed with tandem mass spectrometry and identified

with SEQUEST using the mass of the phosphoryl or glycosyl group as the modification on the specific target amino acids.

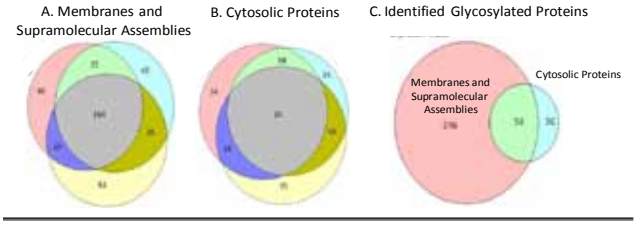



Figure 1. Identification of Glycosylated Proteins. 406 glycosylated proteins in *Synechococcus* 7002. Identified by LC-MS/MS following enrichment using a lectin column. Reproducibility (A, B) of sample identification is shown for (A) membrane and (B) cytosolic protein fractions. (C) Summary of identified glycol modifications identified from three replicate measurements.

*Glycosylation:* To identify how enzyme-mediated glycosylation may regulate metabolism, we have used cell fractionation and lectin affinity chromatography to purify glycosylated proteins from *Synechococcus* 7002 cell lysates. Isolated glycoproteins were digested with trypsin followed by high-resolution LC-MS/MS. A total of 330 proteins (containing ≥2 confidently identified peptides) were identified in the membrane fraction, whereas 130 proteins were found in the cytosolic fraction, with very high reproducibility from three separate replicates (Figure 1). Identified proteins include well studied examples of known glycosylated proteins, including the S-layer like proteins (SYNPCC7002_A2813, SYNPCC7002_A1634, and SYNPCC7002_A1178). A total of 406 different proteins were identified (Figure 1C). Identification of the majority of glycosylated proteins within the membrane fraction is consistent with the hypothesized functional role of glycosyl-transferase proteins in regulating protein localization and cellular function. Identified proteins are being mapped onto their respective cellular pathways, with the goal of correlating changes in glycosylation patterns with environmentally-induced changes in cellular function.

*Phosphorylation:* Phosphoester linkages on serine, threonine, and tyrosine protein side-chains are known to regulate cellular metabolism and intracellular signaling, and are routinely identified at PNNL using metal affinity chromatography (IMAC) enrichment. To characterize phosphorylation in *Synechococcus* 7002 and understand their relationship to biomass production, we implemented an automatic IMAC system for high-throughput phosphopeptides enrichment. In this system, samples are being enriched on a normal flow "large IMAC" column with optimized binding using an autosampler controlled by in-house developed software. The system allows for large quantities of starting material and effectively removes nonspecifically-bound peptides using optimized washing protocol. Initial testing of the new automated platform was carried out for *E. coli* phosphoproteome measurements and we achieved an average absolute mass accuracy of 0.9 ppm. Out of 960 total detected phosphopeptides, we identified 147 unique peptides from 119 *E. coli* proteins which included31 unique phosphopeptides previously identified in the literature. An increased coverage of putative phosphopeptide species provided by the automated system will be critical to our ability to characterize phosphorylation patterns in *Synechococcus* 7002, and, subsequently our understanding of regulatory mechanisms involved in central carbon and energy metabolism.

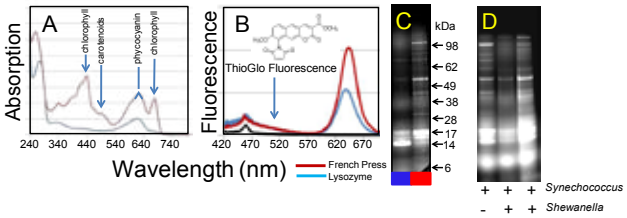



Figure 2. Identification of Thiol-Reactive Proteins within Thylakoids of *Synechococcus*. (A) Absorption spectra, (B) fluorescence emission spectra, and (C) SDS-PAGE showing thiol reactive proteins for lysates from *Synechococcus* before (blue) and after (red) enrichment for thylakoid membranes.

Cysteine Modification: We hypothesize that dithiol "switches" (e.g., Cys-$Xxx_n$-Cys, n < 6) activated by thioredoxin-dependent pathways regulate electron flux and carbon sequestration to modulate energy partitioning in *Synechococcus* PCC 7002. In addition, modifiers of redox-dependent regulation are likely to include reversible modifications of cysteine previously identified to regulate intracellular metabolism, including S-nitrosylatin, glutathiolation, and the formation of sulfenic acid. Identification of redox-dependent dithiols involves two complementary approaches. Using newly synthesized chemical probes, visualization, down-stream purification, and subsequent LC-MS/MS identification of all thiol-reactive proteins are facilitated. Initial results indicate that the majority of thiol-reactive proteins in *Synechococcus* 7002 are located within the thylakoid membrane (Figure 2). These results are consistent with expectations based on homologies with higher plants, where redox-dependent dithiol formation is linked to the regulation of energy metabolism. Identification of these thiol reactive proteins will provide important insights regarding possible relationships between environmental stress, cellular metabolism, and cell growth. Specifically, we hypothesize that modulation of these redox dependent control mechanisms is critical to the maintenance of cell function and will minimize oxidative stress related to redox imbalance.Complementary in vivo approaches to monitor redox-active dithiol reactivity will use recently synthesized cell-permeable monoarsenic probes that selectively bind to reduced disulfides in close proximity to trap available dithiols in living cells prior to cell lysis. The identification of proteins in *Synechococcus* 7002 that undergo disulfide exchange in response to changes in cellular conditions (e.g., light intensity, nutrient availability) involve incubation of these new cell-permeable reagents. Following reaction with proximal thiols, cells are lysed and following reduction of internal disulfides with TCEP, the bound probe will be released permitting capture and identification of redox active proteins by commonly used thiol-capture affinity methods operational at PNNL. Parallel imaging measurements will assess how environmental conditions affect cellular redox

state monitored as reduced dithiols both in model systems and isolated natural microbial communities to facilitate our understanding of possible relationships between environmental changes, cellular redox state, and intracellular metabolism.

# 78

## Comparative Molecular Imaging Analysis of *Brachypodium distachyon* and Its Mutants

Liang Chen,[1]* Ludmila Tyler,[2,3] Andrew N. Tauscher,[1] **John Vogel,[2]** and **Hoi-Ying N. Holman[1]** (hyholman@lbl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]USDA-ARS Western Regional Research Center, Albany, Calif. ; [3]University of California Berkeley, Albany

**Project Goals: Our project aims to (1) develop and apply Synchrotron-Radiation-based FTIR imaging and analysis techniques for high spatial resolution molecular characterization of differences among *Brachypodium distachyon* lines, and (2) develop and provide a high-throughput approach that can help link chemical information to gene functions in *B. distachyon*.**

The grass *Brachypodium distachyon* (*Brachypodium*) is an important model system for bioenergy grasses (e.g. switchgrass and *Miscanthus*) owing to its small genome size (272Mbp), compact physical stature, short lifecycle, self-fertility and transformability. To date, ~10,000 T-DNA insertional mutant lines have been generated for functional genomics studies. The ability to rapidly compare compositions, structures and functions of different *Brachypodium* lines and their mutants is critical to establish causal links between gene functions and chemical composition. Synchrotron-Radiation-based FTIR (SR-FTIR) can play an important role in determining this information because SR-FTIR is a nondestructive label-free spectroscopic method that derives chemical information directly from characteristic vibrational frequencies of various chemical bonds with diffraction-limited spatial resolution. The fast speed and non-invasive nature of this analytical method make it especially suitable for development into a high-throughput chemical analysis tool for *Brachypodium* studies. As investigators of the Berkeley Synchrotron Infrared Structural Biology (BSISB) program, we aim to develop and apply the SR-FTIR spectro-imaging technique together with multivariate statistics to the rapid chemical analysis of the *Brachypodium* cell wall and the comparative analysis of different *Brachypodium* lines and their mutants.

During the first phase of our technology research and development effort, we used six genetically diverse *Brachypodium* lines selected from a larger collection of 187 natural accessions (Figure 1). As a first step, we performed bulk FTIR measurement on ground stems from the six lines. With Principal Component Analysis-Linear Discriminant Analysis (PCA-LDA) software tools specifically developed for the *Brachypodium* project, we found that the lines could be clearly separated into two groups by their mid-infrared signatures (Figure 2), suggesting underlying compositional and/or structural differences among the lines. Based on this promising initial result, we then used SR-FTIR to obtain high-resolution infrared images of thin sections of these six *Brachypodium* lines. This study reveals the spatial heterogeneity in the chemical composition of the plants. Using comparative image analysis methods, we found that there are significant differences in the spatial distribution of cell wall composition between the six *Brachypodium* lines. The genetic basis of these phenotypic differences will be evaluated by making crosses between lines with different compositions/patterns and identifying the genes responsible using a map-based approach. With our optimized methods we also plan to examine the compositional changes in *Brachypodium* cell wall mutants identified by near infrared spectroscopy in the Vogel lab. Our methods will provide the foundation for future studies linking chemical information to gene functions in *Brachypodium*. Our long-term goal is to standardize and enable full automation of our analytical system for high through-put screening experiments for multiple natural and engineered energy crops.

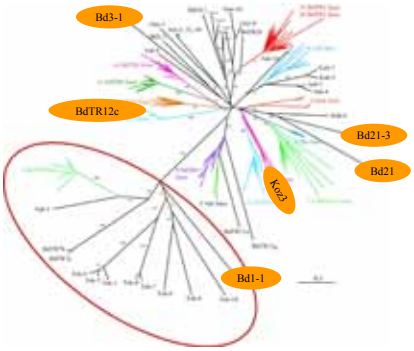



Figure 1. Genotypic diversity in select inbred lines. An unrooted Neighbor-Joining consensus tree of 187 lines based on 100 shared allele bootstrap trees constructed using 43 SSR markers. The lines in the red circle share a number of phenotypic characters (small seeds, nearly hairless lemmas, long vernalization requirements). The lines used in this study are highlighted by orange ovals. Note that Bd21 is the line sequenced for the reference genome. The other five lines used in this study were recently resequenced through the DOE JGI Community Sequencing Program.

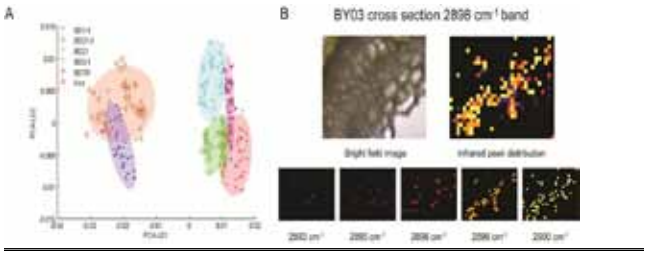



Figure 2. Chemical analysis of *Brachypodium* samples using mid-IR. (A). The PCA-LDA score plot of samples from 6 *Brachypodium* lines. 45 spectra are collected from each line. Spectra data are collected from 600 $cm^{-1}$ to 4000 $cm^{-1}$. The spectra in the region between 800 $cm^{-1}$ and 1200 $cm^{-1}$ are used as input for PCA-LDA analysis. The "Bd1-1" and "Koz-3" samples are clearly separated from the other 4 types of samples. (B). The distribution and shift of the 2898 $cm^{-1}$ band in the cross section of a *Brachypodium* sample thin slice. Each peak position is represented by a different color.

# 79

## Neutrons Reveal Structural Changes During Switchgrass and Poplar Biomass Pretreatment

Sai Venkatesh Pingali[1]* (pingalis@ornl.gov), Volker Urban,[1] William T. Heller,[1] Hugh M. O'Neill,[1] Marcus Foston,[2] Dean A. Myles,[1] Art Ragauskas,[2] Barbara R. Evans,[1] and **Brian H. Davison**[1]

[1]Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [2]Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta

Program: SFA Biofuels

**Project Goals: Lignocellulosic biomass is recalcitrant to deconstruction and saccharification due to its fundamental molecular architecture and multicomponent laminate composition. A fundamental understanding of the structural changes and associations that occur at the molecular level during biosynthesis, deconstruction, and hydrolysis of biomass is essential for improving processing and conversion methods for lignocellulose-based fuels production. 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.**

**Small-angle neutron scattering (SANS) was used to obtain a better understanding of the morphology of the cellulose/lignin composite in switchgrass and poplar to aid in understanding and ultimately selecting biomass pretreatment methods that are required to prepare lignocellulosic biomass for conversion to ethanol.**

Here, we present results from the dilute acid pretreatment method used to break down switchgrass and poplar biomass. Results are from both ex-situ experiments as well as a new in-situ reaction cell which allows direct SANS monitoring of changes during pretreatment.

From our ex-situ studies, the similar, basic trends in structural and morphological changes in switchgrass and poplar biomass during dilute acid pretreatment are: (1) increase in the small-scale structure which can be related to the crystalline core cross-section; (2) decrease in the interconnectivity of the biopolymers; (3) formation of additional distinct structures at length scales 50-200 Å that are due to formation of lignin aggregates which first appears around 125-150˚C; and (4) at length scales larger than 1000 Å, no change in the smooth domain boundaries until 60 min at 160˚C. On the other hand, the inherent differences in the chemical composition, structure and morphologies of biomass grass (switchgrass) when compared to trees (poplar) cause many differences in the details of the structural changes during dilute acid pretreatment, such as: (1) larger increase in the small-scale structure than a single crystalline microfibril indicating coalescence of more neighboring crystalline microfibrils; (2) larger lignin aggregate size formation and growth; and (3) rougher surface morphology of the micron-sized cell walls.

The results from the in-situ study demonstrated the appearance of a distinct additional structure of $R_g$ ~30 Å after attaining 120˚C. A maximum temperature of 180˚C was used for this study. The size of the additional structure, after 30 min at 180˚C was $R_g$ ~700 Å and thereafter was larger than our accessible particle-size length scales. If the additional structure is interpreted as lignin aggregates, these undergo much larger growth during the in-situ process and the first appearance of the lignin aggregates occurs at a slightly lower temperature (120˚C) when compared to the ex-situ study, above 125˚C.

# 80

## Integration of Small-Angle Neutron Scattering and Computational Modeling to Investigate the Structures of Lignin Aggregates and the *Trichoderma reesei* Cellobiohydrolase I

Loukas Petridis[1]* (petridisl@ornl.gov), Sai Venkatesh Pingali,[1] Volker Urban,[1] William T. Heller,[1] Jeremy C. Smith,[1] Hugh M. O'Neill,[1] Marcus Foston,[2] Dean A. Myles,[1] Art Ragauskas,[2] Barbara R. Evans,[1] and **Brian H. Davison**[1]

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee; and [2]Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta

Program: SFA Biofuels

**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.**

**Computational modeling and small-angle neutron scattering (SANS) were used to examine the molecular architecture of lignin across a broad range of length scales. Similarly, this combined approach was used to understand the solution structure of cellobiohydrolase I from *Trichoderma reesei*, which breaks down cellulose into cellobiose, and how the enzyme responds to changes in pH.**

Here, we report on two studies using this combined approach to better understand lignin and the cellobiohydrolase I (Cel7A), the major cellulase enzyme component from the fungus *T. reesei*. Lignin, a major polymeric component of plant cell walls, forms aggregates in vivo after pretreatment of lignocellulosic biomass for ethanol production. The aggregates are thought to reduce ethanol yields by inhibiting enzymatic hydrolysis of cellulose. Extensive atomistic molecular dynamics (MD) simulations using models built with information provided by chemical analysis, $^{13}C$- and $^{1}H/^{2}H$- NMR studies and SANS were used to demonstrate that the surfaces of the aggregates are highly irregular and characterized by a surface fractal dimension that is invariant across length scales from ~1–1000Å. The simulations reveal extensive water penetration of the aggregates and heterogeneous chain dynamics corresponding to a rigid core with a fluid surface. The detailed multiscale structure revealed here should aid in understanding biomass recalcitrance to hydrolysis and in feedstock engineering efforts to improve biofuel yield. The solution structure of Cel7A was studied as a function of pH using SANS and computational modeling. The enzyme, which consists of a large catalytic core connected to a much smaller cellulose binding domain by a flexible linker, has a pH optimum of 4.5. The SANS data demonstrate that a conformational change takes place as the pH is lowered from 7.0 to 4.2. The spatial relationship between the cellulose binding domain and the catalytic core changes between pH 7.0 and 5.3. At pH 4.2, the enzyme transitions to a flexible conformation that is intermediate to a tightly-folded structure and a disordered state. The results suggest that the increased flexibility observed in the structure of the enzyme is necessary for optimum activity of the enzyme. The insight provided by this work could aid in the engineering of enzymes optimized for converting lignocellulosic biomass into biofuels.

# 80A‡

submitted post-press

## Label-Free, Real-Time Monitoring of Biomass Processing with Stimulated Raman Scattering Microscopy

Brian G. Saar,[1] Yining Zeng,[2] Christian W. Freudiger,[1] Yu-San Liu,[2] Michael E. Himmel,[2] **Shi-You Ding**[2]* (Co-PI: Shi.you.Ding@nrel.gov), and **X. Sunney Xie**[1]* (PI: xie@chemistry.harvard.edu)

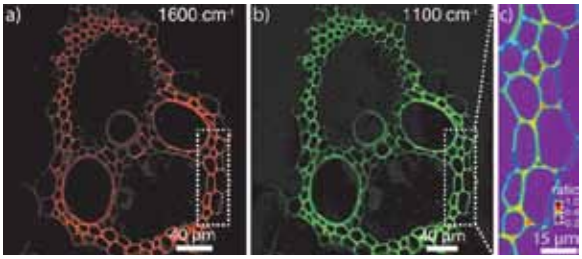
[1]Department of Chemistry and Chemical Biology, Harvard University and [2]Biosciences Center, BioEnergy Science Center, National Renewable Energy Laboratory

**Project Goals: The objective of this project was to capitalize on and to further develop Coherent Raman Scattering (CRS) microscopy in order to characterize and image lignin and polysaccharides in plant cell walls, and to monitor plant cell wall deconstruction in real time in biomass conversion processes from a variety of agricultural products. The specific aims were (1) construction of a new generation of CRS microscope with frequency modulation that offers much improved sensitivity in order to selectively image cell wall constituents in plant cell walls. (2) Use of the new microscope to monitor and study the dynamic structural and compositional changes of plant cell wall macromolecules during their chemical and enzymatic degradation *in situ*. (3) Development of a near-field CRS microscope with nanometer spatial resolution in order to probe the structure and dynamics of plant cell wall processing.**

We demonstrate that stimulated Raman scattering (SRS) microscopy [1,2], a new imaging method, allows real time observation of biomass conversion processes. Current analytical methods, such as gas chromatography–mass spectrometry, electron or scanning-probe microscopy, and fluorescence microscopy, have restrictions. Microscopy based on infrared absorption offers chemical specificity, but spatial resolution is limited by long infrared wavelengths, and penetration depth into aqueous plant samples is small. Raman microspectroscopy offers label-free chemical contrast with high resolution and chemical specificity. However, the Raman scattering effect is weak, and long pixel dwell times are required for imaging plant materials. This means that real-time imaging is challenging, requiring a long data collection time. SRS microscopy offers chemical contrast based on the intrinsic Raman vibrational frequencies in a sample with much shorter imaging time and easier spectral identification. SRS superseded CARS microscopy [3], another nonlinear vibrational imaging method our group has developed in the past. The SRS imaging technique for studying the conversion process *in situ* offers chemical specificity without exogenous labels, non-invasiveness, high spatial resolution, and real-time monitoring capability. We demonstrate its utility for the study of biomass conversion. [4]

**References**
**(publications supported by this project)**

1. Freudiger, Christian W.; Min, Wei; Saar, Brian G.; Lu, Sijia; Holtom, Gary R.; He, Chengwei; Tsai, Jason C; Kang, Jing X.; Xie, X. Sunney "Label-Free Biomedical Imaging with High Sensitivity by Stimulated Raman Scattering Microscopy," *Science*, 322, 1857-1861 (2008).
2. Saar, Brian G.; Freudiger, Christian W.; Reichman, Jay; Stanley, C. Michael; Holtom, Gary R.; Xie, X. Sunney "Video-Rate Molecular Imaging in Vivo with Stimulated Raman Scattering" *Science*, 330, 1368-1370 (2010).
3. Zeng, Yining; Saar, Brian G.; Friedrich, Marcel G.; Chen, Fang; Liu, Yu-San; Dixon, Richard A.; Himmel, Michael E.; Xie, X. Sunney; Ding, Shi-You "Imaging Lignin-Downregulated Alfalfa Using Coherent Anti-Stokes Raman Scattering Microscopy" *Bioenerg. Res.* (2010).
4. Saar, Brian G.; Zeng, Yining; Freudiger, Christian W.; Liu, Yu-San; Himmel, Michael E.; Xie, X. Sunney; Ding, Shi-You "Label-Free, Real-Time Monitoring of Biomass Processing with Stimulated Raman Scattering Microscopy" *Angew. Chem. Int. Ed.* 49, 5476-5479 (2010).




Imaging of lignin and cellulose with SRS microscopy. a) SRS image at 1600 $cm^{-1}$ of a vascular bundle, showing the lignin distribution with a red intensity grade. b) SRS image of the same vascular bundle as in (a), showing the cellulose distribution at 1100 $cm^{-1}$ with a green intensity grade. Both (a) and (b) were obtained with a 50 ms pixel dwell time. These images can be acquired simultaneously using the two-color SRS instrument. c) Ratio of the lignin divided by the cellulose signal at higher magnifi-cation, obtained from the region surrounded by the dotted line in (a) and (b).

# Systems Biology and Metabolic Engineering Approaches for Biological Hydrogen Production

# 81

## A Systems Biology Approach to Energy Flow in $H_2$-Producing Microbial Communities: Multi-Scale Measurement of Metabolic Function in Complex Microbial Ecosystems

Xavier Mayali[1]* (mayali1@llnl.gov), Kevin J. Carpenter[1]* (carpenter37@llnl.gov), Javier Ceja-Navarro[2]* (JCNavarro@lbl.gov), Peter K. Weber,[1] Shalini Mabery,[1] Paul D. Hoeprich,[1] Ulas Karaoz,[2] Todd Z. DeSantis,[2] Meredith M. Blackwell,[3] Stephanie Gross,[3] **Eoin L. Brodie**,[2] and **Jennifer Pett-Ridge**[1]*

[1]Chemical Sciences Division, Lawrence Livermore National Laboratory, Livermore, Calif.; [2]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; and [3]Department of Biological Sciences, Louisiana State University, Baton Rouge

**Project Goals: The objective of this research is to develop an integrated analysis of energy flow in complex microbial communities for the purpose of optimizing biofuel production. We will combine biogeochemical, stable isotope probing, metatranscriptomic and computational approaches, with the aim of understanding nutrient cycling and biofuel (specifically, $H_2$) production in microbial communities. Our ultimate goal is the development of multi-scale models that can predict ecological and biochemical relationships within multi-trophic microbial systems.**

A fundamental goal of microbial ecology is to understand the biogeochemical role of individual microbial taxa in their natural habitat. This is a complex problem because most microbes remain uncultivated and most microbial communities are very diverse. Thus direct analysis of isolated strains results in a skewed and possibly non-representative ex-situ view of functional capabilities while culture-independent (metagenomic) approaches mostly provide hypotheses as to which of the many biogeochemical processes organisms or communities perform.

To directly measure functional roles of uncultivated microbes, we have developed a multi-scale approach. Beginning at the community level, we probe for organisms capable of assimilating substrates in pathways of interest using an approach termed "chip-SIP", a high-sensitivity, high-throughput stable isotope probing (SIP) method performed on a phylogenetic microarray. We incubate microbial communities with isotopically ($^{13}C/^{15}N$) labeled substrates, hybridize community rRNA to a microarray, and measure isotope incorporation—and therefore assimilation—by

secondary ion mass spectrometer imaging (NanoSIMS). Using this approach we quantified amino acid, nucleic acid and fatty acid incorporation by 81 taxa in a marine microbial community. These data enabled the visualization of carbon flow that constrains the functional roles of specific microbial taxa (e.g. substrate generalists vs. specialists) (Fig. 1A).

Our next approach is to directly image substrate assimilation by individual microbial cells to identify the location of activity within a natural system. For this aspect we are currently focusing on protist-bacterial symbioses within the hindgut of lower termites (Fig. 1B). Protists are an extremely diverse, abundant, and important group of microbes with capabilities of interest in alternative energy and climate amongst others. For example, protists of the lower termite hindguts convert lignocellulose into simple carbohydrates while also producing hydrogen gas. Many protist species have tens or hundreds of thousands of bacterial symbionts whose function is poorly understood or unknown. Here we report the first results in exploring these protist-bacterial symbioses using in situ stable isotopic labeling ($^{13}C$ enriched feeding substrates and $^{15}N$ enriched air) followed by analysis of individual protists and their symbiotic bacteria with NanoSIMS.

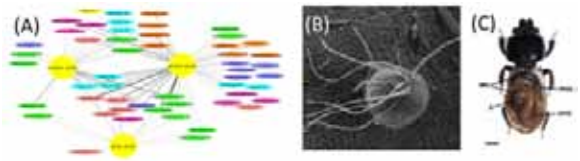



Figure 1. (A) Network diagram showing the uptake and assimilation of nucleic acids, amino acids and fatty acids by a community of microorganisms. (B) Lower termite hindgut protist. (C) Wood-feeding passalid beetle digestive tract.

Our third complementary approach is the analysis of microbial community gene expression in conjunction with micron-scale measurements of physicochemical gradients to provide fundamental physiological context. For this aspect we also focus on wood-ingesting insects (Passalid beetles – Fig. 1C) and the metabolic partitioning through the insect's digestive tract to perform efficient lignocellulose metabolism that results in the production of methane, hydrogen and other potential biofuels. We have adapted mRNA enrichment procedures to deal with the abundance of bacterial, archaeal and eukaryal rRNA prior to meta-transcriptome sequencing. Gene expression data are then used to inform focused isotope-labeling experiments.

Our goal is to combine these approaches into a workflow across the same microbial system to yield a comprehensive model of the flow of energy in these systems.

# 82

## Hydrogen Producers and Consumers in Photosynthetic Microbial Mats Identified by Combining Biogeochemical and Molecular Analysis with Single Cell Techniques

Luke Burow,[1,2] Dagmar Woebken,[1,2] Brad M. Bebout,[2] Lee Prufert-Bebout,[2] Tori M. Hoehler,[2] Steven W. Singer,[3] Alfred M. Spormann,[1] **Jennifer Pett-Ridge**,[4] and **Peter K. Weber**[4]* (weber21@lbl.gov)

[1]Department of Civil and Environmental Engineering, Stanford University, Stanford, Calif.; [2]Exobiology Branch, NASA Ames Research Center, Moffett Field, Calif.; [3]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; and [4]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, Calif.

**Project Goals: We are developing a new method that will provide correlated oligonucleotide, functional enzyme, and metabolic image data, thereby linking identity to function in situ in complex microbial communities. Oligonucleotide and functional enzyme labels, along with biomass labeled by incorporation of stable isotope tracers, will be imaged at the single cell level by nanometer-scale secondary ion mass spectrometry (NanoSIMS). This method, NanoSIP, will improve on the success of previous stable isotope probing (SIP) methodologies in linking microbial identity and function by preserving critical spatial information and allowing finer temporal resolution than previous approaches. NanoSIP will allow us to place coupled genetic and functional information within the context of the physical structure, chemical and physical microgradients, and microbial assemblages of complex communities. NanoSIP will be broadly relevant to DOE's missions in bioenergy, carbon cycling and bioremediation. We will develop our method using Microcoleous and Lyngbya cyanobacterial mats, which are relevant to bioenergy because they are ecological counterparts to the cyanobacterial systems considered for biohydrogen production.**

Hypersaline, photosynthetic microbial mats are diverse communities of microorganisms that can produce hydrogen ($H_2$). The hydrogen ecology of these mats provides a challenging test bed for single cell methods and has the potential to provide insights into microbial interactions relevant to industrial-scale biofuels production. Our investigations of microbial mats to date have focused on characterizing the controls on hydrogen production and consumption, including nitrogen fixation, using a combination of biogeochemical and molecular techniques along with high-resolution imaging mass spectrometry by NanoSIMS. In our poster, we will present data for a microbial mat from Elkhorn Slough, CA that is dominated by *Microcoleus chthonoplastes*. Our data indicate that *M. chthonoplastes* is the primary $H_2$ producer, and that sulfate reducing bacteria (SRB) are the primary $H_2$ consumers. The dominant cyanobacterial nitrogen fixers are

members of a novel clade (UD3), which show significant variability in fixation rates at the single cell level (Fig. 1).

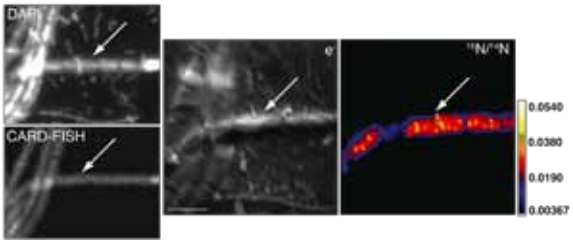



Figure 1. Correlated fluorescence (DAPI and CARD-FISH) and NanoSIMS (secondary $e^-$ and $^{15}N/^{14}N$) images of a microbial mat sample incubated with $^{15}N_2$ and hybridized with a CARD-FISH probe targeting the 16S rRNA of the proposed cyanobacterial clade UD3. The arrow points to a UD3 filament that had a high rate of nitrogen fixation relative to the adjacent UD3 filaments and other microbes (DAPI) based on its enrichment in $^{15}N$ ($^{15}N/^{14}N$ ~0.03). NanoSIMS analysis showed significant variability in N-fixation rates.

Nitrogen fixation has been hypothesized to be a major driver of hydrogen production in microbial mats because of the strong correlation between $N_2$-fixation and hydrogen production in field samples. We performed manipulation experiments to differentiate among potential mechanisms of $H_2$ production. Suppression of $N_2$-fixation by addition of excess ammonium to mats did not change the amount of $H_2$ or organic acids produced, demonstrating that $H_2$ production occurs independently of $N_2$-fixation and likely via fermentation. Depriving mats of daytime sunlight resulted in a ~20-fold decrease in the subsequent nighttime $H_2$ production, suggesting that photosynthate is the energy source for $H_2$-producers. Homogenization of the mat to disruption microbial associations increased $H_2$ production ~2.5-fold, indicating that physical proximity of $H_2$-consumers to $H_2$-producers is an important factor in $H_2$ consumption. Addition of sodium molybdate to inhibit sulfate-reducing bacteria (SRB) also increased net $H_2$ production ~2.5-fold, implying that SRBs are important $H_2$-consumers. Pyrosequencing of small subunit rDNA of the upper strata of nighttime samples revealed the presence of more than 60 phyla across all domains of life (Fig. 2A). Analysis of rRNA:rDNA pyrotags, although not definitive, suggest that OTUs affiliated with Cyanobacteria (including Oscillatoriales) and Chloroflexi (including Chloroflexaceae) are the dominant active microorganisms (Fig 2B). $H_2$-evolving NiFe $H_2$-ases associated with *M. chthonoplastes* dominated clone libraries derived from the upper strata using degenerate primer sets that targeted a diversity of phototrophic organisms. Tiled $H_2$-ase microarray data (~44K probes targeting ~1.2K $H_2$-ases) show that the *M. chthonoplastes* $H_2$-ases are up-regulated 3-fold at night compared to the day.

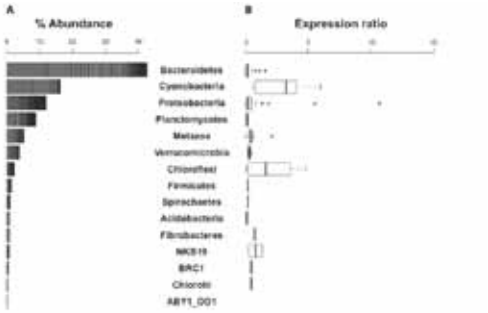



Figure 2. (A) More than 63 phyla across all domains of life were present in the upper 2 mm strata as determined by rDNA sequencing (15 most abundant phyla shown). (B) OTUs belonging to Cyanobacteria and Chloroflexi had relatively higher rRNA:rDNA ratios (expression ratios), suggesting that organisms in these phyla are more active relative to organisms in other phyla present in the mats.

While our manipulations indicate that N-fixation is not the direct driver of $H_2$ production, the possibility still remained that $H_2$ production and N-fixation are performed by the same dominant microbe, *M. chthonoplastes*. However, $^{13}C$-$NaHCO_3$ and $^{15}N_2$ incubations of mat samples followed by high-resolution secondary ion mass spectrometry by NanoSIMS showed that *M. chthonoplastes* has high rates of C-fixation but no detectable N-fixation. Sequencing of the expressed genes encoding for the nitrogenase reductase (*nifH*) showed that 36% of the sequences belonged to a previously unknown cyanobacterium. An enrichment culture of this cyanobacterium (UD3) allowed us to link this *nifH* sequence clade to a 16S rRNA sequence, and 16S rRNA sequence analysis showed that UD3 is only distantly related to any sequence in the databases (max. 92% sequence identity). Catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH) with probes specific for this novel cyanobacterium and subsequent NanoSIMS analyses showed that 69% of the UD3 related population in the microbial mats incorporated $^{15}N_2$ (Fig. 1).

# 83

## Developing Techniques for Bioenergy Research Using Plant and Microbial Cultures

Catherine Lacayo[1]* (lacayo2@llnl.gov), Yongqin Jiao[1]* (jiao1@llnl.gov), Ali Navid,[1] Ben Stewart,[1] Mona Hwang,[1] Shi-You Ding,[2] Alexander Malkin,[1] Jennifer Pett-Ridge,[1] and **Michael P. Thelen**[1]

Lawrence Livermore National Laboratory, Livermore, Calif.; and [2]National Renewable Energy Laboratory, Golden, Colo.

**Project Goals: Our research aims to develop an integrated analysis of energy flow in complex microbial communities. We are combining biogeochemical, stable isotope probing, metatranscriptomic and computational approaches, to understand nutrient cycling and biofuel ($H_2$) produc-**

**tion production in complex microbial communities. A comprehensive understanding of such communities is needed to develop efficient, industrial-scale processes for microbial $H_2$ production and lignocellulose degradation. Experimentally, we are focusing on natural models for $H_2$ production and degradation of plant/microbial biomass: $H_2$-producing cyanobacterial mats and insect hindgut communities, and polysaccharide biogenesis and deconstruction in plant cells and microbial mats. Our ultimate goal is the development of multi-scale models that can predict ecological and biochemical relationships within multi-trophic microbial systems. Our work includes both technique development and biological research components.**

To gain a better understanding of the mechanisms of plant cell wall breakdown, and to accelerate efforts to convert lignocellulosic materials to biofuels, we examined two distinct elements involving degradation followed by fermentation. To visualize deconstruction of natural plant lignocellulose, we monitored the enzymatic degradation of the wall of wood cells in culture. For biofuel production, we investigated the interaction of two disparate bacteria in a co-culture system that degrades cellulose and produces hydrogen gas.

To monitor the response of plant cell walls to chemical, enzymatic and microbial degradation, we are imaging single cultured wood cells from the plant *Zinnia elegans* using both optical and topographical methods. We have evaluated the effects of exogenous enzymes, such as fungal cellulases and hemicellulases, on the cell wall as a function of time. Loss of cellulose in the wall was measured using fluorescent probes, such as carbohydrate-binding modules (CBMs) from *Clostridium thermocellum*. Using atomic force microscopy, we observed structural changes consistent with the gradual removal of cellulose fibrils from the wall.

As a consolidated bioprocessing approach, cellulose-degrading organisms are often co-cultured with organic acid-consuming microbes. However, the metabolic interactions between the organisms are poorly understood. In a co-culture containing *Clostridium cellulolyticum* and *Rhodopseudomonas palustris*, we analyzed the kinetics of cellular growth as well as carbon fluxes in both organisms using a constraint-based metabolic model. We found that acetate, lactate, ethanol and pyruvate are likely the major metabolites involved in carbon transfer, with pyruvate being the most inhibitory for the growth of *C. cellulolyticum*. Model-based analyses of *R. palustris* examined the effect of pyruvate and suggested possible ways to improve pyruvate consumption.

LLNL-ABS-469072. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

# 84

## Altered Sequences and Modifications to Proteins from Organisms of Interest to the Environment and Energy Production

**Rachel R. Ogorzalek Loo**[1,2*] (rloo@mednet.ucla.edu), Hong Nguyen,[1,3] Deborah Francoleon,[1,3] Housna Mouttaki,[5] Joseph A. Loo,[1,3] Lars Rohlin,[1,4] **Robert Gunsalus**,[1,4] **Sabeeha Merchant**,[1,3] and Michael McInerney[5]

[1]UCLA-DOE Institute for Genomics and Proteomics; [2]Department of Biological Chemistry, [3]Department of Chemistry and Biochemistry, [4]Department of Microbiology, Immunology, and Molecular Genetics, University of California-Los Angeles; and [5]Department of Botany and Microbiology, University of Oklahoma

http://www.doe-mbi.ucla.edu/overview

**Project Goals:**

- **Define *Chlamydomonas reinhardtii* proteomics**
- **Characterize proteome of *Syntrophus aciditrophicus***
- **Elucidate surface glycoproteins of methanogens and link to morphology**

**Introduction**

Living organisms transcribe and translate genetic information many ways and often embellish the products co- and post-translationally. The average error rate for translating proteins in normally growing mammalian and bacterial cells, $3x10^{-4}$ misreads per codon, was determined by elegant radiolabeling experiments performed 2-3 decades ago. Related experiments established that the error rate increased with exposure to high $Mg^{2+}$ concentrations, low temperature, high pH, or various organic solvents. Starvation for certain amino acids in animal and bacterial cells elevates the level of translation errors to one readily visualized as a trail of spots on 2-D gels, an effect referred to as "stuttering". An associated phenomenon, recoding, describes the local reprogramming of mRNA translation to discard standard translational rules and decode non-canonically. The products of this stimulated process, proteins incorporating shifts in reading frame, bypassed nucleotide segments, and/or unexpected amino acids have been observed in all 3 kingdoms and are known to contribute importantly to posttranscriptional regulation of gene expression. Environmental organisms are often challenged severely and likely to respond in these unusual ways, but the resulting proteomic changes are invisible to many experimental strategies. Proteomic methods can be tailored to reveal and characterize these unanticipated modifications and sequence alterations.

**Methods**

*Syntrophus aciditrophicus*, a gram-negative anaerobe surviving at the limit of thermodynamic feasibility (McInerney *et al.*, *PNAS* 2007, 104, 7600-7605), was cultivated anaerobically on crotonate or benzoate. Alternatively it was co-cultured with *Methanospirillum hungatei* on crotonate, benzoate, or cyclohexane carbonate. Proteins were separated by 2D gel

electrophoresis; protein-containing spots were trypsin-digested and analyzed by LC-MS/MS.

Streptavidin affinity chromatography, SDS-PAGE and near-western blotting were used to visualize and capture biotin-tagged surface-exposed proteins, which were trypsin-digested, and analyzed by LC-MS/MS. Lectin blotting suggested sugars present in glycosylated bands. Concanavalin A lectin affinity chromatography was used to enrich the S-layer protein. ConA eluates were digested with trypsin, GluC and AspN individually and double-digested with immobilized Pronase or Proteinase-K. Glycopeptides were enriched by doubled ConA affinity chromatography, ERLIC (HILIC/WAX) fractionation, C12 fractionation, and nonspecific protease treatment. Glycopeptides were measured by LC-MS/MS (QTOF) and ESI-MS/MS (ECD, IRMPD or CAD with an LTQ-FT Ultra). Normal phase and epifluorescence microscopy were used to visualize cell surface glycosylation.

**Results**
In the proteome of Syntrophus aciditrophicus we found many examples of translational frameshifting or recoding. De novo sequencing MS/MS spectra unassigned by automated protein and 6-frame translated DNA searches revealed multiple instances where >1 sequence appeared to arise from a single mRNA transcript; e.g., 6 unpredicted variants of the peptide TATPDQAQEAHAFIR in phosphoglycerate mutase. For the predicted peptide R.AQNLHDAK, the peptides AQNLHDAK, LGAQNLHDAK, and pyroQSAQNLHDAK were observed. All together, 14 peptides did not match the phosphoglycerate mutase sequence in any one reading frame.

Standard searching of LC-MS/MS data from a digested spot containing branched chain aminotransferase assigned 27/67 spectra, but after completing our analysis, 61/67 spectra were assigned, revealing 14 peptides unmatched in any one reading frame.

*Methanosarcina mazei* and *Methanosarcina acetivorans* surface-exposed proteins and modifications were identified, employing methods specially developed for compatibility with their fragile cells. *Methanosarcina* S-layer proteins MM1976 and MA0829, do not correspond to those predicted by bioinformatics. Unusual glycosylation patterns were also recovered.

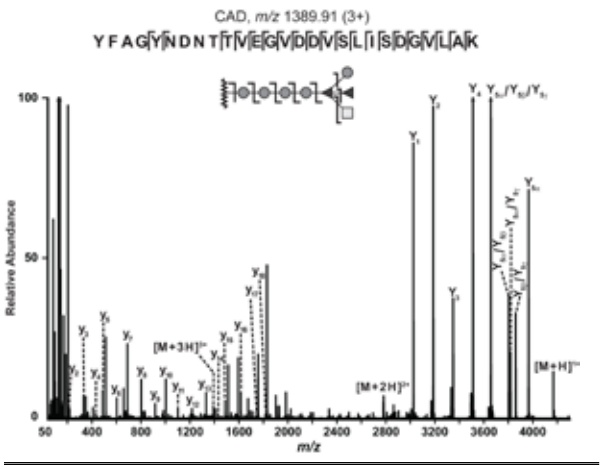



Figure 1. Mass spectrometry analysis of a glycopeptide in *Methanosarcina mazei* MM1364. Glycan symbol code: gray circle, Hexose; black triangle, Deoxyhexose; gray square, N-acetyl Hexosamine; wiggly line, peptide.

Student Oral Presentation–Monday

# 85

## Novel Hydrogen Production Systems Operative at Thermodynamic Extremes

**Michael J. McInerney**,[1] **Ralph S. Tanner**,[1] Jessica R. Sieber,[1] Kimberley L. Thomas,[1]* Matthew W. Maune,[1] Gregory Hurst,[2] Lars Rohlin,[3] and **Robert P. Gunsalus**[3]* (robg@microbio.ucla.edu)

[1]University of Oklahoma; [2]Oak Ridge National Laboratory; and [3]University of California, Los Angeles

http://www.mimg.ucla.edu/faculty/gunsalus/

**Project Goals: The goals of this collaborative project are to develop and utilize new research strategies to address Genomic Science program needs in the area of biohydrogen production. This includes the delineation of the molecular machinery involved in hydrogen production from thermodynamically difficult substrates, as well as the characterization of new microbial model systems that generate high $H_2$ concentrations approaching 17% of the gas phase.**

To identify the strategies used for hydrogen production in one model butyrate-degrading organism, we are performing genomic, proteomic, and transcript analysis on *Syntrophomonas wolfei*. This microbe is representative of an important but poorly understood class of hydrogen-producing organisms that are capable of syntrophic fatty and aromatic acid metabolism when co-cultured with suitable microbial partner(s). $H_2$ production from fatty and aromatic acids requires energy input. The *S. wolfei* genome possesses genes for three cytoplasmic and two externally located formate dehydrogenases plus two cytoplasmic and one externally located hydrogenase. By implication, either hydrogen or formate could be produced by *S. wolfei* during syntrophic growth. Interestingly, three cytoplasmic formate

dehydrogenases and one cytoplasmic hydrogenase appear to be NADH-linked because the respective gene clusters contain NADH:quinone oxidoreductases E and F subunit genes. The gene arrangement suggests that *S. wolfei*, like anaerobes known to produce high molar ratios of hydrogen from glucose, may produce $H_2$ and/or formate from NADH by an electron bifurcation mechanism. The proteome of *S. wolfei* grown syntrophically and axenically was analyzed to determine proteins involved in key redox reactions. Whole cell-derived peptide mixtures were analyzed with two-dimensional liquid chromatography/tandem mass spectrometry (2D LC-MS-MS) via the MudPIT approach. Proteomic analysis identified ~1090 polypeptides in all three growth conditions. Thirty-four polypeptides, including a zinc-dependent dehydrogenase with a GroES domain and a putative membrane-bound hydrogenase, were detected only in syntrophically grown *S. wolfei*, e.g., with *Methanosprillum hungatei* JF1 on butyrate or crotonate. Multiple systems for interspecies electron transfer and reverse electron transport including a confurcating hydrogenase and a novel FeS oxidoreductase thought to serve as an ETF:quinone oxidoreductase were detected. Multiple beta-oxidation enzymes and several electron transfer flavoproteins were detected in each growth condition. The number of proteins involved in coenzyme and amino acid transport and metabolism increased under syntrophic conditions while the number of proteins involved in lipid transport and metabolism increased during crotonate axenic growth. *S. wolfei*, expressed multiple enzyme systems for fatty acid metabolism, interspecies electron transfer, and energy conservation under all growth conditions. Only a relatively few proteins (34) were unique to the syntrophic lifestyle.

In a companion project, we are characterizing the genetic, biochemical, and physiological properties of a newly isolated anaerobic bacterium called *Anaerobaculum hydrogenoformans* strain OS1 that generates $H_2$ at concentrations up to 17% with glucose. It is a member of the proposed phylum 'Synergistetes', a poorly studied microbial group composed of many environmental strains for which little is understood at the ecological, biochemical, and physiological levels. *A. hydrogenoformans* has fermentation yields > 3 $H_2$/glucose, and grows with and produces high levels of $H_2$ from several other carbohydrates, at least nine different amino acids, and complex substrates like Casamino acids, tryptone and yeast extract. Strain OS1 also generates $H_2$ from undefined substrates like raw sewage and other waste streams that are potential commercial substrates for $H_2$ production. Labeling, metabolomic, and enzymatic analyses showed that OS1 degrades glucose by a modified Embden-Myerhof-Parnas (EMP) where a ferredoxin-dependent, glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR) is used instead of a $NAD^+$-linked glyceraldehyde-3-phosphate dehydrogenase. Pyruvate formed by this pathway is then metabolized by pyruvate:ferredoxin oxidoreductase. These results show that OS1 uses a ferredoxin-dependent metabolism to generate high levels of $H_2$. The OS1 genome is approximately 2.38 MB in size with 2,557 orfs, and has a GC content of 46.5%. The current assembly consists of 151 contigs with about 2.2 MB contained in the top three contigs. Machine annotation and manual curation are currently in progress to support a metabolic reconstruction of the cellular metabolism leading to hydrogen formation in strain OS1 when grown on hexoses and pentoses. Since strain OS1can also grow syntrophically in the presence of a $H_2$-consuming methanogen, it suggests the ability for a more complex alternative lifestyle.

# 86

## Biohydrogenesis in the Thermotogales

Andrew Frock[1]* (adfrock@ncsu.edu), Amanda Petrus[2]* (akpetrus@gmail.com), Derrick White[3]* (derrick.white@yahoo.com), Raghuveer Singh[3]* (raghuveer28m@yahoo.com), Andrew Loder,[1] **Paul Blum,[3] Kenneth Noll,[2]** and **Robert Kelly[1]**

[1]Dept. of Chem. and Biomolecular Engineering, North Carolina State University, Raleigh; [2]Department of Molecular and Cell Biology, University of Connecticut, Storrs; and [3]Beadle Center for Genetics, University of Nebraska, Lincoln
All contributed equally to this work.

**Project Goals:**

1. **Examine the regulation of substrate catabolic proteins and pathways as this relates to carbon partitioning, disposition of reducing power, and $H_2$ generation in *Thermotoga maritima* (Tma).**
2. **Dissect catabolic and regulatory pathways using genetic approaches based on past success with other hyperthermophiles.**
3. **Thermotogales biodiversity arises from adaptive specialization that expands on a conserved minimal genome; physiological characterization of selected novel traits will be done to expand understanding of biohydrogenesis.**

The bacterial order Thermotogales consists of obligate fermentative anaerobes that grow optimally in the range of 65-80°C. These bacteria are characterized by their unique "toga"-like outer membrane, which plays a role in breaking down a wide range of complex polysaccharides. Thermotogales are also capable of producing hydrogen with a high yield approximately twice that of mesophilic bacteria (3-4 mol $H_2$/mol glucose compared to 1-2). Study of all aspects of Thermotogales physiology would be furthered by the development of a genetic system. The objectives of this project are stated above under "Project Goals."

To investigate these objectives, genetic, biochemical and functional genomic methods were used and revealed considerable new information about this deep understudied phylum of hydrogen evolving organisms.

A genetic system for Tma has been developed and provided fundamental information about recombination, competence and the minimal promoter for transcription. The system includes three genetic markers, methods for transformation, cell line genotyping, and cell banking. Initial studies using

the genetic system focused on nonessential genetic targets in anticipation of studies focusing on metabolic genes. Three types of Tma recombinants have been constructed: gene replacements exchanging wild type for mutant (1) or mutant for wild type genetic alleles (2), and cell lines with chromosomally integrated disruption constructs (3). In each case, PCR and DNA sequencing was used to verify genotypes of clonal cell lines. (1) Gene replacement involving repair of a 128 nt pyrE deletion rescued uracil auxotrophy and provided the first estimate of regions (500 bp) required for homologous recombination in Tma. (2) Gene replacement of the DNA gyrase subunit encoded by gyrB, introduced a synthetic gyrB allele arising from a G to A transition mutation at nt 401 conferring resistance to novobiocin. Incorporation of flanking synonymous third position codon changes (nt 399, 405) was used to discriminate against spontaneous drug resistance mutations and provided the first estimate of recombination rates in Tma (102/ug DNA). (3) Integration of an arabinose isomerase (araA) disruption cassette was accomplished by single crossover. This recombinant was produced using a groES promoter fused to a thermostable kanamycin marker and provided the first demonstration of a minimal sequence for transcription initiation in Tma. Currently, the genetic system is being used to target genes involved in carbon catabolism (xylose isomerase), organic acid excretion (ackA, ldh), hydrogen formation (hydA), hydrogen consumption (hydA, hydB, hydC; bifurcating hydrogenase), and several non-metabolic targets.

The most unique and visually striking aspect possessed by the Thermotogales is the outer toga structure of the cell envelope. The toga presents an interesting subject on the merit of its structural novelty within the prokaryote domain. This structure associates closely with the cell membrane and by ballooning at the poles creates a notable periplasmic space. It serves as the interface at which these organisms interact with their environment. It is also proposed to act as an extracellular matrix for the organization of enzymes and proteins, allowing the utilization of insoluble carbon sources. Therefore understanding the composition of the toga will provide essential information regarding carbohydrate utilization and the role of this outer envelope. Studies on the identification of proteins present in the toga are currently being run on maltose-grown cells. Future studies will include cells grown a variety of carbon sources to gauge their effect on protein speciation in the toga.

To investigate the functionality and composition of the toga, samples were isolated by several techniques for proteomic analysis. In our first technique, outer envelope fractions were collected from lysed cell samples by sucrose gradient fractionation. Toga sheath material collected from fractions of higher sucrose densities yielded a sample with at least fifteen protein bands visualized by SDS-PAGE analysis. This sample contains bands within the molecular weight range of the two verified structural proteins present only in the toga, Ompα and Ompβ. These fractions are being sequenced by mass spectrometry to identify toga proteins. This sample was also further purified to obtain the band matching the weight of the Ompβ trimer and is currently being sequenced to identify the gene encoding Ompβ, which was not identified in the published genome annotation.

Outer envelope extraction from intact cells was carried out in three separate methods; (1) "trypsin shaving" of peptides from proteins with portions exposed on the cell surface, (2) biotinylation of exposed cell surface proteins using sulfo-NHS-S-S-biotin followed by avidin chromatography purification and (3) an organic extraction utilizing a chloroform/methanol solution that extracts surface proteins while leaving the cell membrane intact. This approach ensures that outer envelope proteins are being collected without contamination by cytoplasmic or cell membrane proteins.

Genome sequences are available for six members of the Thermotoga genus, and although some of these species are very closely related, key differences can impact carbohydrate utilization and hydrogen production. For example, Thermotoga species generally use ATP-binding cassette (ABC) transporters for sugar transport, but a single phosphotransferase system (PTS) transporter is found in *T. naphthophila* and *T.* sp. RQ2. During growth on a mixture of monosaccharides, *T.* sp. RQ2 showed a preference for fructose (the putative PTS substrate) that was not observed for *T. neapolitana*, *T. petrophila*, or *Tma*. These highly similar species offer a unique opportunity to directly investigate the role that the transport mechanism plays in energy conservation and hydrogen production. Analysis of gene transcription during growth on glucose and a mix of complex polysaccharides for four Thermotoga species (*Tma*, *T. neapolitana*, *T. petrophila*, and *T.* sp. RQ2) showed that although responses to these growth subtrates were mostly conserved from one species to another, *T. petrophila* exhibited a unique glucose-dependent induction of xylan/xyloside utilization genes.

Interspecies interactions and population dynamics during mixed culture are also being investigated. A real-time PCR assay has been developed to quantify the relative amounts of the same four species present in a mixed culture. Under the conditions initially tested, *T.* sp. RQ2 dominates the culture after a relatively short period of time. This observation may be attributable to a faster maximum growth rate, but interspecies interactions have not been ruled out as a factor. Mixed culture experiments up to this point have been performed in batch culture, so it remains to be seen how the population dynamics are affected by a controlled growth rate in a chemostat culture.

# 87

## Pathways and Regulatory Network of Hydrogen Production from Cellulose by *Clostridium thermocellum*

Jonathan Millen,[1] Ji Hye Jo,[2] Shiv Thammannagowda,[2] Yuan Zhang,[1] Ryan Saverins,[1] Pin-Ching Maness,[2] Mary Lipton,[3] and **J.H. David Wu**[1]* (davidwu@che.rochester.edu)

[1]University of Rochester, N.Y.; [2]National Renewable Energy Laboratory, Golden, Colo.; and [3]Pacific Northwest National Laboratory, Richland, Wash.

**Project Goals: The overall objective of this research is to understand $H_2$ metabolic pathways in *Clostridium thermocellum* and the underlying regulatory network at the molecular and systems levels. We hypothesize that: multiple hydrogenases work concertedly to contribute to hydrogen evolution; transcription factors control hydrogenase expression at the transcription level; and hydrogen metabolism is coupled to cellulolysis and fermentative pathways in this bacterium. We are determining the hydrogenase expression and metabolic network nodes on cells subjected to different culture conditions and metabolic pathway inhibitors, to probe differential expression of the various hydrogenases and their interrelationship with other cellular metabolic pathways. We are also mapping connections in the transcription factor network controlling linked metabolic pathways. Finally, we will purify FeFe-hydrogenases from its native producer and heterologously expressed *E. coli* to determine their subunit compositions.**

*Clostridium thermocellum*, a thermophilic, ethanogenic, and cellulolytic anaerobe, produces a complex extracellular cellulolytic organelle called the cellulosome. The cellulosome contains various depolymerizing enzymes that are arrayed on a protein scaffold and effectively degrades complex cellulosic substrates. During cellulose fermentation, the bacterium evolves hydrogen at a high rate. Analysis of its genome sequence reveals the existence of at least three putative hydrogenases (CtHydA1, CtHydA2 and CtHydA3) central to hydrogen metabolism. Furthermore, at least 20 genes are potentially related to hydrogen metabolism. The bacterium is thus remarkably versatile in employing various enzymes, some of which are potentially novel, for hydrogen metabolism. The versatility indicates the significance of this biological process. Yet little is known concerning the pathway for hydrogen production and the underlying regulatory mechanism/network that control these hydrogenase and the related genes as well as cellulolytic process and other metabolic pathways in the organism.

To probe the physiological functions of these three hydrogenases, we performed protein immunoblots with *C. thermocellum* 27405 (acquired from ATCC) cultured in cellobiose medium. We observed the expression of CtHydA1 and CtHydA2, but not CtHydA3. Sequencing the CtHydA3-encoding gene uncovered a mutation converting an amino acid residue to a stop codon and yielding a truncated protein. This mutation is confirmed by the RT-PCR data. Moreover, a second mutation was found in its upstream gene, Ct_3004 encoding an [FeS]- protein, the putative redox partner of CtHydA3. This mutation was also confirmed by RT-PCR. The mutations yield an inactive chimeric protein fusing Ct_3004 [FeS]-protein with the N-terminus of the CtHydA3 (mutant). However, none of these mutations were detected in another stock culture of the 27405 strain. At similar stages of growth, we consistently detected more hydrogen production in WT than mutant, assayed both *in vivo* (10% more) and *in vitro* (25% more), the latter mediated by reduced methyl viologen. The results suggest that CtHydA3 contributes to hydrogen production *in vivo*. At similar stages of growth, the mutant produced slightly more ethanol than WT, presumably by redirecting the cellular flux toward the reduced end product in lieu of hydrogen.

To identify transcription factors controlling hydrogen metabolic pathways, we developed an affinity purification method by immobilizing promoter DNA sequences to a solid support. DNA-binding proteins from the *C. thermocellum* cell lysate, eluted from the affinity columns, were identified by the MALDI-TOF or LC-MS-MS techniques. Several transcription factor candidates were identified. One of them binds to the promoter region of CtHydA1. The binding was confirmed by EMSA (electrophoretic mobility shift assay). Furthermore, the EMSA results indicate that binding of the protein to the promoter sequence is inhibited by NADH. Additional EMSA work using various oligonucleotides corresponding the promoter region revealed that the binding site consists of a 22 bp sequence containing a palindrome. Searching the entire *C. thermocellum* genome for the palindromic binding site sequences, we identified many potential binding sites in the intergenic sequences. EMSA analysis of these potential binding sites show that at least 5 of these promoter regions are indeed recognized by the regulatory protein. These results indicate that CtHydA1 expression is regulated by a transcription circuit that responds to the cellular redox status.

The studies will provide important insights into the pathway and regulatory mechanism/network controlling hydrogen metabolism and cellulolysis as well as other pertinent metabolic pathways in this very intriguing cellulolytic and thermophilic bacterium, which catalyzes the rate-limiting cellulose-degradation reaction in a single-step process of biomass conversion (or CBP, Consolidated Bioprocessing). Detailed understanding of the pathway and regulatory mechanism/network will ultimately provide rationales for engineering, alternating, or deregulating the organism for biomass conversion to liquid and hydrogen fuels.

# 88

## A Systems Approach to Uncovering the Metabolic and Regulatory Networks Surrounding Hydrogen Production in *Pyrococcus furiosus*

Michael P. Thorgersen* (mthorger@uga.edu), Gerrit J. Schut,* Gina L. Lipscomb, Jennifer R. Huddleston, Yixuan Zhu, Karen Stirrett, William J. Nixon, Janet Westpheling, Michael W.W. Adams, and **Robert A. Scott**

PI: Robert A. Scott (rscott@uga.edu)
Co-PIs: Michael W. W. Adams, and Janet Westpheling
Presenters: Michael P. Thorgersen (mthorger@uga.edu), and Gerrit J. Schut (gerti@uga.edu)

Department of Biochemistry and Molecular Biology, University of Georgia, Athens

http://adams.bmb.uga.edu/

**Project Goals: The goal of SAPHyRe (Systems Approach to Probing Hydrogen Regulation) is to characterize the regulatory network surrounding hydrogen production pathways in *Pyrococcus furiosus* as it relates to carbon and nitrogen sources, metal availability, and oxidative stress. The outcome of this project will serve two purposes: 1) it will bring us one step closer to utilizing *P. furiosus* in development of alternative energy sources and 2) it will serve as a model methodology for investigating the regulatory pathways of hydrogen production in other organisms.**

The efficient production of hydrogen ($H_2$) by the hyperthermophilic archaeon *Pyrococcus furiosus* is controlled by a complex regulatory and metabolic network. Hydrogen production can therefore be affected by environmental changes such as different carbon sources, the presence of electron acceptors such as sulfur, stress from reactive oxygen species, and decreased metal availability. The goal of our research is to characterize this network and identify key control points for $H_2$ production. We developed a genetic system for *P. furiosus* and constructed several mutant strains with gene deletions predicted to have an impact on hydrogen production either directly or indirectly. *P. furiosus* contains one membrane bound ferredoxin-dependent hydrogenase and two cytosolic NADPH-dependent hydrogenases. Studies of hydrogenase deletion mutants show that the membrane bound hydrogenase functions as the major $H_2$ producing enzyme, while the cytosolic hydrogenases only play minor roles in hydrogen metabolism. The expression of the hydrogenase genes is greatly decreased in the presence of $S^0$, and previous *in vitro* studies identified the redox active transcriptional regulator SurR to be involved in this response. *In vivo* studies confirmed the key role for SurR in activation of the hydrogenase operons in the absence of $S^0$ and in mediating the primary response to $S^0$ that results in a metabolic shift from production of $H_2$ to $H_2S$. Also, a connection between dioxygen ($O_2$) stress and decreased $H_2$ production in *P. furiosus* was found. *P. furiosus* redirects electron flow away from $H_2$ production to deal with oxidative stress, and we identified an enzyme (FdpA) that appears to scavenge $O_2$ *in vivo*. Using genetic technologies in *P. furiosus* is leading to a better understanding of the regulatory pathways influencing $H_2$ production and will allow fine-tuned control at the environmental and genetic levels for improved $H_2$ production.

# 89

## Development of *Cyanothece* as a New Model Organism for Photobiological Hydrogen Production

Anindita Bandyopadhyay,[1]* Jana Stöckel,[1] Lawrence Page,[1] Xueyang Feng,[2] Yinjie Tang,[2] Sujata Mishra,[3] Xiaohui Zhang,[3] Louis A. Sherman,[3] Uma K. Aryal,[4] Stephen J. Callister,[4] Janani I. Shutthanandan,[4] Carrie D. Nicora,[4] Ljiljana Pasa-Tolic,[5] Dave W. Koppenaal,[5] Richard D. Smith,[4] and **Himadri B. Pakrasi**[1,2]

[1]Depts. of Biology and [2]Energy, Environmental and Chemical Engineering, Washington University, St. Louis, Mo.; [3]Dept. of Biological Sciences, Purdue University, West Lafayette, Ind.; [4]Biological Sciences Division; and [5]Environmental Molecular Science Laboratory, Pacific Northwest National Laboratory, Richland, Wash.

http://sysbio.wustl.edu/Pakrasi/projects/hydrogen.php

**Project Goals: The aim of this project is to develop the cyanobacterium *Cyanothece* as a model organism for photobiological hydrogen production. Members of the genus *Cyanothece* are unicellular oxygenic prokaryotes with the ability to fix atmospheric nitrogen. Our long-term goal is to develop a deep understanding of the metabolism of these microbes as it pertains to $H_2$ evolution. Specifically, we are using genome sequencing, transcriptomics, proteomics, metabolomics, mutagenesis, biochemical analysis and physiological approaches, all of which are encased in a systems biology framework.**

*Hydrogen Production by Cyanothece*: We established a two stage system for photobiological $H_2$ production in the unicellular cyanobacterium *Cyanothece* ATCC 51142. Using this system, we demonstrated high rates of nitrogenase-mediated photobiological $H_2$ production under aerobic conditions (150 μmoles of $H_2$/mg Chl.h). $H_2$ production was dependent on the cellular glycogen reservoir and could be enhanced with an external carbon source like high $CO_2$ and glycerol. A batch culture of *Cyanothece* 51142 in the presence of glycerol could produce more than 900 ml of $H_2$ per liter of culture over a period of two days. These rates are several folds higher compared to the rates reported for any other wild type model $H_2$ producing strain. $H_2$ production in *Cyanothece* 51142 was largely driven by light even in the presence of glycerol, suggesting that the observed rates were not solely due to glycerol fermentation. However,

photosystem II did not seem to be involved in the process as observed from studies in the presence and absence of the photosystem II inhibitor DCMU (3-(3,4-dichlorophenyl)-1,1-dimethylurea). The strain is being analyzed further at the systems level to unravel the mechanism of $H_2$ production and to obtain insights into possible ways of further improving yields. $H_2$ production has also been optimized in six other sequenced *Cyanothece* strains. Analysis of nitrogenase activity and $H_2$ production in these strains revealed the ability of aerobic nitrogen fixation and $H_2$ production in five of the six *Cyanothece* strains. *Cyanothece* PCC 7425 exhibited $H_2$ production only under anaerobic incubation conditions. These results will be presented and discussed in detail.

*Cyanothece* ATCC 51472 and PCC 7822 were assayed for $H_2$ production as well as for the storage of carbon (in the form of glycogen, PHA and EPS) in detail. *Cyanothece* 7822, but not 51472, produces copious amounts of EPS and PHA and both strains make substantial levels of glycogen. Cultures were grown under ambient and elevated (0.4%) $CO_2$ concentrations, as well as in the presence of glycerol. *Cyanothece* 51472 grows well under high light (400 $\mu mol.m^{-2}s^{-1}$). Both the species had enhanced growth under increased $CO_2$ and under photomixotrophic conditions (with glycerol), and glycogen content was enhanced in 51472 under increased $CO_2$ concentration. Glycerol enhanced the nitrogenase activity and glycogen content in both the species. The level of $H_2$ evolved was positively correlated with higher levels of intracellular glycogen in both *Cyanothece* strains. In addition, our results demonstrated that *Cyanothece* sp 7822 was more efficient than 51472 in $H_2$ production both under photoautotrophic and photomixotrophic conditions. Although both strains produce high levels of $H_2$ (150-200 μmoles $H_2$ evolved/mg Chl/h), neither strain is as productive as *Cyanothece* 51142.

*Comparative Genomics*: Complete genome sequences of six *Cyanothece* strains (ATCC 51142, PCC 7424, PCC 7425, PCC 7822, PCC 8801, and PCC 8802) are currently available and one more (ATCC 51472) is in the process of completion at the DOE Joint Genome Institute. The sequencing revealed the presence of one linear chromosome in *Cyanothece* 51142 and 3 linear elements in *Cyanothece* 7822. This feature is unique to *Cyanothece* strains compared to other sequenced cyanobacteria and suggest the presence of distinctive metabolic traits in members of this group. A comparison of the genomes of the six *Cyanothece* strains revealed the presence of several pathways analogous to non-oxygenic microbes in these strains, an observation which complies with their ability to maintain a suboxic intracellular environment for a significant part of a diurnal cycle. These characteristics suggest that the group *Cyanothece* can be appealing as model organisms for studies pertaining to biohydrogen production.

*Cyanothece Genetics*: *Cyanothece* 7822 was successfully transformed using a single-stranded DNA technique and a *nifK* knockout mutant was generated. A stable line of this mutant is being further analyzed. In addition, we constructed a mutant *(DhupL,* a deletion of the gene encoding the large subunit of uptake hydrogenase) in *Cyanothece* 7822 by inserting a neomycin/kanamycin antibiotic resistance cassette in the *hupL* gene. The growth rate of the mutant strain was about half the rate of the wild type strain in BG11 medium with nitrate, and the *DhupL* strain grew very poorly in nitrate-free BG11. In contrast to uptake hydrogenase mutants in filamentous cyanobacterial strains we observed no $H_2$ production and no nitrogenase activity in *DhupL*. The *in vitro* uptake hydrogenase activity is zero in the mutant. The mutant cells are rounder and fatter in morphology and 1.7 times larger in volume relative to the wild type cells. The knockout of *hupL* caused a defect of nitrogenase in terms of conversion of $N_2$ to $NH_3$, as well as a defect in $H_2$ production. In the unicellular cell, HupL may play an important role in balancing *in vivo* gas metabolism.

*Metabolomic and Fluxomic Studies*: The previous $^{13}C$-assisted metabolism analyses have identified unique metabolic features (*i.e.* citramalate pathway and $CO_2$ fixation regulations) in *Cyanothece* 51142. Based on these discoveries, a constraint-based genome-scale flux balance model for *Cyanothece* 51142 metabolism is under development. The model can be used not only to analyze functional pathways under different cultivation conditions, but also for *in silico* genetic manipulations (*e.g.* knock-out and over-expression) and to predict metabolic behavior in mutant strains. To facilitate our study on other cyanobacterial species, our model development mainly focuses on a general platform for metabolic network reconstruction and flux balance analysis. Such user-friendly and website-based software is able to automatically reconstruct genome scale metabolic network for different cyanobacterial species based on the "KEGG" database. Then the customers can manually redefine functional pathways and flux boundaries. Users can also select their own objective functions (such as maximum biomass productions, minimal enzyme usage, etc). In addition, the software can perform dynamic flux analysis by integrating kinetics and fluxomics via static optimization approach. This software can be potentially used for comparing the physiologies of different cyanobacterial species and providing the guidelines for rational design of metabolic network for $CO_2$ fixation and $H_2$ production.

*Proteomic Studies*: The proteomes of six *Cyanothece* strains (ATCC 51142, PCC 7822, PCC 7424, PCC 7425, PCC 8801 and PCC 8802) were analyzed and compared to identify proteins common to all strains as well as unique proteins characteristic of an individual strain. The observed coverage of predicted proteins (based upon genome annotation) ranged from about 47% of the predicted genome in *Cyanothece* PCC 7822 to 67% in *Cyanothece* ATCC 51142. Details of these analyses will be presented and discussed.

Project number - DE-FC02-07ER64694

# 90

## Systems Biology of Hydrogen Regulation in *Methanococcus maripaludis*

Sung Ho Yoon,[1] Qiangwei Xia,[2] Erik L. Hendrickson,[2] June Burn,[4] **William B. Whitman**,[3] **Murray Hackett**,[2] **Nitin Baliga**,[1] and **John A. Leigh**[4]* (leighj@u.washington.edu)
**Co-PIs:** Nitin Baliga,[1] Murray Hackett,[2] and William B. Whitman[3]

[1]Institute for Systems Biology, Seattle, Wash.; [2]Department of Chemical Engineering, University of Washington, Seattle; [3]Department of Microbiology, University of Georgia, Athens; and [4]Department of Microbiology, University of Washington, Seattle

http://baliga.systemsbiology.net/enigma//content/hydrogen-regulation
http://faculty.washington.edu/leighj/
http://depts.washington.edu/mhlab/

**Project Goals: Generate a regulatory network model for a hydrogenotrophic methanogen. Determine the response of the model species *Methanococcus maripaludis* to changes in hydrogen, nitrogen, and phosphate conditions.**

**Background**
We are engaged in a long-term effort to understand regulatory networks in hydrogenotrophic methanogens, members of the Archaea whose energy metabolism specializes in the use of $H_2$ to reduce $CO_2$ to methane. Our studies focus on *Methanococcus maripaludis*, a model species with good laboratory growth characteristics, facile genetic tools, and a tractable genome of 1722 annotated ORFs. Much of our work to date has focused on the response that occurs when supplies of essential nutrients are decreased to growth-limiting levels. Thus, we have studied the responses to $H_2$ limitation, nitrogen limitation, phosphate limitation, and leucine limitation (using a leucine auxotroph) (1-3). A key aspect of our approach is the use of continuous culture for maintaining defined nutrient conditions (4).

**Transcriptome structure**
We have determined the transcriptome structure of *Methanococcus maripaludis* using a tiling array. We mapped transcription start and transcription termination sites for 1,025 transcription units. In many cases, we observed conditional activation of promoters inside operons and even inside coding sequences. We identified 29 antisense RNAs, 8 additional non-coding RNAs, 6 new protein-coding genes, and 5 putative overlapping protein-coding genes. A poster on the comparative evolution of transcriptome structures in four different Archaea will be presented separately (S. H. Yoon et. al.).

**Regulatory network inference**
We have generated quantitative transcriptome data over time courses during four transitions in chemotats: phosphate limitation to hydrogen limitation, nitrogen limitation to hydrogen limitation, high hydrogen to low hydrogen, and low hydrogen to high hydrogen. Regulatory network inference based on this data is under way. In addition, analysis of steady-state conditions has identified numerous previously unannotated transcripts that are regulated by hydrogen limitation or generally by nutrient limitation.

**Proteomics**
As a complement to our previous proteomic analysis (1), we have measured protein abundances in response to nutrient conditions, using the same strain and same samples that were analyzed at the transcriptome level. A comparison of regulatory trends at the transcriptome and proteome levels is under way. An analysis of local false discovery rates in the proteomic data will be presented in a separate poster (M. Hackett et al.). Chemostat samples are being generated for differential analysis of post-translational modifications under all three nutrient limitation conditions, using both computational and chemical approaches.

**Harvesting, quenching, and metabolite measurements**
Rapid and non-destructive methods for harvesting cell samples are important for the accurate measurement of metabolite levels and potentially for the detection of conditional post-translational modifications. We have implemented a rapid-cooling method using cold methanol to prevent metabolic perturbations during harvesting. One metabolite we are focusing on is the coenzyme $F_{420}$, an important electron carrier in methanogenesis. We have developed a rapid, anaerobic, acetone extraction method, which we follow with a fluorometric assay to determine levels of oxidized and reduced $F_{420}$ in cells.

This research was supported by the Office of Science (BER), U.S. Department of Energy, Award No. DE-FG02-08ER64685.

**References**

1. Xia Q, *et al.* (2009) Quantitative proteomics of nutrient limitation in the hydrogenotrophic methanogen *Methanococcus maripaludis*. *BMC Microbiol* 9:149.
2. Hendrickson EL, *et al.* (2008) Global responses of *Methanococcus maripaludis* to specific nutrient limitations and growth rate. *J Bacteriol* 190:2198-2205.
3. Hendrickson EL, Haydock AK, Moore BC, Whitman WB, and Leigh JA (2007) Functionally distinct genes regulated by hydrogen limitation and growth rate in methanogenic Archaea. *Proc Natl Acad Sci U S A* 104:8930-8934.
4. Haydock AK, Porat I, Whitman WB, and Leigh JA (2004) Continuous culture of *Methanococcus maripaludis* under defined nutrient conditions. *FEMS Microbiol Lett* 238:85-91.

# 91

## Quantitative Local False Discovery Rates, Deep Sampling and Protein Abundance Change for *Methanococcus maripaludis*

Qiangwei Xia,[1,2†] David A. C. Beck,[3] Tiansong Wang,[2] E.L. Hendrickson,[1] **John A. Leigh**[2] (leighj@u.washington.edu), and **Murray Hackett**[1*] (mhackett@u.washington.edu)

[1]Department of Chemical Engineering, [2]Department of Microbiology, and [3]eScience Institute, University of Washington, Seattle
†Current address: Regeneron Pharmaceuticals, Inc., Tarrytown, N.Y.

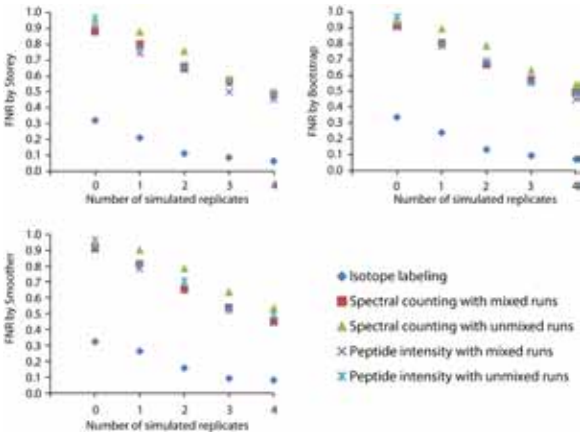
**PI:** John A. Leigh, University of Washington,
**Co-PIs:** Murray Hackett, University of Washington; Nitin Baliga, Institute for Systems Biology; William B. Whitman, University of Georgia

**Project Goals: Proteomics is a rapidly changing field, and in addition to our primary goals that involve probing questions of gene regulation as they apply to nutrient limitation, methanogenesis and the potential for hydrogen production, we are also interested in improving our analytical procedures so as to get the maximum return on the resources invested. This is especially important with respect to whole cell proteomes where the need to be as comprehensive as possible must be balanced by considerations of cost and available instrument time. To that end we have been pursuing three interrelated goals.**

1. **To find the best balance between statistical power to detect protein abundance change and the need to decrease the time required for a complete proteome analysis for *Methanococcus maripaludis* and related organisms.**
2. **To establish just how much sampling is required for spectral counting to become as efficient for generating protein abundance ratios relative to traditional metabolic stable isotope labeling. The answer to this question is heavily dependent on the specific mass spectrometry instrumentation and methods used in the investigation, both of which have experienced several upgrades and improvements since we last addressed this question in the peer reviewed literature in 2006.**
3. **To establish the most efficient data reduction and transformation procedures for the use of local false discovery rates (lfdr) with spectral counting data. Lfdr is one approach among several to multiple hypothesis testing, based on the more general concept of false discovery rate.**

***Methanococcus maripaludis* Proteomics**
Protein abundance ratios were measured using five different approaches for the Archaeon *Methanococcus maripaludis*. Multidimensional capillary HPLC coupled with tandem mass spectrometry was used for analysis of heavy ($^{15}N$) and natural abundance ($^{14}N$) tryptic digests of *M. maripaludis* grown in chemostats. Here we report our comparison of abundance ratios based on heavy and light proteomes mixed prior to mass spectrometry; spectral counting of heavy and light proteomes mixed; spectral counting of heavy and light proteomes analyzed separately; summed signal intensities for mixed heavy and light proteomes; and summed signal intensities for heavy and light proteomes analyzed separately. Protein identifications were saturated and proteome penetration maximized at ~91% of the predicted protein-encoding open reading frames. False discovery rates (fdr) and local false discovery rates (lfdr) were compared as complementary approaches to multiple hypothesis testing for quantitative significance.

**Results for calculations using fdr and lfdr theory**
At the limit of deep sampling frequency based measurements are competitive but lack the same power to detect relative abundance change when compared with metabolic stable isotope labeling, based on a conventional *t*-test for two sample comparisons assuming unequal variances (see figure below, lower FNR (false negative rate) means higher power, first entries on the left are real data, followed by Monte Carlo simulations of additional replicates. In this example a two-fold change cutoff was used). This result was driven primarily by the large numbers of heavy-light pairs used in the *t*-test calculations, sometimes several thousand for a single calculation for a given ORF observed in a two-state comparison. In general the lfdr calculations were difficult to implement with the *M. maripaludis* nutrient limitation data due to the relatively poor fit of the p-values and (or) z-scores used as inputs and the strict requirements of existing lfdr algorithms with respect discontinuities in the input distributions.




This research was supported by the Office of Science (BER), U.S. Department of Energy, Award No. DE-FG02-08ER64685

# 92

## Development of Biologically-Based Assays to Study Rate-Limiting Factors in Algal Hydrogen Photoproduction

**Maria L. Ghirardi**[1]* (maria.ghirardi@nrel.gov), **Alexandra Dubini**[1] (alexandra.dubini@nrel.gov), and **Matthew Wecker**[2] (matt.wecker@nrel.gov)

[1]Biosciences Center, National Renewable Energy Laboratory, Golden, Colo.; and [2]Genebiologics, LLC, Boulder, Colo.

**Project Goals: To develop techniques that will provide a deeper understanding of algal $H_2$ metabolism and accelerate the development of future photobiological $H_2$-producing catalysts and organisms.**

Photobiological $H_2$ production from water is a clean, non-polluting and renewable technology. The efficiency of light conversion into $H_2$ by biological organisms is theoretically high (about 10%). However, the system is currently limited by biochemical and engineering constraints, including the extreme $O_2$ sensitivity of the biological hydrogenases and the low availability of reductants to the hydrogenase due to the predominance of competing metabolic pathways. We address the $O_2$ sensitivity issue by developing an assay to screen large microbial populations for improved $H_2$-production properties with the eventual goal of evolving $O_2$ tolerance in [FeFe]-hydrogenases. The assay that we have created is based on the $H_2$ sensing system of the photosynthetic bacteria *Rhodobacter capsulatus* and it is tied to the production of a green fluorescence protein in response to $H_2$. The issue of competitive pathways utilizes the yeast two-hybrid assay to measure and deconvolute the interactions between different ferredoxin isoforms present in the green alga, *Chlamydomonas reinhardtii*, with their known electron acceptors. Identified interactions are then quantified by isothermal titration calorimetry.

We have demonstrated that the *Rhodobacter*-based $H_2$-sensing assay is remarkably sensitive to $H_2$, even at levels below the background level of $H_2$ in air, thus rivaling the sensitivity of the best known chemical sensors to $H_2$. We have shown that $H_2$-reporting bacteria, co-cultured with *Chlamydomonas reinhardtii* under anaerobic conditions fluoresce in direct relation to the amount of $H_2$ produced by the algae. This suggests a useful means of selecting *Chlamydomonas* $H_2$-production mutants with improved $H_2$ production activity. We will focus next on adapting the assay for operation under photobiological $H_2$-production conditions and use it to screen for $O_2$-tolerant [FeFe]-hydrogenases generated through directed-evolution techniques. The hydrogenases of *Clostridium acetobutylicum* and *C. reinhardtii* and hydrogenase assembly proteins from *Bacteroides thetaiotaomicron* have been introduced into a hydrogenase-negative *R. capsulatus* strain and are being tested for heterologous $H_2$ production.

To address the issue of competitive metabolic pathways with $H_2$ production, we have performed a yeast two-hybrid assay to screen for proteins interacting with the 6 ferredoxin (FDX) isoforms and the 2 hydrogenases (HYD) present in *Chlamydomonas reinhardtii* to characterize the pathways involved in shuffling electrons away from or directly to the hydrogenases. The results have yielded a list of genes encoding for different proteins that are going to be the primary targets for future metabolic engineering and used also to understand how the reductants are allocated to the HYDs. Each FDX and HYD has been shown to interact with specific binding partners. We have confirmed those interactions by library screening and pairwise interaction studies. We have now a metabolic network map available from those studies where the FDXs and the HYDs have been assigned a potential function/interaction in the newly discovered pathways. The major information that resulted from this study is showing that FDX1 is interacting with both HYDs, confirming its role in hydrogen photoproduction. The next step is to study those interactions in vitro and to define their affinity binding using the IsoThermal Calorimetry technique. Currently, HYDs and FDX1 are being purified and will be the first candidates to be tested for their binding constant.

**Publications:**

1. Matt S.A. Wecker, Jon Meuser, Matthew Posewitz and Maria L. Ghirardi "Design of a new biosensor for algal $H_2$ production based on the $H_2$-sensing system of *R. capsulatus*" J. Int. Hydrogen Energy, submitted.

# 93

## Phototrophic Metabolism of Organic Compounds Generates Excess Reducing Power That Can Be Redirected To Produce $H_2$ as a Biofuel

James B. McKinlay* (mckinla1@uw.edu), Yasuhiro Oda, Colin R. Lappala, and **Caroline S. Harwood**

Department of Microbiology, University of Washington, Seattle

https://depts.washington.edu/cshlab

Program title: Hydrogen production

**Project Goals: The goals of this project are (i) to use $^{13}$C-metabolic flux analysis and other approaches to identify metabolic factors that influence the phototrophic production of $H_2$ from organic compounds by *R. palustris* and (ii) to use the resulting information to guide the engineering of *R. palustris* for improved $H_2$ production characteristics.**

There is currently a pressing need for renewable fuels to negate the adverse impacts of burning fossil fuels. $H_2$ is a promising fuel, having about three-times the energy content of gasoline. Although most manufactured $H_2$ comes from fossil fuels, $H_2$ can also be produced biologically. *Rhodop-*

*seudomonas palustris* uses energy from sunlight and electrons from organic waste to produce $H_2$ via nitrogenase. In order to understand and improve this process we used $^{13}C$-substrates having various oxidation states to track and compare central metabolic fluxes in non-$H_2$ producing wild-type *R. palustris* and an $H_2$-producing mutant. The pathways by which substrates were oxidized generated excessive amounts of reducing power such that only 40-60% could be used for biosynthesis, depending on the growth substrate. Wild-type cells relied heavily on the $CO_2$-fixing Calvin cycle to oxidize the excess reduced electron carriers, using $CO_2$ produced from the organic substrates by other metabolic reactions. The $H_2$-producing mutant used a combination of $CO_2$ fixation and $H_2$ production to oxidize excess reduced electron carriers. The majority of electrons for $H_2$ production were diverted away from $CO_2$ fixation for all substrates. Microarray and qRT-PCR analyses indicated that this shift of electrons towards $H_2$ involved transcriptional control of Calvin cycle gene expression. These observations pointed to the Calvin cycle as a convenient single target to disrupt to force more electrons towards $H_2$ production. Blocking Calvin cycle flux by mutation in the $H_2$-producing strain resulted in higher $H_2$ yields for all substrates. The increase in $H_2$ yield was proportional to the Calvin cycle flux in the parent strain for most substrates. These results demonstrate how systems level approaches, such as $^{13}C$-metabolic flux analysis, can lead to effective strategies to improve product yield. Furthermore, our results underscore that the Calvin cycle and nitrogenase have important electron-accepting roles separate from their better known roles in biomass generation and ammonia production.

# 94

## Large Scale Functional Genomic Analysis of Hydrogen-Producing *Rhodopseudomonas palustris*

Yasuhiro Oda[1]* (yasuhiro@uw.edu), Christopher D. Armour,[2] Somsak Phattarasukol,[1] Colin R. Lappala,[1] John C. Castle,[3] Matthew C. Biery,[2] Christopher K. Raymond,[2] and **Caroline S. Harwood**[1]

[1]Department of Microbiology, University of Washington, Seattle; [2]NuGEN Technologies, Inc., San Carlos, Calif.; and [3]Institute for Translational Oncology and Immunology, Mainz, Germany

**Project Goals: To use a systems level approach to dissect metabolic and regulatory networks that are critical for nitrogenase-catalyzed hydrogen production by a phototrophic bacterium *Rhodopseudomonas palustris*.**

Hydrogen gas has good potential as an alternative fuel since it is clean burning and has a high energy content. Bacteria can produce hydrogen using hydrogenases or nitrogenases. The photosynthetic bacterium *Rhodopseudomonas palustris* produces copious amounts of hydrogen along with ammonia via the enzyme nitrogenase. Nitrogenase-catalyzed hydrogen production requires large amounts of ATP and electrons, which *R. palustris* can obtain from sunlight and biomass (such as agricultural and industrial waste), respectively, by fairly complex metabolic routes. Thus, hydrogen production involves the appropriate integration of dozens of metabolic reactions. Our long-term goal is to integrate genomic, transcriptomic, and phenotypic data from up to 100 *R. palustris* stains using Bayesian network analysis to identify all genes that are involved in hydrogen production. This would include genes that may change very little in expression and are therefore not easily recognized when just a few strains are analyzed by conventional transcriptomic techniques.

The emergence of next generation sequencing technology has opened up new opportunities to obtain large scale systems biology information. The success of high-throughput sequence-based transcriptome profiling depends on the availability of both efficient methods for the construction of high complexity cDNA libraries and computational data analysis tools. Here, we describe molecular and computational tools for high-throughput analysis of bacterial transcriptomes. Our cDNA library construction method utilizes computationally designed hexamers to selectively enrich non-rRNA transcripts during cDNA synthesis. This method requires only 500 ng of total RNA and enables the construction of barcoded cDNA libraries within 6h. When we applied this method to characterize transcriptome profile of *R. palustris* CGA009, on the order of 50% of the total sequencing reads were mapped to non-rRNA transcripts. We also developed a sequence analysis pipeline that provides global mapping statistics, gene-level expression profiles, and chromosomal expression maps as standard outputs. Using our tools, we characterized the transcriptomes of five *R. palustris* strains grown under the conditions of nitrogen-fixing (hydrogen-producing)-high light, nitrogen-fixing-low light, and ammonia-high light. Under these conditions, there were significant differences in growth rates, hydrogen production, and nitrogenase activities among strains tested. Genes that were previously identified as highly expressed under hydrogen-producing conditions by Affymetrix GeneChip were also expressed at higher levels in all strains, including the molybdenum-nitrogenase gene cluster, PII nitrogen regulatory genes, and ammonium transporter genes. Interestingly, levels of expression among these genes varied significantly from one strain to another, indicating that these expression changes may result in the observed phenotypic differences among the strains. We also identified a ferredoxin gene (which may donate electrons to nitrogenase) and several hypothetical genes that were highly expressed among the strains that had higher hydrogen production and nitrogenase activities.

These results demonstrate that our molecular and computational tools are a powerful way to characterize bacterial transcriptomes. We further demonstrated the integration of large scale functional genomic and phenotypic data derived from hydrogen-producing *R. palustris* strains. These studies have revealed novel insights into bacterial transcriptome dynamics and have led to the identification of several genes that may be important for hydrogen production.

Student Oral Presentation–Monday

# 95

## Xpression: A Complete Pipeline for Analyzing RNA-Seq Data

Somsak Phattarasukol[1]* (sphatt@u.washington.edu), Matthew Radey,[2] Mitchell Brittnacher,[2] and **Caroline S. Harwood**[1]

[1]Dept. of Microbiology and [2]Northwest Regional Center of Excellence, University of Washington, Seattle

**Project Goals: To use a systems level approach to dissect metabolic and regulatory networks critical for nitrogenase-catalyzed hydrogen production by the phototrophic bacterium *Rhodopseudomonas palustris*.**

RNA-seq is a revolutionary tool for transcriptome analyses. It has a number of advantages over traditional microarray-based technologies (e.g. annotation-independent detection of transcription, improved sensitivity and increased dynamic range), which collectively results in a much more precise measurement of gene expression levels. Despite these clear benefits, wide spread adoption of RNA-seq by researchers is impeded by the lack of an integrated software to analyze the data. Currently, millions of raw sequence reads are generated for each RNA-seq experiment, which makes it impossible to interpret these data without using computational tools.

We have developed Xpression, a complete pipeline for analyzing RNA-seq data generated from the Illumina platform. The workflow starts involves 1) filtering out low-quality sequence reads in raw sequence files, 2) aligning the sequence reads against a reference genome using the Burrows-Wheeler Alignment (BWA) tool, 3) computing statistics of sequence mapping, 4) reporting the number of sequence reads uniquely mapped to a particular region of the genome to indicate expression levels, and 5) generating plots to illustrate expression data (as normalized graphs of sequence read coverage across a genome) in platforms such as Artemis or Integrated Genome Browser.

Note that outputs from Xpression can be conveniently used in further downstream analyses. For example, users can additionally apply statistical softwares such as edgeR, DEseq or baySeq on expression data to identify differentially expressed genes, or open an alignment file in specialized visualization softwares such as Tablet. Also, Xpression can be configured to match certain requirements. For example, users may specify what barcodes are attached to biological samples, the types of library preparation protocol being used and whether strandness of sequencing reads is conserved.

# 96

## Reconstructing the Metabolic Network of *Rhodobacter sphaeroides 2.4.1*

Saheed Imam[1]* (srimam@wisc.edu), Safak L. Yilmaz,[2] Ugur Sohmen,[2] Timothy J. Donohue[3] (tdonohue@bact.wisc.edu), and **Daniel R. Noguera**[2] (noguera@engr.wisc.edu)

[1]Program in Cellular and Molecular Biology, [2]Department of Civil and Environmental Engineering, and [3]Department of Bacteriology, University of Wisconsin, Madison

**Project Goals: The goals of this project include: (i) creating a genome-scale metabolic reconstruction for *R. sphaeroides* 2.4.1; (ii) determination of its biomass composition to generate suitable biomass objective functions for modeling; (iii) validating the generated model using constraint based analysis and comparing predictions to experimental data; and (iv) employing the model in hypothesis generation.**

**Background:** *Rhodobacter sphaeroides* represents one of the best studied members of the group of purple non-sulfur photosynthetic bacteria and serves as an excellent model for the study of the complex and versatile metabolic capabilities of this group of organisms. The ability of *R. sphaeroides* to produce large amounts of hydrogen gas ($H_2$), polyhydroxybutyrate (PHB) and fatty acids photoheterotrophically, as well as its ability to utilize atmospheric carbon dioxide ($CO_2$) as a carbon source during photoautotrophic growth, make it an excellent candidate of potential use in a wide variety of biotechnological applications. To effectively harness its biotechnological potential, a thorough understanding of its metabolic capabilities is necessary. The goals of this project included: (i) creating a genome-scale metabolic reconstruction for *R. sphaeroides 2.4.1*; (ii) determination of its biomass composition to generate suitable biomass objective functions for modeling; (iii) validating the generated model using constraint based analysis and comparing predictions to experimental data; and (iv) employing the model in hypothesis generation.

**Methods:** Network reconstruction was conducted using all archived *R. sphaeroides 2.4.1* metabolic and genomic information from KEGG database combined with metaSHARK analysis. The model generated from the reconstruction was used in simulations employing constraint based approaches including flux balance analysis (FBA), flux variability analysis (FVA) and alternate optima analysis.

*R. sphaeroides* cells were grown photosynthetically in continuous culture and the biomass composition of cells determined using established laboratory techniques. Gas production was measured by respirometry, with $H_2$ and $CO_2$ contents determined via gas chromatography.

**Results:** The reconstructed metabolic network of *R. sphaeroides 2.4.1* consists of 796 metabolites, 856 transformation

reactions and 300 transport reactions. The reconstruction accounts for 1095 genes, covering about 25% of the *R. sphaeroides* open reading frames. The subsystem distribution of the reactions in the reconstruction is summarized in Fig 1. The empirically determined biomass composition of photosynthetically growing *R. sphaeroides* (Fig 2), was used in generating a biomass objective function for constraint based analysis. FBA simulations, which allowed for prediction of metabolic flux distributions during aerobic, photoheterotrophic and photoautotrophic growth, showed good qualitative agreement with experimental data (Table 1). Model predictions of growth rate, PHB synthesis, $H_2$ and $CO_2$ also showed good agreement with experimental data. Finally the model was also used in the identification of other potential pathways for carbon assimilation and cellular redox balancing, thus serving a source of testable hypotheses.

**Conclusions:** Our model of *R. sphaeroides* metabolic network serves as an important starting point in our goal of reconstructing and integrating the metabolic, regulatory and signaling networks of this complex microbe. The metabolic model shows good qualitative and quantitative agreement with experimental data and thus should provide an excellent framework for future metabolic modeling of *R. sphaeroides* and other related bacteria.

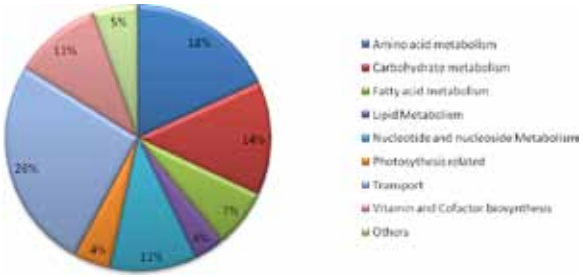



Figure 1. Subsystem distribution of reactions in reconstruction. Pie chart depicts the subsystem distribution of the model's reactions, with the percent contribution of each subsystem of reactions indicated in the corresponding section of the chart. It can be seen that amino acid, carbohydrate and nucleotide metabolism dominate the transformation reactions present in the model.

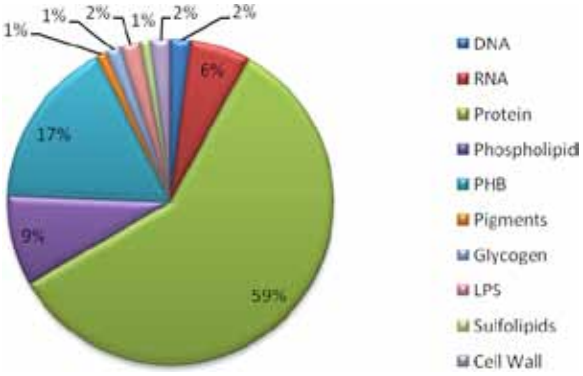



Figure 2. Biomass compostion of photosynthetically grown *R. sphaeroides*.

| | Light | Dark | | |
|---|---|---|---|---|
| **Carbon/nitrogen source** | | **$O_2$** | **DMSO** | **None** |
| Succinate + $NH_3$ | +/+ | +/- | +/- | -/- |
| Succinate + Glutamate | +/+ | +/- | +/- | -/- |
| Lactate + $NH_3$ | +/+ | +/- | +/- | -/- |
| Glutamate only | +/- | +/- | +/- | -/- |
| $CO_2$ + $H_2$ + $NH_3$ | +/- | -/- | -/- | -/- |
| $CO_2$ + $H_2$ + $N_2$ | +/- | -/- | -/- | -/- |

Table 1: Qualitative assessment of metabolic model. Table summarizes model predictions during aerobic and photosynthetic growth. +/+ Growth and $H_2$ production predicted; +/- Growth but no $H_2$ production predicted; -/- No growth

# 97

## Systems Level Approaches to Understanding and Manipulating Heterocyst Differentiation in *Nostoc punctiforme*: Sites of Hydrogenase and Nitrogenase Synthesis and Activity

Harry Christman,[1] Elsie Campbell,[1] Douglas Rissser,[1] Brett Phinney,[2] Wan-Ling Chiu,[3] and **John C. Meeks[1]*** (jcmeeks@ucdavis.edu)

[1]Department of Microbiology and [2]Genome Center, University of California, Davis; and [3]Department of Biology, Virginia Commonwealth University, Richmond

**Project Goals: Heterocyst-forming cyanobacteria, such as *Nostoc punctiforme*, are applicable for cost-effective photo-biohydrogen production, providing the frequency of heterocysts in the filaments can be increased by genetic manipulation and the metabolic end product, $H_2$, is uncoupled from growth. The goals of this project are to apply transcriptomic and proteomic analyses to free-living and symbiotically grown wild type and mutants strains to identify regulatory elements and pathways that can be manipulated.**

We hypothesize that if the frequency of heterocysts, sites of nitrogen fixation and hydrogen evolution, can be increased approximately 3-fold above the less than 10% normally found in filaments, then heterocyst-forming cyanobacteria would be applicable for cost effective photo-biohydrogen production. In *Nostoc punctiforme*, the heterocyst frequency of 8% in the free-living state is increased to 30-35% when in symbiotic association with terrestrial plants, such as the hornwort *Anthoceros* spp. and the angiosperm *Gunnera* spp. There is an increase in symbiotic nitrogen fixation that parallels the increased heterocyst frequency and about 85-90% of the metabolic end product, ammonium, is excreted to the plant partner. This is the kind of metabolic uncoupling we seek for biohydrogen production. We model heterocyst differentiation in the context of establishing the pattern of

heterocyst spacing, maintaining the pattern during growth on $N_2$ and symbiotic disruption of the spacing pattern. Establishment of the pattern following combined nitrogen deprivation is further modeled as a two stage process: biased initiation where a cluster of cells sense nitrogen starvation and initiate the differentiation process, followed by competitive resolution through interactions between HetR and PatS, where a single cell in the cluster completes the differentiation and maturation into a mature functional heterocyst. We do not know whether the same or similar regulatory networks are involved in establishment, maintenance and disruption of the pattern. We are applying transctriptomic and proteomic analyses with wild-type and mutant strains to identify the regulatory circuits of free-living heterocyst differentiation and how those circuits may have been co-opted during symbiotic growth.

**Genetics.** We have isolated an exogenously induced transposon mutant of *N. punctiforme* that displays a heterocyst frequency (> 30% of the total cells) and spacing pattern (multiple singular heterocysts) that is essentially the same as that in symbiotic association. This is the heterocyst frequency and spacing pattern we hypothesize would be an experimental platform for enhanced hydrogen production. The mutant does grow with $N_2$ as the sole nitrogen source, but at a slower rate than the wild type and with about 30% of the rate of nitrogen fixation (acetylene reduction). Thus, the mutant physiology does not parallel that of the symbiotic growth state. The targeted gene encodes a protein unique to heterocyst-forming cyanobacteria that we have designated PatN. PatN has four distinct domains: a 30 amino acid cytoplasmic region, a transmembrane domain, a glutamine rich coiled-coil domain in the periplasm, and a signal sequence. Our working hypothesis is that PatN may be involved in the translocation of a negative acting heterocyst differentiation element, such as, for example, the RGSGR C-terminal pentapeptide of PatS. We are setting up a microscope for fluorescence recovery after photobleaching (FRAP) analysis to examine translocation properties of the mutant relative to the wild type. The RGSGR pentapeptide is also present in HetN, which is modeled to be involved in maintenance of the heterocyst spacing pattern. The putative HetN protein in *N. punctiforme* lacks the RGSGR motif. Bioinformatic analysis revealed two additional proteins encoded by *N. punctiforme* as containing the RGSGR motif. Single deletion mutants do not result in a multiple contiguous heterocyst (Mch) spacing pattern as is characteristic of *hetN* mutants of *Anabaena* sp. strain PCC 7120. Moreover, *patS* deletion mutants of *N. punctiforme* do not show an extensive Mch pattern. These results imply considerably more cross-talk or complementation in negative acting regulatory elements in *N. punctiforme* compared to *Anabaena* 7120. To confirm this suggestion we are constructing multiple deletion mutants.

**Transcriptomics.** To more rigorously analyze time course microarray data, we have adapted the Bayesian Analysis of Time Series (BATS) software. BATS analyses have reduced the number of statistically significant genes that are differentially expressed during heterocyst differentiation in the wild type and mutant strains. The reanalysis confirmed that the wild type (511 genes) and mutants impaired in HetR (1150 genes) and HetF (722 genes) (two positive acting regulatory elements) function have different patterns and numbers of genes differentially transcribed during nitrogen starvation. We interpret these data to verify that the primary role of HetR and HetF, directly or indirectly, is to influence gene transcription. Conversely, other than *patN* itself, no other genes are altered in their patterns of differential transcription in the *patN* mutant. This observation implies that PatN has a structural or functional as opposed to transcriptional role in the differentiation of heterocysts and supports the FRAP analysis above to define that role.

**Proteomics.** We have completed three replicates of three dimensional SDS-PAGE and LC fractionation of $N_2$-grown, wild type *N. punctiforme* and protein identification by MudPit MS/MS. These analyses yielded 1,210 identified proteins in the membrane (40,000 x g pellet), molecular complex (150,000 x g pellet) and soluble (150,000 x g supernatant) fractions. Known heterocyst structural and catalytic proteins were present in all fractions. We have completed two replicates of symbiotic colonies isolated from the hornwort *Anthoceros punctatus.* Due to the small amount of biomass, the symbiotic cell extracts were not subjected to centrifugal fractionation. These analyses yielded only 306 identified proteins. However, the identified proteins were highly enriched in cell envelope, energy metabolism and transporters. Present were PsbB and PsbC, but not PsbA, which may account for the lack of photosynthetic oxygen evolution in heterocysts and the lower rate of symbiotic $CO_2$ fixation by *N. punctiforme*. We are currently exploring methods to fractionate the small-volume, symbiotic cell extracts for better resolution of less abundant proteins.

# 98

## Pathway of Fermentative Hydrogen Production by Sulfate-Reducing Bacteria

Kimberly L. Keller[1]* (kellerkl@missouri.edu), Steve D. Brown,[2,3] Iris Porat,[2] Adam Deutschbauer,[4,3] Jennifer Kuhl,[4,3] Grant M. Zane,[1,3] and **Judy D. Wall**[1,3] (WallJ@missouri.edu)

[1]University of Missouri, Columbia; [2]Oak Ridge National Lab, Oak Ridge, Tenn.; [3]ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies), Berkeley, Calif.; and [4]Lawrence Berkeley National Laboratory, Berkeley, Calif.

**Project Goals: The production of hydrogen by fermentative pathways of the anaerobic sulfate-reducing bacteria (SRB) of the genus *Desulfovibrio* is the focus of the project. The limitations to hydrogen production identified in these model organisms may be informative for those microbes chosen for industrial hydrogen generation. We propose to determine the contribution of substrate-level phosphorylation to respiratory growth on sulfate and the contribution of respiration to fermentation of pyruvate. The enzymes for pyruvate oxidation will be established**

**in two strains of *Desulfovibrio* to identify the reduced product available for hydrogen generation. Electron sinks potentially competing with protons will be eliminated individually and together to determine the plasticity of electron flow to hydrogen.**

We are exploring the production of hydrogen by fermentative pathways of the anaerobic sulfate-reducing bacteria (SRB) of the genus *Desulfovibrio*. The SRB have not been considered model organisms for hydrogen production, yet they have a multifaceted hydrogen metabolism. Strains of *Desulfovibrio* can ferment organic acids in the absence of terminal electron acceptors and produce rather large amounts of hydrogen. We believe a study of the limitations to hydrogen production in these model organisms (in particular, *Desulfovibrio* G20 and *Desulfovibrio vulgaris* Hildenborough [DvH]) may be informative to decipher the flow of electrons in those organisms chosen for industrial application for hydrogen production.

**Aim 1:** We proposed to tease apart the contribution of fermentation to the respiratory energy budget to determine the dependence of the bacterium on this process. We want to eliminate substrate-level phosphorylation by creating a deletion of the gene encoding acetate kinase to confirm that this enzyme is essential for substrate-level phosphorylation during pyruvate fermentation. Determine the effect of this deletion on the efficiency of pyruvate and lactate respiration.

The acetate kinase in both organisms is annotated to be in a 10 gene operon. Transposon libraries of G20 and DvH selected under conditions that would allow mutations in genes non-essential for respiration lacked mutations in the gene, *ack*. Proteomic data from *Desulfovibrio* G20 indicated that the abundance of the enzymes involved in substrate-level phosphorylation was actually decreased when cultures were fermenting pyruvate compared to growth on lactate/sulfate. Microarray data from *D. vulgaris* indicated expression of acetate kinase was decreased when growing with hydrogen as a electron donor where substrate-level phosphorylation is not possible. Efforts are being made to use these growth modes to obtain a deletion of *ack* in *D. vulgaris*. Once obtained, the mutant will be characterized to determine the contribution of this pathway to the overall energy budget of the SRB.

**Aim 2:** We proposed to identify the enzyme(s) responsible for oxidizing pyruvate during fermentation and the role of formate, if any, in pyruvate fermentation. This information will distinguish the potential pathways to hydrogen from substrate oxidation. Two different enzymes are annotated in the genome, pyruvate:ferredoxin oxidoreductase and pyruvate formate lyase, each of which has multiple possible orthologs/paralogs.

Transposon mutants of the two annotated pyruvate formate-lyase genes have been obtained in both *Desulfovibrio* strains. Because of a developed markerless deletion system, work has begun to delete the two candidates, singly and in combination, monitoring the products of the metabolism of the mutants. Transposon mutants in three of the four operons encoding proteins annotated as pyruvate:ferredoxin oxidoreductases have also been obtained in both *Desulfovibrio* strains. Fitness experiments with a mini-pool of 1,163 transposon mutants of *Desulfovibrio* G20 indicated that one pyruvate ferredoxin oxidoreductase (Dde_1639) mutant did have decreased fitness in all experiments performed to date and Dde_3237 had no transposon insertions.

**Aim 3:** We proposed to channel electrons from alternative sinks to hydrogen during fermentation and determine the effects of removal of those sinks on the fermentation efficiency. We will attempt to eliminate alternative electron sinks that are likely to compete for protons, reducing the overall yield of hydrogen.

Proteomic and transcriptomic data from *Desulfovibrio* G20 indicate the importance of the fumarate reductase in growing cultures, especially in the later stages of growth. Transposon mutants lacking fumarate reduction to succinate are being studied since fumarate may be a possible electron sink in competition with protons and loss of this activity may alter the energy budget of the cells. Growth on fumarate is inhibited by the presence of formate, hydrogen, or carbon dioxide; whereas, formate is only accumulated during growth of G20 on fumarate inhibited with hydrogen. The hydrogen inhibition may be due to an inability to reoxidize ferredoxin or a possible blockage of proton pumping. In addition, the histidine kinase (Dde_1261) and the response regulator (Dde_1260) upstream of the fumarate reductase operon appear to play a criticial role in the utilization of fumarate as transposon mutants in either of these two genes renders the cells unable to grow on fumarate.

# 99

## Defining the Metabolic Landscape of *Chlamydomonas reinhardtii* by Large-Scale Annotation of its Encoded Open Reading Frames and Metabolic Network Modeling

Lila Ghamsari,[1] Balaji Santhanam,[1] Roger L. Chang,[2] Paul Jensen,[3] Patrick Ng,[4] Ani Manichaikul,[3] Erik F.Y. Hom,[5] Dawit Balcha,[1] Yun Shen,[1] Changyu Fan,[1] Tong Hao,[1] Bernhard Ø. Palsson,[2] Haiyuan Yu,[4] **Jason A. Papin**,[3] and **Kourosh Salehi-Ashtiani**[1,6*] (ksa3@nyu.edu)

[1]Center for Cancer Systems Biology and Department of Cancer Biology, Dana-Farber Cancer Institute, and Department of Genetics, Harvard Medical School, Boston, Mass.; [2]Department of Bioengineering, University of California, San Diego; [3]Department of Biomedical Engineering, University of Virginia, Charlottesville; [4]Department of Biological Statistics and Computational Biology and Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, N.Y.; [5]Department

of Molecular and Cellular Biology, Harvard University, Cambridge, Mass.; and [6]New York University Abu Dhabi, Abu Dhabi, UAE, and Center for Genomics and Systems Biology, Department of Biology, New York University, New York

**Goals: The release of the complete genome sequence of *Chlamydomonas reinhardtii* has made this unicellular alga an ideal model for metabolic engineering; however, the annotation of the relevant genes has not been verified yet and the much-needed metabolic network model is currently unavailable. Using the integrated annotation and metabolic network modeling that we recently established (*Manichaikul et al., Nature Methods 2009*), we are engaged in efforts to: 1) assign enzymatic functions to the annotated proteome of *C. reinhardtii*, 2) experimentally verify or refine the structure of the annotated open reading frames (ORFs), and 3) build a genome-wide metabolic network model for the organism based on the assigned metabolic functions.**

**Results:** We used the new *JGI* "filtered transcript models" (Chlre4_best_transcripts and Chlre4_best_proteins), and the *Augustus 5* models released through the *JGI* portal (http://genome.jgi-psf.org/Chlre4/Chlre4.home.html) for both functional assignments and structural annotation verifications. Enzymatic functional assignments were made by associating Enzyme Classification (EC) numbers through reciprocal blast searches against UniProt (and AraCyc) enzyme database (with over 100,000 protein entries). The best match for each translated ORF was identified (with an e-value threshold of $10^{-3}$) and the EC number from the UniProt best match was transferred on to the ORF. We extended the EC assignments to the respective paralogs of the ORFs by clustering ORFs using BLASTCLUST (sequence identity cut-off of 35% and sequence length cut-off of 70%) within each annotation group (i.e., *Augustus* 5 and *JGI filtered models*). Altogether, we were able to assign over 900 enzyme annotations to 1,427 *JGI* and to 1,877 *Augustus* models. Over 93% of the EC terms were assigned to both *JGI* and *Augustus* models. We then carried out all possible pairwise alignments between the *JGI* and *Augustus* transcripts that had been assigned the same EC numbers by the above-mentioned procedure. In contrast to the high overlap between the two models in terms of EC assignments, less than half of each set were found to be 100% identical in sequence, indicating that the structural annotation of many of the two sets differ from one another.

To experimentally verify the structure of both *JGI* and *Augustus* ORF models, we carried out open reading frame (ORF) verification by RT-PCR on all ORFs that we had assigned EC numbers to (as well as a set of positive control ORFs). Following optimization of the RT-PCR procedure for high GC content of the *C. reinhardtii* transcriptome, we tested the structure of the metabolic-related ORF models by reverse transcription-PCR of the functionally annotated ORFs. Following cloning, we carried out 454FLX sequencing of the ORFs. Based on alignment of the 454FLX reads to the ORF predicted sequences, we obtained more than 90% coverage for 80% of the metabolic ORFs. Only 99 ORFs were not verified using this experimental pipeline. We obtained expression evidence for 93% of the metabolic ORFs in the algal cells grown under constant light and in the presence of acetate.

Using our in-house generated functional annotation (*described above*), combined with literature and publicly available database resources, we have reconstructed the first genome-scale reconstruction of *C. reinhardtii* metabolic network, accounting for all pathways and metabolic functions indicated. The reconstruction accounts for 1,080 genes, associated with 2,190 reactions and 1,068 unique metabolites. Our reconstruction accounts for multiple wavelengths of light and includes considerable expansion of fatty acid metabolism over previous reconstructions. Further, the metabolic network reconstruction provides a greater level of compartmentalization than existing reconstructions of *C. reinhardtii*, with the inclusion of the lumen as a distinct component of the chloroplast for photosynthetic functionality, and the eyespot used to guide the flagella in phototaxis.

**Conclusion:** Our validated and comprehensive genome-scale reconstruction of *C. reinhardtii* metabolism provides a valuable quantitative and predictive resource for metabolic engineering toward improved production of biofuels and other commercial targets. The verified metabolic ORF clones will provide the experimental resource needed for downstream experiments and will be made available to the research community.

The DOE support for this research was generously made through a grant from the Office of Science (Biological and Environmental Research), U.S. Department of Energy, grant No. DE-FG02-07ER64496 to J.P. and K.S.-A.

# Computing for Bioenergy

## 100

### Multiscale Coarse-Grain Simulation Studies of Cellulosic Biomass

Goundla Srinivas[1]* (srinivasg@ornl.gov), Dennis Glass,[1] Xiaolin Cheng,[1] Jeremy Smith,[1,2] Mark S. Gordon,[3] Yana Kholod,[3] Monica H. Lamm,[3] Sergiy Markutsya,[3] John Baluyut,[3] and Theresa L. Windus

[1]Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [2]Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville; and [3]Applied Mathematics and Computational Science, Ames Laboratory, Department of Chemistry, Iowa State University, Ames

**Project Goals:**

1. **To develop computationally affordable large scale coarse-grain force field for cellulosic biomass.**
2. **Study structure and dynamics of cellulosic biomass from a micro and macroscopic viewpoint in order to understand biomass recalcitrance.**

Understanding cellulose structure and dynamics is important for improving the process of conversion of biomass to ethanol. Here, we explore cellulose fibril structure and dynamics and catalytic degradable pathways by developing multiscale methods for extending the time and length scales accessible to biomolecular simulation on massively parallel supercomputers. For this purpose, high level quantum mechanics calculations are performed using the fragment molecular orbital (FMO) method in GAMESS (General Atomic and Molecular Electronic Structure System). The FMO simulations are used to study the relevant catalyzed reaction paths, with the solvent represented by the effective fragment potential (EFP) method. The combined FMO/EFP methods are used with molecular dynamic simulations to provide coarse grained (CG) potentials using the force matching method. Further, accurate parameterization of glucose monomers derived from quantum mechanical calculations is used as input for classical molecular dynamics which in turn is utilized to develop a large-scale force field for the cellulose fibril. Without the use of constraints the CG crystalline fibril is found to remain stable (>1μs). Analyzing the static coherent structure factors reveals the ability of the present CG model to differentiate between intra-plane and inter-plane interactions in the fibril. The model is successfully extended to represent various amorphous cellulose fibrils as well. Further, we have carried out a REACH CG analysis of the cellulose fibril using information on correlated motions. Together these simulation results contribute to our understanding of biomass recalcitrance to hydrolysis.

# 101

## Bayesian Computational Predictions of Gene Regulation in the α-Proteobacteria

**Charles E. Lawrence**[1] (charles_lawrence@brown.edu), **Lee Newberg**[2,3] (lee.newberg@wadsworth.org), William Thompson[1] (william_thompson_1@brown.edu), and **Lee Ann McCue**[4]* (leeann.mccue@pnl.gov)

[1]Center for Computational Molecular Biology and the Division of Applied Mathematics, Brown University, Providence, R.I.; [2]The Wadsworth Center, New York State Department of Health, Albany; [3]Department of Computer Science, Rensselaer Polytechnic Institute, Troy, N.Y.; and [4]Pacific Northwest National Laboratory, Richland, Wash.

http://www.brown.edu/Research/CCMB/

**Project Goals: see below**

Decreasing America's dependence on foreign energy sources and reducing the emission of greenhouse gases through the development of biofuels are important national priorities. These priorities have catalyzed research on cellulosic ethanol as a clean, renewable energy source to replace fossil fuels, and biohydrogen as a carbon-free energy carrier. Turning these biofuels into viable alternative energy sources requires further research into the degradation of cellulose and fermentation of the resulting sugars, and the metabolic and regulatory networks of biohydrogen production. The genomes of many of the microbial species capable of these processes have been sequenced by the GSP and other programs, and many more are expected soon. These sequence data provide a wealth of information to explore nature's solutions for the production of biofuels. In particular, among the α-proteobacterial species with genome sequence data available are several species with metabolic capabilities of interest, including efficient fermentation of sugars to ethanol and the ability to produce hydrogen. Understanding the regulatory mechanisms and complex interplay of metabolic processes in these species is key to realizing the promise of biofuels. Thus, our research goal is to identify the ensemble of solutions that have been explored by the α-proteobacteria to regulate the metabolic processes key to biofuel production. We have predicted putative regulatory sites and motifs in clades from 63 α-proteobacterial species, to identify regulatory mechanisms and reconstruct the ancestral states of the regulatory networks for the efficient fermentation of sugars to ethanol and the production of biohydrogen.

# 102

## Vertically Integrated Metabolic Engineering Process: Computational and Experimental Optimization of *Escherichia coli* Toward Fatty Acid Production

**George Church** (guido@genetics.med.harvard.edu), Nicholas Guido,[1]* and Graham Rockwell[1,2]

[1]Dept. of Genetics, Harvard Medical School; and [2]Bioinformatics Program, Boston University

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

**Project Goals: The goal for this work is to integrate computational, mutagenesis and selection tools toward a generalized method for the optimization of production strains synthesizing high-value products. We are modeling and mutagenizing strains of *Escherichia coli* to increase fatty acid production as an initial test case.**

Advancements in both computational and metabolic selection are required to complement new in vivo mutagenesis technology enabling metabolic engineering from design to strain production. Here we build new in silico strategies, based on flux balance analysis (FBA), to provide gene level engineering targets for the optimization of complex metabolic processes at the genome scale, previously unobtainable by other methods. Fatty acid production is the test case. These engineering instructions are brought to realization via multiplexed recombination (MAGE), optimizing transcription and translation levels, demonstrating that the model successfully predicts metabolic targets to increase fatty acids. We also develop a unique in vivo selection method to select for cells that have increased fatty acid production as a result of our genetic manipulations. This work shows that computational instructions can be used in conjunction with

mutagenesis and specific selections to create a process of metabolic engineering which represents a general method for the optimization of biological metabolite production.

# 103

## Computational Study of Gating Elements for Proton Pumping in Cytochrome c Oxidase

S. Yang* (syang@chem.wisc.edu), J**. Mitchell**, and Q**. Cui**

BACTER Graduate Program, University of Wisconsin, Madison

Cytochrome c Oxidase (CcO) is the terminal component of the electron transfer pathway in the cellular respiration. It reduces oxygen to water and utilizes the released chemical free energy to pump protons across the membrane in a stoichiometric and efficient fashion. After decades of structural, kinetic, mutagenesis and computational analysis, the identity of the gating element(s) in CcO remains hotly debated. Two studies have been published recently, which focused on the conformational preference of a key side-chain and water molecules. In one of them, the side-chain of a key branching residue, Glu 286 (*R. sphaeroides* number) was found to strongly prefer the downward orientation when deprotonated, which presumably shut off the back flow of protons. Therefore, it was proposed that Glu 286 is the essential valve that minimizes leakage of the pump. In the other study, the orientation of the water molecules in the active site switched depending on the redox state of heme a and the binuclear center. This led to the proposal that water wire reorientation is another major element that controls the appropriate branching of proton transfers, which was shown by elegant kinetic analysis as crucial to the thermodynamic efficiency of proton pumping.

In this report, we systematically analyze the relevant energetics using elaborate molecular models for CcO. The results show that neither Glu 286 rotation nor water wire reorientations acts as the robust gating factor, and the observations from the previous studies are most likely due to the use of simplified models of CcO, which highlights the importance of carefully benchmarking the molecular model for an analysis of complex proton pumping systems.

# 104

## Uncovering Mechanisms of Cellobiose Hydration

Madeleine Pincu[1]* (mpincu@uci.edu), John P. Simons,[2] and **Robert B. Gerber**[3]

[1]University of California-Irvine, Department of Chemistry; [2]University of Oxford, Chemistry Department, Physical and Theoretical Chemistry Laboratory, Oxford, United Kingdom; and [3]The Hebrew University, The Fritz Haber Research Center, Jerusalem, Israel, and University of California-Irvine, Department of Chemistry

**Project Goals: We aim to study with theoretical tools saccharides—the most abundant bio-molecules on earth. Our studies are designed to shed light on the fundamental structure, energy configuration and temperature-dependent dynamic behaviour of saccharides in isolation and in the hydrated state. We also seek to understand the effect of ions and salts on these compounds as these have important implications to the understanding of biological systems. To validate our theoretical findings, we collaborate with several experimentalists, in particular with Professors J. P. Simons (of Oxford) and Ilana Bar (of Beer Sheva University), who provide us with spectroscopy data for some of our systems. We expect this work to illuminate our understanding of the chemistry of polymers of sugar, in particular of cellulose—a source of bio-energy. In the process, we will develop and validate new computational tools designed to simulate moderate to large size bio-molecular polymers within reasonable time frames.**

We explore conformational and structural dynamics of cellobiose and their micro-hydrated complexes isolated in the gas phase, with a combination of theoretical and experimental tools. Their structures at low temperature have been determined through double resonance, IR-UV vibrational spectroscopy conducted under molecular beam conditions, substituting $D_2O$ for $H_2O$ to separate isotopically, the carbohydrate (OH) bands from the hydration (OD) bands. Car-Parrinello (CP2K) simulations, employing dispersion corrected DFT potentials and conducted 'on-the–fly' from ~20K to ~300K, were used to explore the consequences of raising the temperature on the infra-red vibration spectrum.

Our findings are: (1) Good agreement between experiment and theory, which gives us confidence in our theoretical methods. (2) Increasing hydration from 1 to 10 $H_2O$ molecules reveals a persistent motif developing, in which hydration proceeds with water forming a bridging network across the trans glycosidic bond, (from OH6' to OH4'). This "bridge" remains stable for the 5+ps duration of the (CP2K) dynamic simulation at ~300K, despite individual fluctuations in the intra- and inter-molecular hydrogen bonding. (3) Although in nature the trans-cellobiose conformer is more stable than cis, in solution at room temperature, and although our calculations predict that their micro-hydrated energies should equalize for a water cluster containing ~12 molecules, we find that a cluster of 10 water molecules organizes with either cis or trans cellobiose at surfactant position relative to that of the water cluster.

The connection between hydration in small complexes and in the extended cluster is discussed and important implications of the results for properties of saccharides are suggested.

# 105

## The Transcriptional Architecture of *Thermotoga maritima*

H. Latif[1]* (halatif@ucsd.edu), V.A. Portnoy,[1] Y. Tarasova,[1] A.C. Schrimpe-Rutledge,[2] H. Nagarajan,[1] R.D. Smith,[2] J.N. Adkins,[2] D.H. Lee,[1] Y. Qiu,[1] **B.O. Palsson**,[1] and K**. Zengler**[1]

[1]Department of Bioengineering, University of California San Diego, La Jolla; and [2]Pacific Northwest National Laboratory, Richland, Wash.

**Project Goals: Validate and improve the existing genome-scale metabolic reconstruction of *T. maritima*. Reconstruct the transcriptional regulatory network (TRN) of *T. maritima*. Generate and use the integrated in silico model of *T. maritima* to assess growth conditions for optimal production of hydrogen.**

The hydrogen producing, hyperthermophilic microorganism *Thermotoga maritima* is the focus of this work. Previously, the *T. maritima* metabolic model has been constructed containing 479 metabolic genes, 565 metabolites and 646 reactions. The metabolic model was then updated to include 478 protein structures yielding the first three dimensional metabolic reconstruction. In addition, we have extended the metabolic model by including all reactions for transcription and translation (see abstract by Lerman et al.), laying the ground work for *in silico* gene expression predictions. Fundamental for exploring the flow of information from genes, to transcripts, to proteins is the structural and operational genome annotation of *T. maritima*. This multi-level annotation is being accomplished through the development and integration of various genome-wide, experimentally derived data. Customized high density, whole genome microarrays together with RNAseq data were utilized for elucidation of the *T. maritima* transcriptome. Growth conditions included exponential phase, hydrogen induced stationary phase, heat shock, acid shock and carbon limiting stationary phase. Combining these data types resulted in expression of >97% of the genome. Furthermore, proteomic data was generated from exponential phase and stationary phase cultures using LC-MS/MS and mapped against a stop-to-stop database. Over 41,000 unique peptides were mapped to ~1,400 coding regions, covering over 73% of the entire proteome. The integration of proteome and transcriptome data provided a highly improved structural annotation (ORFs) of the *T. maritima* genome. The operational annotation, consisting of operons and transcription units, was subsequently resolved by including genome-wide transcription start site (TSS) data, determined at single base pair resolution. Integrating transcriptomics, proteomics, and TSS data yielded the transcriptional architecture of *T. maritima.* This data now provides the basis upon which the transcriptional regulatory network for this hyperthermophilic bacterium will be built.

# 106

## Discovery of Phenotype-Related Biochemical Processes with Application to Biological Hydrogen Production

Andrea M. Rocha[1]* (amrocha@mail.usf.edu), Matthew C. Schmidt,[2,3] William Hendrix,[2,3] Kanchana Padmanabhan,[2,3] Yekaterina Shpanskaya,[2] James R. Mihelcic,[1] and **Nagiza F. Samatova**[2,3] (samatovan@ornl.gov)

[1]University of South Florida, Tampa; [2]North Carolina State University, Raleigh; and [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.

**Project Goals: The purpose of this project is to develop computational approaches to identify phenotype-related cellular subsystems and demonstrate predictability of these techniques to hydrogen production by dark fermentative and acid-tolerant bacteria.**

Many microbial communities in natural environments exhibit phenotypes that directly cause particular diseases, convert biomass or wastewater to energy, or degrade various environmental contaminants. Understanding of how interacting biochemical pathways in these communities realize specific phenotypic traits (e.g., carbon fixation, hydrogen production) is critical for addressing health, bioremediation, or bioenergy problems that, arguably, cannot be solved by experiments alone.

To complement experimental approaches, in this study, we develop graph-theoretical and statistical methods for *in silico* prediction of the cellular subsystems that are related to the expression of a target phenotype. First, our Network Instance-Based Biased Subgraph Search (**NIBBS**) is capable of comparing hundreds of genome-scale metabolic networks to identify *metabolic subsystems* that are statistically biased toward phenotype-expressing organisms. NIBBS accurately approximates the set of all *biased* network motifs in a set of metabolic networks. From the results obtained, for example, we were able to predict a number of metabolic subsystems that are likely related to biohydrogen production, such as acetate and butyrate fermentation, fatty acid biosynthesis, amino acid metabolism, and nitrogen metabolism. Also, the genome-scale comparative network analysis enabled us to predict pathway cross-talks, including those involved in production of Acetyl-CoA.

Second, our $\alpha,\beta$*-motifs* approach allows for identification of *functional modules* that, in addition to metabolic subsystems, could include their regulators, sensors, transporters, and even uncharacterized proteins that are predicted to be related to the target phenotype. By comparing hundreds of genome-scale networks of functionally associated proteins, our method identifies those functional modules that are enriched in at least $\alpha$ networks of phenotype-expressing organisms but may still appear in no more than $\beta$ networks of organisms that do not exhibit the target phenotype. Using $\alpha,\beta$*-motifs* approach, for instance, we were able to identify

clusters of genes responsible for synthesis, metal insertion, or regulation of hydrogenase and nitrogenase enzymes complexes. Within hydrogen producers, these two complexes play important roles in production of hydrogen.

Third, our Dense ENriched Subgraph Enumeration (DENSE) algorithm allows for incorporating partial *prior* knowledge about the proteins involved in a phenotype-related process and enriches that knowledge with newly identified sets of functionally associated proteins present in individual phenotype expressing organisms. When applied to a network of functionally associated proteins in the dark fermentative, hydrogen producing bacterium, *Clostridium acetobutylicum*, we were able to predict known and novel relationships including those with regulatory, signaling, and uncharacterized proteins.

Finally, we integrate the information obtained from application of these complementary approaches not only to allow for further understanding of networks related to hydrogen production, but also to provide insights into metabolic networks and system controls involved in expression of microbial traits. Such information is necessary for advancing engineering approaches that result in more efficient biohydrogen production.

This research is supported by both the Office of Biological and Environmental Research and by the Office of Advanced Scientific Computing Research of the U.S. Department of Energy. The work by A.M.R. was supported by the Delores Auzenne Fellowship and the Alfred P. Sloan Minority PhD Scholarship Program.

# 107

## Ultrascale Computational Modeling of Phenotype-Specific Metabolic Processes in Microbial Communities

**Nagiza F. Samatova**[1,2]* (samatovan@ornl.gov), **Chongle Pan**[2] (pcl@ornl.gov), **Robert (Bob) L. Hettich**[2] (hettichrl@ornl.gov), and **Jill Banfield**[3] (jbanfield@berkeley.edu)

[1]North Carolina State University, Raleigh; [2]Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [3]University of California, Berkeley

**Project Goals: Computational modeling methods are aimed to be developed for revealing phenotype-specific metabolic processes and their crosstalks and applied to the critical DOE problem of acid mine drainage (AMD). The apex of the project is a procedure for: (1) identification and expression-level characterization of phenotype-related genes; (2) reconstruction of phenotype-specific metabolic pathways; and (3) elucidation of symbiotic/competing interplays between these pathways.**

Many microbial communities in natural environments exhibit phenotypes of interest to DOE, such as oxidization of pyrite ore that leads to acid mine drainage, breaking down the lignocellulosic barrier of biomass, and the biodegradation of various environmental contaminants. Addressing bioremediation and bioenergy problems will require an understanding of how interacting biochemical pathways in these communities realize specific phenotypic traits (nitrogen and carbon fixation, resistance to heavy metals, tolerance to pH perturbations, etc.). This problem cannot be solved by experiments alone. There is a need for computational modeling methods that will reveal phenotype-related "signals'" and their combinatorial interplay by comparing potentially hundreds of microorganisms with millions of genes organized into thousands of metabolic pathways, which are uncertainly defined. These methods are being applied to the acid mine drainage (AMD) community to help answer long-standing questions regarding the role of fine-scale variation in adaptation to dynamic environmental conditions and community composition.

We performed and published quantitative proteomics comparison of field AMD biofilm to laboratory AMD biofilm. To enable laboratory studies of growth, production, and ecology of AMD microbial communities, a culturing system was designed to reproduce natural biofilms, including organisms that are recalcitrant to cultivation. A comprehensive metabolic labeling-based quantitative proteomic analysis was utilized to verify that natural and laboratory communities were comparable at the functional level. Results confirmed that the composition and core metabolic activities of laboratory-grown communities were similar to a natural community, including the presence of active, low abundance bacteria and archaea that have not yet been isolated. However, laboratory growth rates were slow compared to natural communities, and this correlated with increased abundance of stress response proteins for the dominant bacteria in laboratory communities. Modification of cultivation conditions reduced the abundance of stress response proteins and increased laboratory community growth rates. This was the first application of a metabolic labeling-based quantitative proteomic analysis at the community level and resulted in a model microbial community system ideal for testing physiological and ecological hypotheses

We also performed and published quantitative proteomics study of pH perturbation to functionally characterize laboratory-cultivated acidophilic communities sustained in pH 1.45 or 0.85 conditions. The distributions of all proteins identified for individual organisms indicated biases for either high or low pH, and suggests pH-specific niche partitioning for low abundance bacteria and archaea. Although the proteome of the dominant bacterium, *Leptospirillum* group II, was largely unaffected by pH treatments, analysis of functional categories indicated proteins involved in amino acid and nucleotide metabolism, as well as cell membrane/envelope biogenesis were generally more abundant at high pH. Results indicate solution pH may play an important role in shaping community membership and biofilm structure. Proteomic analysis of communities also revealed differences in the number of phage proteins detected across biological replicates. Stochastic spatial heterogeneity of viral outbreaks may also play a role in shaping community

structure. Quantitative proteomic comparisons showed distinct differences in community composition and metabolic function of individual organisms during different pH treatments, and confirms the importance of specific geochemical parameters that 'fine-tune' acidophilic microbial community structure and function at the species and strain level.

In support of these proteomics studies, we developed, published, and released a *de novo* sequencing algorithm, **Vonode**, to exploit the potential of high-resolution MS/MS data by using a unique tag scoring function and a novel type of spectrum graphs. When compared to an established *de novo* sequence algorithm, PepNovo v2.0, the Vonode algorithm inferred sequence tags for 11,422 (vs. 2,573) spectra at an average length of 5.5 (vs. 6.0) residues with 84% (vs. 65%) accuracy of inferred consensus sequence tags.

We also developed and released a module for **ProRata**, a data analysis algorithm for quantitative proteomics, to address the following two critical needs: (1) to combine multiple replicates and to assess the reproducibility of measurements to obtain reliable quantification information and (2) to compare two unlabeled field samples of interest to a labeled reference sample grown in the laboratory because we cannot label metabolic labeling to a field sample.

We developed graph-theoretical and statistical methods for *in silico* prediction of the cellular subsystems that are related to the expression of a target phenotype: (1) the Network Instance-Based Biased Subgraph Search (NIBBS) is capable of comparing hundreds of genome-scale metabolic networks to identify *metabolic subsystems* that are statistically biased toward phenotype-expressing organisms; (2) ther α,β-motifs approach allows for identification of *functional modules* that, in addition to metabolic subsystems, could include their regulators, sensors, transporters, and even uncharacterized proteins that are predicted to be related to the target phenotype; and (3) the Dense ENriched Subgraph Enumeration (DENSE) algorithm allows for incorporating partial *prior* knowledge about the proteins involved in a phenotype-related process and enriches that knowledge with newly identified sets of functionally associated proteins present in individual phenotype expressing organisms.

From the results obtained, for example, we were able to predict cellular subsystems that are likely related to various phenotypes such as acid tolerance, biohydrogen production, aerobic and anaerobic respiration, etc. Also, the genome-scale comparative network analysis enabled us to predict pathway crosstalks and to perform a systematic study on various mechanisms underlying crosstalks.

This research is supported by both the Office of Biological and Environmental Research and by the Office of Advanced Scientific Computing Research of the U.S. Department of Energy.

# 108

## Functional Annotation of Hierarchical Modularity

Kanchana Padmanabhan[1,2]* (kpadman @ncsu.edu), Kuangyu Wang,[1] and **Nagiza F. Samatova**[1,2] (samatovan@ornl.gov)

[1]North Carolina State University, Raleigh; and [2]Oak Ridge National Laboratory, Oak Ridge, Tenn.

**Project Goals: The goals of this project are (1) To develop a method that would assess functional coherence and provide annotation of hierarchically structured modules; (2) To design a biologically relevant functional coherence scoring metric; (3) To reconstruct a hierarchical modularity of cellular organization from a "bag of genes" allowing multi-functional genes or proteins to be part of multiple functional modules in the hierarchy.**

Network motifs are recurring, statistically significant patterns of node interactions that act as building blocks of complex networks. In biological networks of molecular interactions in a cell, such as protein-protein interaction networks or gene transcriptional regulatory networks, network motifs that are biologically relevant are also functionally coherent, or form functional modules, such as a ribosomal module synthesizing proteins or a signal transduction system governing bacterial chemotaxis. These functionally homogenous modules combine in a hierarchical manner into larger, less cohesive subsystems, thus revealing one of the essential design principles of system-level cellular organization and function—*hierarchical modularity*.

Arguably, hierarchical modularity has not been explicitly taken into consideration by most, if not all, functional annotation systems. Instead, a functional module is traditionally viewed as a "bag of genes," and methods that assess its functional coherence, or provide functional annotation, analyze this bag in its entirety. As a result, the existing methods would often fail to assign a statistically significant functional coherence score to biologically relevant molecular machines (see Table 1).

To address this gap, we developed a methodology for hierarchical functional annotation of biological cal network motifs. Given the hierarchical taxonomy of functional concepts (e.g., Gene Ontology) and the association of individual genes or proteins with these concepts (e.g., GO terms), our method will assign a *Hierarchical Modularity Score* (**HMS**) to each node in the hierarchy of functional modules; the HMS score and its *p*-value measure functional coherence of each module in the hierarchy. While existing methods annotate each module with a set of "enriched" functional terms in a bag of genes, our complementary method provides hierarchical functional annotation of the modules and their components that are hierarchically organized.

A hierarchical organization of functional modules often comes as a bi-product of cluster analysis of gene expression

Table 1: Statistical significance of protein pairs' functional coherence in *Saccharomyces cerevisiae*.

| Protein Pair | | | | *p*-value (pair/module/module size) | | | | | Ref |
|---|---|---|---|---|---|---|---|---|---|
| ID | Description | ID | Description | HMS | [1] | [2] | [3] | [4] | |
| SRB2 | Subunit of the RNA polymerase II mediator complex | RPB9 | RNA polymerase II subunit B12.6 | 0.022/ 2.2 $e^{-16}$// 57 | 0.48 | 0.12/ 1.0/ 57 | 0.333/ 1.0/ 57 | 0.98/ 1.0/ 57 | [5] |
| SNU13 | RNA binding protein | DIB1 | 17-kDa component of the U4/ U6aU5 tri-snRNP | 0.003/ 0.003/ 2 | 0.23 | 0.01/ 0.01/ 2 | 0.1/ 0.1/ 2 | 0.42/ 0.42/ 2 | [6] |
| HAP1 | Zinc finger transcription factor involved in the complex regulation of gene expression in response to levels of heme and oxygen | RPM2 | Protein subunit of mitochondrial RNase P | 0.006/ 3.207$e^{-8}$ / 3 | 0.214 | 0.1/ 1.0/ 3 | 0.02/ 1.0/ 3 | 0.51/ 1.0/ 3 | [8] |
| NSR1 | Nucleolar protein that binds nuclear localization sequences | DBP2 | Essential ATP-dependent RNA helicase of the DEAD-box protein family | 0.012/ 0.012/ 2 | 0.44 | 0.1/ 0.1/ 2 | 0.13/ 0.13/ 2 | 0.74/ 0.74/ 2 | [7] |

Table 2: Skill metrics for *Saccharomyces cerevisiae* KEGG experiments.

| | KEGG | Heidke Score | Pierce Score | Gerrity Score |
|---|---|---|---|---|
| HMS | Level 1 | 1 | 1 | 1 |
| [1] | | 1 | 1 | 1 |
| HMS | Level 2 | 0.917 | 0.923 | 0.887 |
| [1] | | 0.61 | 0.79 | 0.65 |
| HMS | Level 3 | 0.916 | 0.923 | 0.931 |
| [1] | | 0.71 | 0.73 | 0.73 |

data or protein interaction data. Otherwise, our method will automatically build such a hierarchy by directly incorporating the functional taxonomy information into the hierarchy search process and by allowing multi-functional genes to be part of more than one component in the hierarchy. In addition, its underlying HMS scoring metric ensures that functional specificity of the terms across different levels of the hierarchical taxonomy is properly treated.

We have evaluated our method using *Saccharomyces cerevisiae* data from KEGG and MIPS using GO ontology as the underlying hierarchical taxonomy of functional concepts. Table 2 illustrates biological relevance of the hierarchical modularity built by our method from a set of genes in various KEGG pathways: at various levels of the hierarchy, the corresponding modules match quite well with the manually-curated hierarchy of pathways in KEGG. We obtained similar results for the protein complexes in the MIPS database. We provide literature evidence for several functional modules that have been identified by HMS as significant both at the protein pairs and at the module levels but have been missed by some existing methods (see examples in Table 1).

This research is supported by both the Office of Biological and Environmental Research and by the Office of Advanced Scientific Computing Research of the U.S. Department of Energy.

### References


1. Pandey J., *et al.* Functional characterization and topological modularity of molecular interaction networks. BMC *Bioinformatics*, 2010, vol: 11, pp: S35.
2. Bauer S., *et al.* Ontologizer 2.0-a multifunctional tool for go term enrichment analysis and data exploration. *Bioinformatics*, 2008, vol: 24(14), pp: 1650.
3. Boyle E. I., *et al.* GO::TermFinder–open source software for accessing gene ontology information and finding significantly enriched gene ontology terms associated with a list of genes. *Bioinformatics*, 2004, vol: 20(18), pp: 3710.
4. Huang D. W., *et al.* Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 2009, vol: 4(1), pp: 44.
5. Myer V. E., *et al.* RNA polymerase II holoenzymes and subcomplexes. *The Journal of Biological Chemistry*, 1998, vol: 273, pp: 27757.
6. Scott W. S., *et al.* Purification of the yeast U4/U6U5 small nuclear ribonucleoprotein particle and identification of its proteins. *Proceedings of the National Academy of Sciences*, 1999, vol: 96(13), pp: 7226.
7. Tai S.L., *et al.* Acclimation of *Saccharomyces cerevisiae* to low temperature: A chemostat-based transcriptome analysis. *Molecular Biology of the Cell*, 2007, vol: 18, pp: 5100.
8. Hach A., *et al.* A new class of repression modules is critical for heme regulation of the yeast transcriptional activator Hap1. *Molecular and Cellular Biology*, 1999, vol: 19, pp: 4324.


Student Oral Presentation–Tuesday

# 109

## Elucidation of Symbiotic and Competing Metabolic Pathway Crosstalks

Kuangyu Wang[1]* (kwang2@ncsu.edu), Kanchana Padmanabhan,[1,2] and **Nagiza F. Samatova**[1,2] (samatovan@ornl.gov)

[1]North Carolina State University, Raleigh; and [2]Oak Ridge National Laboratory, Oak Ridge, Tenn.

**Project Goals: This project is focused on a systems-level understanding of biological nature of metabolic pathway crosstalks in bacterial organisms. The overall goal is 1) to provide a systematic characterization of metabolic pathway crosstalk mechanisms, such as those due to physical and genetic interactions, 2) to construct pathway crosstalk networks based on various types of functional genomics data and provide biological interpretations for discovered patterns in such networks, and 3) to confirm and validate the predicted crosstalks using literature searches and gene expression data across different conditions using *Escherichia coli* as our model organism.**

Given the complex nature of a phenotype in a microbial community, it is likely that not a single metabolic pathway but a group of phenotype-related pathways jointly function to accomplish a particular phenotype. For example, in a community, interplays between pathways may exist not only in a cell, but also across cells of different species. From a metabolite-centric perspective, pathways can interact cooperatively by exchanging intermediates, competitively by inputting or outputting common intermediates, or incompatibly by requiring different conditions to function. From a gene-centric perspective, genes in pathways can be regulated in a correlated manner or an anti-correlated manner. In this study, we aim to systematically characterize the mechanisms underlying various metabolic pathway cross-talks.

Traditionally, metabolic pathways are viewed as a linear sequence of metabolic reactions, where the product metabolites of one reaction are used as the substrate metabolites of the next reaction. Arguably, such an abstraction de-emphasizes the prevalence of crosstalks (non-additive interactions) between individual pathways. A better understanding of underlying mechanisms of pathway crosstalks will help predict processing of various environmental signals and conduct metabolic engineering for the purpose of bioenergy production. Using *Escherichia coli* as our model organism, here we present a systematic study that fuses genomics and proteomics information in order to predict crosstalks between metabolic pathways such as those extracted from KEGG.

A large number of non-additive interactions or crosstalks exist globally; however, the mechanisms underlying pathway crosstalks are limited. Although there is no universal categorization of biological mechanisms that underlie crosstalks, in this study, we have made attempts towards a possible classification that takes into consideration various types of evidences such as 1) physical via direct binding, 2) biochemical via phosphorylation, and 3) functional via transcriptional regulation.

By analyzing various types of biological networks, we infer the clues about putative metabolic pathway crosstalks. We further analyze those clues for possible positive and negative correlations and identify those that are hypothesized to be orthogonal. We devise methodology for understanding the higher-level organization of pathway crosstalk networks derived from these clues and provide biological interpretation of and literature evidence for the discovered patterns in those networks.

This research is supported by both the Office of Biological and Environmental Research and by the Office of Advanced Scientific Computing Research of the U.S. Department of Energy.

# 110

## Numerical Optimization Algorithms and Software for Systems Biology: von Bertalanffy 1.0 : A COBRA Toolbox Extension to Thermodynamically Constrain Metabolic Models

Ronan M.T. Fleming[1]* (ronan.mt.fleming@gmail.com), Ines Thiele,[2] and **Michael A. Saunders**[3]

[1]Science Institute and Center for Systems Biology, University of Iceland, Reykjavik; [2]Center for Systems Biology and Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Reykjavik; and [3]Dept of Management Science and Engineering, Stanford University

http://opencobra.sourceforge.net

**Project Goals: This project aims to reconstruct genome-scale models of metabolism and macromolecular synthesis and to develop algorithms capable of solving the resulting large, stiff and ill-scaled matrices. We aim to combine state of the art reconstruction and constraint-based modeling and analysis tools with high-end linear optimization solvers and convex flux balance analysis. The incorporation of thermodynamic information in addition to environmental constraints will allow an accurate assessment of feasible steady states. While we will prototype the reconstruction and algorithm developments with *Escherichia coli*, we will employ the resulting networks to determine thermodynamically favorable pathways for hydrogen production by *Thermotoga maritima*.**

In flux balance analysis of genome scale stoichiometric models of metabolism, the principal constraints are uptake or secretion rates, the steady state mass conservation assumption and reaction directionality. Here, we introduce an algorithmic pipeline for quantitative assignment of reaction directionality in multicompartmental genome scale models based on an application of the second law of thermodynamics to each reaction. Given experimental or computationally estimated standard metabolite species Gibbs energy and metabolite concentrations, the algorithms bounds reaction Gibbs energy, which is transformed to *in vivo* pH, temperature, ionic strength and electrical potential. This toolbox may be used to distinguish between thermodynamically feasible and thermodynamically infeasible metabolic pathways. This is a critical first step in the rational design of novel strategies for production of biofuels. This cross platform MATLAB extension to the COnstraint-Based Reconsruction and Analysis (COBRA) toolbox, is computationally efficient, extensively documented and open source.

Student Oral Presentation–Tuesday

# 111

## Numerical Optimization Algorithms and Software for Systems Biology: Existence of Positive Equilibria for Mass Conserving and Mass-Action Biochemical Reaction Networks with a Single-Terminal-Linkage Class

Santiago Akle[1]* (akle@stanford.edu), Onkar Dalal[1] (onkar@stanford.edu), Ronan Fleming[2] (ronan.mt.fleming@gmail.com), **Michael Saunders**[3] (saunders@stanford.edu), Nicole Taheri[1]* (ntaheri@stanford.edu), and **Yinyu Ye**[3] (yinyu-ye@stanford.edu)

[1]Institute for Computational and Mathematical Engineering, Stanford University, Stanford, Calif.; [2]Science Institute and Center for Systems Biology, University of Iceland, Reykjavik; and [3]Department of Management Science and Engineering, Stanford University, Stanford, Calif.

**Project Goals: Develop a convex optimization algorithm for computing thermodynamically feasible reaction fluxes in a general instance of a genome-scale integrated metabolic and macromolecular biosynthetic network.**

A steady state of a chemical reaction network is a set of chemical concentrations that remain constant for the induced reaction rates. In this work we assume the *law of mass-action* governs the rate of the reactions, i.e. the rate of a reaction is proportional to the concentrations of the participating species. More specifically, if $Y_{ij}$ is the stoichiometric coefficient for species $i$ in reaction $j$, $k_j$ is the thermodynamically feasible rate constant for reaction $j$, and $c_i$ is the concentration of species $i$, then the rate of reaction $j$ is

$$v_j = k_j \prod_i c_i^{Y_{ij}}.$$

Assuming the reactions are *mass conserving*, and that the directed graph corresponding to the set of reactions forms a strongly connected component, we show that, regardless of the rate constants, there exists at least one steady state where all concentrations are positive.

We also establish the parallel between steady states and a fixed point of a mapping that arises from solving a strictly convex optimization problem, which allows us to find such steady states in randomly constructed networks.

Research supported in part by DOE Grant DE-SC0002009
Research supported in part by NSF Grant GOALI 0800151and DOE Grant DE-SC0002009

# 112

## *Zea mays i*RS1563: A Comprehensive Genome Scale Model of Maize Metabolism

Rajib Saha* (rus184@psu.edu), Patrick F. Suthers, and **Costas D. Maranas**

Department of Chemical Engineering, The Pennsylvania State University, University Park

**Project Goals: Develop a genome-scale model for maize that meets rigorous standards on gene-protein-reaction (GPR) associations, elementally and charged balanced reactions and a biomass reaction abstracting the relative contribution of all biomass constituents.**

The scope and breadth of genome-scale metabolic reconstructions has continued to expand over the last decade. Herein, we introduce a genome-scale model for a plant with direct applications to food and bioenergy production (i.e., maize). Maize annotation is still underway which introduces significant challenges in the association of metabolic functions to genes. The developed model *Zea mays i*RS1563 (see Figure 1) is designed to meet rigorous standards on gene-protein-reaction (GPR) associations, elementally and charged balanced reactions and a biomass reaction abstracting the relative contribution of all biomass constituents. *Zea mays i*RS1563 can be viewed as a mathematically structured database of maize metabolism. The metabolic network contains 1,563 genes and 1,825 metabolites involved in 1,985 reactions from primary and secondary maize metabolism. For approximately 42% of the reactions direct literature evidence for the participation of the reaction in maize was found. As many as 445 reactions and 369 metabolites are unique to the maize model compared to the AraGEM model for *A. thaliana*. 674 metabolites and 893 reactions are present in *Zea mays i*RS1563 that are not accounted for in maize C4GEM. All reactions are elementally and charged balanced and localized into six different compartments (i.e., cytoplasm, mitochondrion, plastid, peroxisome, vacuole and extracellular). *Zea mays i*RS1563 accounts for the fact that photosynthesis in maize (i.e., a C4 plant) occurs in two separate cell types (i.e., mesophyll cell and bundle sheath cell). A biomass equation is established that quantifies the relative abundance of different constituents of dry plant cell biomass. GPR associations are also established based on the functional annotation information and homology prediction accounting for monofunctional, multifunctional and multimeric proteins, isozymes and protein complexes. We describe results from performing flux balance analysis under different physiological conditions, (i.e., photosynthesis, photorespiration and respiration) of a C4 plant and also explore model predictions against experimental observations for two naturally occurring mutants (i.e., *bm1* and *bm3*). The developed model corresponds to the largest and more complete to-date effort at cataloguing metabolism for a plant species.

By making use of high throughput enzymatic assays, proteomic and transcriptomic data across different parts of the maize plant, *Zea mays i*RS1563 could serve as the starting

point for the development of tissue-specific maize models as well as for other important $C_4$ plants such as Sorghum and switch grass. By accounting for both primary and some secondary metabolism pathways of maize, *Zea mays* *i*RS1563 can be used to explore *in silico* the effect of genetic modifications aimed at plant cell wall modification and/or starch storage on the overall metabolic state of the plant (e.g., biomass precursor availability, cofactor balancing, redox state, etc.). By taking full inventory of plant metabolism optimal gene modifications could be pursued for a variety of targets in coordination with experimental techniques. These may include (i) increase cellulose and hemicellulose production, (ii) starch yield, (iii) disruption of recalcitrant lignin subunits, and (v) nitrogen efficiency.

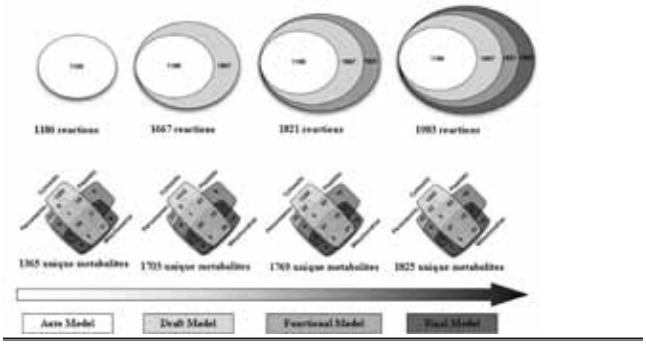


Figure 1. Successive development of *Zea mays* *i*RS1563: evolution of total number of reactions and metabolites moving from Auto, Draft, Functional and Final models

# 113

## OptForce Directed *Escherichia coli* Genetic Modifications for Maximizing Malonyl-CoA Availability

Sridhar Ranganathan[1]* (sur152@psu.edu), Peng Xu,[2] Hila Dvora,[2] Mattheos A.G. Koffas,[2] and **Costas D. Maranas**[3]

[1]Huck Institutes of Life Sciences, [2]Chemical and Biological Engineering, and [3]Department of Chemical Engineering, The Pennsylvania State University, University Park

**Project Goals: Develop an integrated computational and experimental study aimed at improving the availability of malonyl-CoA in *Escherichia coli* by deploying our OptForce methodology to predict the minimal set of genetic manipulations.**

Malonyl-CoA is an important precursor metabolite for the biosynthesis polyketides, fatty acids, biofuels (microdiesel) and plant-specific secondary metabolites. However, malonyl-CoA is directly consumed for the production of amino acids, phospholipids and biomass leaving behind only a residual amount available for overproduction targets. Metabolic engineering of microorganisms for products derived from malonyl-CoA requires achieving a fine balance between malonyl-CoA available for cellular growth and product synthesis. Efficiently harnessing malonyl-CoA continues to be one of the key barriers in the biosynthesis of long-chain biofuels ($<C_6$) and pharmaceutical compounds. Metabolic engineering efforts aimed at improving intracellular malonyl-CoA availability have so far focused on reaction steps adjacent to malonyl-CoA. Examples include overexpression of acetyl-CoA carboxylase that directly produces malonyl-CoA and elimination of competing bioconversions catalyzed by acetate kinase and alcohol dehydrogenase. Even though existing genetic manipulations have managed to significantly improve malonyl-CoA availability, numerous engineering possibilities are yet to be explored. Our group recently introduced the computational strain design procedure OptForce[1] that can hierarchically pinpoint genetic interventions that lead to yield improvements.

In this work, we present milestones achieved from an integrated computational and experimental study aimed at improving the availability of malonyl-CoA in *Escherichia coli*. We deployed our OptForce methodology to predict the minimal set of genetic manipulations in *E. coli* wild-type strain BL21 Star™ that overproduces malonyl-CoA. Using OptForce we identified a hierarchy of metabolic modifications that cooperatively redirect carbon flux towards malonyl-CoA while ensuring biomass production at pre-specified rates. Interventions predicted by OptForce can be ranked based on their quantitative impact towards achieving the overproduction target thus providing a way to prioritize the implementation of genetic interventions. In this work, up-regulations for glycolytic reactions, glyceraldehyde-3-phosphate dehydrogenase (GAPD) and phosphoglycerate kinase (PGK), pyruvate dehydrogenase (PDH), and acetyl-CoA carboxylase (ACCOAC) were predicted as the most important. All of these interventions directly contribute towards precursors of the malonyl-CoA pathway. OptForce suggested reducing the activity of TCA reactions (malate dehydrogenase (MDH), fumarase (FUM) and aconitase (ACONTa/b)) instead of eliminating them (to ensure production of all biomass components). In addition, knockouts for succinyl-CoA synthetase (SUCOAS) and propionyl-CoA:succinyl-CoA transferase (PPCSCT) were predicted that reduce the drain of malonyl-CoA towards by-products. The complete set of engineering interventions and alternatives suggested by OptForce can be represented in the form of a logic decision tree (see Figure 1).

By successively implementing the hierarchy of OptForce suggestions, we have successfully constructed a recombinant strain of *E. coli* that exhibits improved malonyl-CoA levels. We demonstrate the efficacy of this strain by incorporating a set of heterologous pathways that use three moles of malonyl-CoA for the production of naringenin. Naringenin is a flavanone that is a low-molecular weight plant-specific polyphenolic compound. Upon translating the OptForce predictions to the gene level using the gene-protein-reaction (GPR) associations for these reactions, we implemented genetic modifications (see Figure 1) for overproducing naringenin in *E. coli*. The evolution of naringenin yield as more interventions are accumulated highlights the synergistic effect of combining beneficial mutants (Δ*fumC* and Δ*sucC*) and overexpression targets (*acc*, *pgk*, *gapA* and *pdh*) predicted

by OptForce. Specifically, a titer of 474 mg/L of naringenin production was observed which is up-to-date the highest yield achieved in a lab-scale fermentation process.

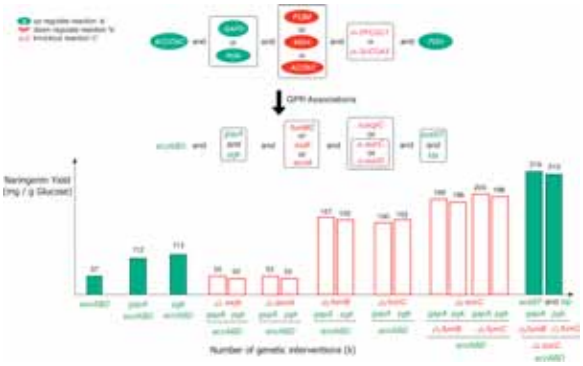


Figure 1. Reaction interventions predicted by OptForce procedure and the corresponding gene associations for the overproduction of malonyl-CoA and naringenin synthesis.

**Reference**

1. Ranganathan, S., Suthers, P. F., Maranas, C. D., OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol* 2010, 6, e1000744.

Student Oral Presentation–Monday

# 114

## MetRxn: Reaction and Metabolite Standardization and Congruency across Databases and Genome-Scale Metabolic Models

Akhil Kumar,[2]* Patrick Suthers,[1] and **Costas D. Maranas**[1]

[1]Department of Chemical Engineering, and [2]Department of Computer Science and Engineering, The Pennsylvania State University, University Park

**Project Goals: Create a knowledgebase for biochemical transformations with standardized metabolite and reaction entries that encompasses existing databases and all publicly available genome-scale metabolic models.**

The ever-accelerating pace of DNA sequencing and annotation information generation is spearheading the global inventorying of metabolic functions across all kingdoms of life. Increasingly, metabolite and reaction information is organized in the form of community, organism, or even tissue-specific genome-scale metabolic reconstructions. These reconstructions account for reaction stoichiometry and directionality, gene to protein to reaction associations, organelle reaction localization, transporter information, transcriptional regulation and biomass composition. Already over 35 genome-scale models are available for eukaryotic, prokaryotic and archaeal species and are becoming indispensable for computationally driving engineering interventions in microbial strains for targeted overproductions, elucidating the organizing principles of metabolism and even pinpointing drug targets. A key barrier to the pace of extraction of metabolic knowledge from data is our inability to directly make use of metabolite/reaction information from databases (e.g., BRENDA, KEGG, BioCyc, UM-BBD, PubChem, ChEBI, Reactome.org, Rhea, etc.) or other metabolic models due to incompatibilities of representation, duplications and errors. Therefore, the inadvertent inclusion of multiple replicates of the same metabolite, stoichiometrically inconsistent and/or elementally/charge unbalanced reactions can lead to erroneous model predictions and missed opportunities to reveal (synthetic) lethal gene deletions, repair network gaps and quantify metabolic flows. There have already been a number of efforts aimed at addressing some of these limitations. The Rhea database aggregates reaction data primarily from IntEnz and ENZYME whereas Reactome.org is a collection of reactions primarily focused on human metabolism. Research towards integrating genome-scale metabolic models with large databases has so far been even more limited. An important step forward is Model SEED which is a web resource that generates draft genome-scale metabolic models drawing from an internal database that integrates KEGG with 13 genome scale models (including six of the models in the BiGG database). Motivated by this challenge we recently carried out an initial construction of the web-based resource MetRxn that integrates, using internally consistent descriptions, metabolite and reaction information from 6 databases and 34 metabolic models. The MetRxn content generation follows the general steps outlined in Figure 1. Metabolite and reaction data was first downloaded from BRENDA, KEGG and BioCyc using a variety of methods based on protocols such as SOAP, FTP and HTTP. We subsequently pre-processed the data into flat files that were imported into MetRxn. All original information pertaining to metabolite name, abbreviations, metabolite geometry, related reactions, catalyzing enzyme and organism name, gene-protein-reaction associations, and compartmentalization was retained. For all 34 genome-scale models ancillary information culled form the corresponding publications was also imported. The "raw data" from both databases and models was unified using standard SQL scripts on a MySQL server. We used Marvin (Chemaxon) to analyze all 231,085 raw metabolite entries containing structural information (out of a total of 322,936 entries). Metabolite atom bond connectivity was calculated at a fixed pH of 7.2 and converted into standard Isomeric SMILES format. Metabolites were also annotated with Canonical SMILES using the OpenBabel Interface from Chemspider. Metabolites with missing structural information were revisited during the reaction reconciliation step. After generating the initial metabolite associations, we identified reaction overlaps using the reaction synonyms and reaction strings along with the metabolite SMILES representations. During this step, reactions were flagged as single-compartment or two-compartment (i.e., transport reactions). Using the corrected metabolite elemental composition and protonation states, reactions are evaluated for charge and elementally balance. We used a linear optimization program to charge and elementally balance all reactions. Currently, the MetRxn knowledgebase is stored on the Open Source MySQL database system using the MyISAM storage engine.

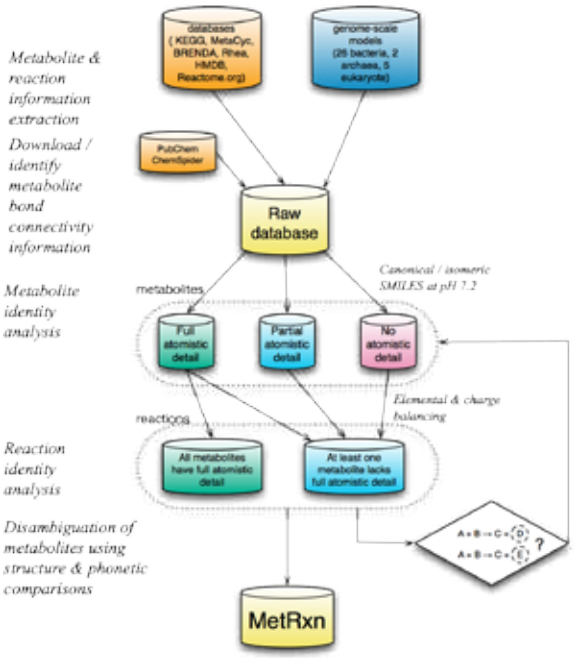



Figure 1. Flowchart abstracting the iterative workflow implemented in constructing the first-phase of the MetRxn knowledgebase. MetRxn contains, so far, over 62,345 distinct metabolites and 56,142 reactions that are charge and elementally balanced.

# 115

## Improving the *i*MM904 *S. Cerevisiae* Metabolic Model Using Essentiality and Synthetic Lethality Data

Ali R. Zomorrodi[1]* (zomorrodi@engr.psu.edu) and **Costas D. Maranas**[1]

[1]Department of Chemical Engineering, the Pennsylvania State University, University Park

**Project Goals: Use gene essentiality and synthetic lethality predictions to improve genome-scale models of metabolism.**

*Saccharomyces cerevisiae* is the first eukaryotic organism for which a multi-compartment genome-scale metabolic model was constructed. Since then a sequence of improved metabolic reconstructions for yeast has been introduced. These metabolic models have been extensively used to elucidate the organizational principles of yeast metabolism and drive yeast strain engineering strategies for targeted overproductions. They have also served as a starting point and a benchmark for the reconstruction of genome-scale metabolic models for other eukaryotic organisms. Despite the successive improvements in the details of the described metabolic processes, even the recent yeast models (e.g., *i*MM904 and Yeast 4.0) remain significantly less predictive than the latest *E. coli* model (i.e., *i*AF1260). This is manifested by its significantly lower specificity in predicting the outcome of grow/no grow experiments in comparison to the microbial models. Contrasting the predicted growth phenotype of single mutant strains with the available experimental data under various growth conditions is the established standard for testing the accuracy of genome-scale metabolic models. These comparisons result in four different outcomes: GG or NGNG when both model and experimental data either imply growth (G) or no growth (NG) for the mutant strain, NGG when the model predicts that the gene deletion is lethal but the experiment shows that it is viable, and finally GNG when the model predicts that the mutant strain would be viable but *in vivo* observations show a lethal effect. Our group recently introduced a mathematical procedure termed GrowMatch for reconciling both NGG and GNG growth prediction inconsistencies across different substrates.

Here, we demonstrate that additional layers of correction and improvement can be gleaned by making use of synthetic gene lethal information. As shown in Figure 1, comparisons of model predicted synthetic lethal interactions with available experimental data reveal a number of additional ways that model and experiment may disagree. Notably, the "no growth" phenotype in this case could be due to either essentiality (ES) or synthetic lethality (SL) of the gene deletions. For example, GES and GSL inconsistencies refer to cases where the *in silico* deletion of a gene pair is not lethal (i.e., *G*rowth) but *in vivo* they are lethal due to gene essentiality or synthetic lethality (i.e., *ES*sential or *S*ynthetic *L*ethal). Similarly, ESG and ESSL represent mismatches where the single deletion of one of the genes *in silico* is lethal (i.e., *ES*sential), however, their simultaneous deletion *in vivo* results in either a viable strain (i.e., *G*rowth) or a lethal phenotype (i.e., *S*ynthetic *L*ethal), respectively. Finally, SLG and SLES denote inconsistencies where the model implies that only the double gene mutation is lethal (i.e., *S*ynthetic *L*ethal) but experimental observations support either growth (G) or lethality of any of the two single gene deletions (i.e., *ES*sential), respectively.

In this project we make use of the automated GrowMatch procedure for restoring consistency with single gene deletion experiments in yeast and extend the procedure to make use of synthetic lethality data using the genome-scale model *i*MM904 as a basis. In addition to essentiality and synthetic lethality we also explored disagreements in auxotrophy complementation, where model predicted supplementation rescue (i.e., auxotrophy) scenarios are inconsistent with experimental data. Overall, we identified and vetted using literature sources 90 distinct model modifications along with 30 regulatory constraints for minimal and YP media. The incorporation of the suggested modifications led to a substantial increase in the fraction of correctly predicted lethal knockouts (i.e., specificity) from 38.84% (87 out of 224) to 53.57% (120 out of 224) for the minimal medium and from 24.73% (45 out of 182) to 40.11% (73 out of 182) for the YP medium. Synthetic lethality predictions improved from 12.03% (16 out of 133) to 23.31% (31 out of 133) for the minimal medium and from 6.96% (8 out of 115) to 13.04% (15 out of 115) for the YP medium. Given that these improvements in the model were achieved using only

the partial list of synthetic lethality data currently available in literature, a far larger contribution of synthetic lethals in providing model refinement strategies is expected as more synthetic lethality data are becoming available. Overall, this study provides a roadmap for the computationally driven correction of multi-compartment genome-scale metabolic models and demonstrates the value of synthetic lethals as curation agents.

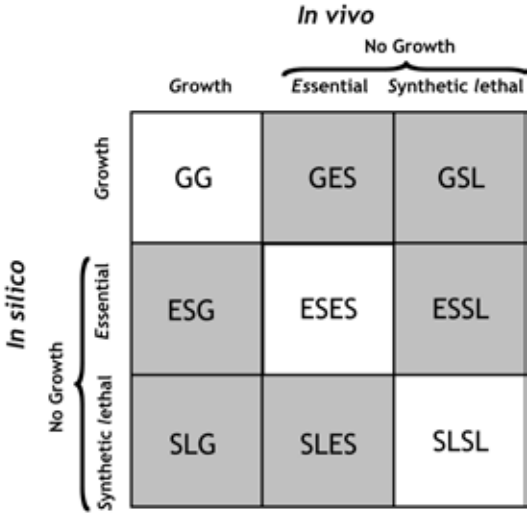



Figure 1. Different types of mismatches between *in silico* predictions and *in vivo* observations for double gene perturbations. The abbreviations G, ES and SL in this figure refer to ***Growth***, ***Essential*** and ***Synthetic Lethal***, respectively. Here, 'No Growth' can be due to either essentiality or synthetic lethality of single or double gene deletions.

submitted post-press

# 115A‡

## Construction of an *E. coli* Genome-Scale Atom Mapping Model for MFA

Prabhasa Ravikirthi[1]* (pxr173@psu.edu), Patrick F. Suthers,[2] and **Costas D. Maranas**[2]

[1]Department of Cell and Developmental Biology and [2]Department of Chemical Engineering, The Pennsylvania State University, University Park

**Project Goals: To generate a genome-scale atom mapping model of *E. coli* for metabolic flux analysis.**

Metabolic flux analysis (MFA) has so far been restricted to lumped networks lacking many important pathways, partly due to the difficulty in automatically generating isotope mapping matrices for genome-scale metabolic networks. Here we describe a procedure that uses a compound matching algorithm based on the graph theoretical concept of pattern recognition along with relevant reaction information to automatically generate genome-scale atom mappings which trace the path of atoms from reactants to products for every reaction. The procedure is applied to the *i*AF1260 metabolic reconstruction of *Escherichia coli* yielding the genome-scale *i*sotope *m*apping model imPR90068. This model maps 90,068 non-hydrogen atoms that span all 2,077 reactions present in *i*AF1260 and contains a total of 1.37 x $10^{157}$ isotopomers (with 8.34 x $10^{93}$ $^{13}$C isotopomers).

The isotope mapping model imPR90068 contains mappings for reactions that were previously lumped or completely absent from earlier isotope mapping models (even in imPS1485 (Suthers et al. 2007)). These new additions include 68 reactions involved in the metabolism of amino acids (see Figure 1), 65 reactions involved in central metabolism, 153 reactions in nucleotide biosynthesis and salvage pathways, 225 reactions in glycerophospholipid metabolism, 160 reactions in cofactor and prosthetic group biosynthesis and 181 reactions in alternate carbon metabolism. The incorporation of more than 1,100 new reactions involved in various parts of *E. coli* metabolism together with the inclusion of more than 800 metabolites compared to the previous largest imPR1485 model implies that alternate metabolic routes could now fully be taken into account during flux elucidation using MFA. Additionally, it allows for the possible labeling of substrates other than glucose.

This paper also introduced the computational infrastructure for tracing all atoms present in every reaction in the *i*AF1260 metabolic reconstruction of *E. coli* from reactants to products to create a genome-scale mapping database. This automated procedure can be efficiently leveraged for genome-scale models of other organisms to create isotope mapping databases. Common reactions already present in *i*AF1260 can be directly culled from the imPR90068 reaction-mappings database thus significantly reducing the effort needed to construct other organism-specific mapping models. The potential to improve our understanding of flux allocation in different organisms is alluded by the gap in the size of genome scale vs. isotope mapping models. For example, there exists a 50-fold difference in the size of the genome-scale reconstruction of *Bacilus subtilis* that spans 1,020 reactions (Oh et al. 2007) and its current isotope mapping model (Dauner et al. 2001) that accounts for only 25 reactions (all from central metabolism). Approximately 70% of reactions in its genome-scale model have an exact match to reactions in *i*AF1260. It is expected that incorporating reactions into the mapping model already present in the genome-scale model could shed light onto metabolic pathway usage patterns. We can do the same for organisms with multiple compartments such as *Saccharomyces cerevisiae*.

However, the ability to elucidate fluxes using the full complement of reactions and metabolites present in genome-scale level reconstructions comes at the expense of requiring additional labeling data. While lumped isotope models (Antoniewicz et al. 2007b; Kim et al. 2008; Suthers et al. 2007) typically require the analysis of spectra (i.e., NMR or GC/MS) for only about 20-50 fragments, using the totality of mapped isotopomers in imPR90068 will require significantly higher numbers of carefully chosen labeled fragments. This makes even more pertinent the use of methods such as OptMeas (Chang et al. 2008; Suthers et

al. 2010), EMU representations (Antoniewicz et al. 2007a) and systematic reaction step aggregation techniques (e.g., SLIPs (Quek 2009)), as well as advances in metabolomics.

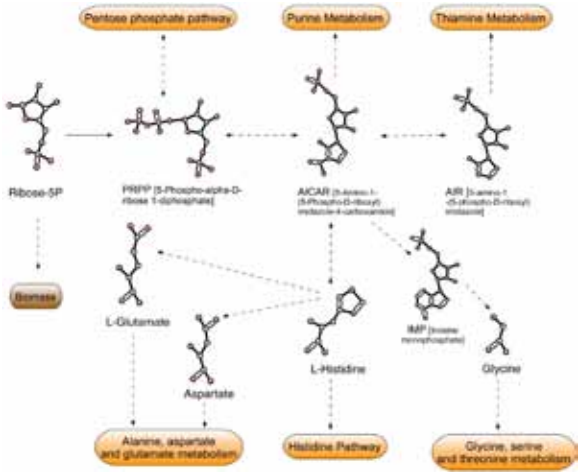



Figure 1. An example of the expanded scope of the genome-scale isotope mapping model imPR90068. In imPS1485 Ribose-5P production was directly routed to biomass as a stand-in substitute for histidine. In imPR90068 R5P downstream conversion is linked to other amino acid synthesis pathways.

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

# 116

## Microbioreactor Technology for Obligate Anaerobes

**Harry Lee** (harrylee@pharyx.com) and Paolo Boccazzi* (boccazzi@pharyx.com)

Pharyx, Inc., Boston, Mass.

http://www.pharyx.com

**Project Goals: see below**

Anaerobic microorganisms have evolved biochemical pathways that can be exploited for industrial applications. These include the ability to breakdown environmental pollutants for bioremediation, the breakdown of cellulose into simple sugars for biofuels, and the production of specialty chemicals. However, there remains a tremendous challenge to the scale-up of bioenzymatic activities to industrial processes. While systems biology approaches and metabolic engineering promise to contribute to our understanding of these systems, a key bottleneck is in conducting controlled experiments to ground these approaches with high quality data. Thus far, experiments are frustrated by the laborious set-up and operation of stirred tank bioreactor systems, which for anaerobic microbiology is further encumbered by the requirement of an anaerobic environment. The absence of easy to use systems also holds back more traditional microbiology approaches such as mutagenesis and screening and directed evolution.

We are developing a parallel bioreactor system, based on microfluidic integration technology and disposable microbioreactor modules, with application specific customizations for anaerobic fermentation. These customizations are aimed to enable up to 32 simultaneous anaerobic fermentations under controlled conditions, with online monitoring of growth kinetics and other phenotypes such as enzyme activity. A unique feature of this system is the ability to operate it in ambient air through careful inoculation port and reactor and control module design, or to operate it within an anaerobic bag, taking advantage of its compact size.

**Project goals are 1) determine microbioreactor designs that will support anaerobic inoculation and fermentation, 2) identify optimal materials for fabricating anaerobic bioreactors, 3) determine the range of process parameters where microbioreactor data corresponds to serum tubes and stirred tank fermentors, 4) monitor enzyme activity on-line.**

In our initial work, we demonstrated passive anaerobic fermentations in simple microbioreactor devices. These experiments highlighted problems including gas bubble generation by anaerobes, which spoiled optical density measurements, and the significance of absorbed oxygen in plastic materials. This motivated a new anaerobic device design with anaerobic mixing, bubble removal, and fluid injection to enable pH controlled anerobic fermentations. Figure 1 shows bubble-free online measured optical density in the microbioreactor and at line measured samples from a stirred tank bioreactor for uncontrolled *C. acetobutylicum* fermentations. Figure 2 shows a comparison between controlled and uncontrolled fermentations of *B. fibrisolvens* in both a stirred tank and the microbioreactor.

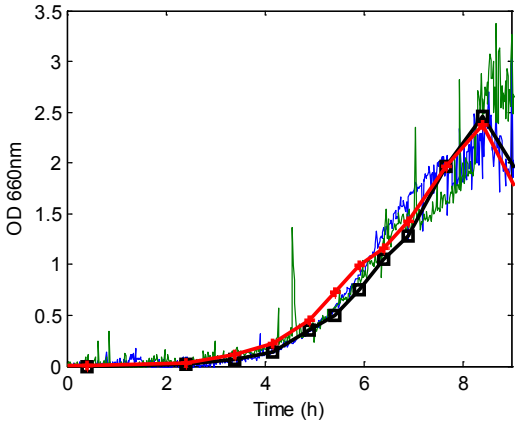



Figure 1. Microbioreactor (solid lines) and bench top stirred tank (+ and square markers) fermentations of *C. acetobutylicum*. With time alignment to account for varying lag phase, and calibration between microbioreactor and spectrophotometer optical density, there is excellent correspondence between both fermentation systems.

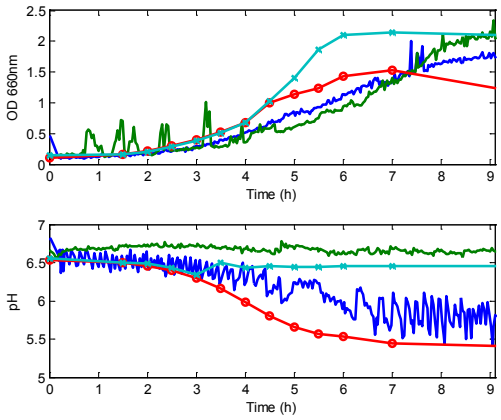



Figure 2. Optical density and pH curves for Microbioreactor (solid lines) and bench top stirred tank bioreactor (lines with x and o markers) of *B. fibrisolvens*. In the stirred tank, there was a clear effect of pH control, which maintained a steady growth rate to the end of the fermentation while the uncontrolled case showed a slower growth rate after 4.5 hours. Similar behavior was observed in the microbioreactor.

# 117

## Semantic Indexing of the Green Technology Patent Literature

**George M. Garrity*** (garrity@msu.edu), Charles T. Parker, and Catherine Lyons

NamesforLife, LLC and Michigan State University

**Project Goals: NamesforLife, LLC has developed a novel technology that resolves uncertainty about the meaning of biological names or other dynamic terminologies. It uses those terms to create persistent links to related information, goods, and services available on the Internet, even if the terms have changed.**

Under a Phase I/II STTR, NamesforLife, LLC created a suite of software tools and techniques to manage dynamic terminologies and an underlying term set (an up-to-date list of over 14,000 validly published names of bacteria and archaea, including all of the synonyms and homonyms, links to appropriate taxonomic literature, key genetic and genomic data). The company's N4L tools can automatically detect and tag bacterial names in HTML and XML documents with a high degree of precision. An interactive browser-based application (N4LGuide) provides end users direct access to correct nomenclatural information along with links to key data (16S sequence and genome sequence data) while reading the literature. It uses ISO standard Digital Object Identifier (DOI) technology to create links at each occurrence of a validly published name in HTML documents. The company has also developed batch tools (the N4L Semantic Tagger) that can embed N4L-DOIs into XML versions of scientific articles that are created as part of the contemporary publishing workflow and used to create human readable content in various forms (e.g., HTML, PDF, ink-on-paper). The company has also developed a unique way of tracking the occurrence of biological names in the literature, based on the usage of our tools (the N4L Contextual Index).

While initially intended as a tool for readers, authors, and publishers of scientific literature, N4L tools can also be applied to other documents where bacterial names appear. As proof of principle, the company processed approximately 250,000 U.S. patents and patent applications with the Semantic Tagger and then indexed the tagged documents using Apache Lucene to provide end users with additional search and retrieval capabilities. Simple graphical tools were added to support limited on-demand analyses of search results. These tools are designed to support data mining by non-commercial organizations, highlighting trends in commercialization of biodiversity research. This work also led to the discovery of "terminological fingerprints" that could be used to classify patents and other documents using externally managed term sets.

To validate the concept of "terminological fingerprinting", the company processed the EPO Green Technology collection of patents, which consists of approximately 362,000 documents. In addition to detecting bacterial names, the N4L Semantic Tagger was modified to recover patent classifications (IPC and ECLA), applicants, assignees, inventors, patent references, patent metadata, and patent titles, as is common in patent landscaping.

A total of 3,845 patents were found that made reference to 3,385 distinct bacterial and archaeal names held in the NamesforLife database. Of these, 626 names were unique to non-U.S. patents. The number of names per patent (name vectors) ranged from 1 – 1,290, with an average of 13 names and a median of 5 names. In addition to name occurrence,

frequency data for each name occurrence per patent was tabulated. The resulting name vectors were then used to further examine the associations among the patents based on the IPC and ECLA classifications. Simple associations could be derived directly from the captured data. However, more complex patterns involving multiple many-to-many relationships could only be ascertained from the cross-products of underlying contingency and frequency data.

The results were then examined using routine approaches for exploratory data analysis and visualization (e.g., principal components analysis, robust clustering, 2D scatter plots, 3D spin plots and heatmaps). Each of these methods revealed strong evidence of terminological fingerprints in the patents. However, those methods did not scale well or suffered from other limitations. Hexagonal binning was, however, found to be suitable for visualizing the complex relationships inherent in the patent data. The company is currently developing interactive hexagonal bin plots as a means of selecting subsets of patents that involve related technologies and microorganisms.

As DOE research on biofuels, bioremediation and carbon sequestration moves from the laboratory into production or commercial environments, a number of important policy and business decisions must be made that demand correct information. These include establishing the patentability of a given technology, freedom to operate, and potential infringement of patents held by competitors, both in the U.S. and abroad. Failure to pay careful attention to these issues can have serious consequences beyond the payment of stiff penalties for infringement. These include lost opportunities arising for technology licensing, failure to detect and understand regional disparities, rapid growth in patent coverage of technologies by competitors and migration of technology across international borders. The scientific and technical literature provides an incomplete view of any field having commercial potential because the underlying technologies are typically not revealed in public until absolutely necessary, and then only after patent applications have been filed. While patents with corresponding papers are not uncommon as a means of announcing important new developments, they are not obligatory. Therefore, an awareness of developments in the field requires a thorough review of both bodies of literature. NamesforLife is building tools to simplify such searches, using its proven approach to indexing through the creation of persistent links to externally managed terminologies that common to both bodies of literature. This approach integrates well with existing commercial, academic and USPTO data mining capabilities.

This research is supported by the Office of Biological and Environmental Research of the U.S. Department of Energy under Phase I and II STTR Awards DE-FG02-07ER86321 A001 - A005.

# 118

## Accelerating Metagenomics Using Graphics Processing Units

**Matthew Hudson*** (mhudson@illinois.edu)

University of Illinois

**Project Goals: Development of faster database searching algorithms for high throughput sequencing.**

The goal of this project is to develop software for DNA sequence processing from DOE projects that runs on the graphics processing units available in modern computers. In particular, the aim is to accelerate the matching of sequence reads from experiments using high-throughput next-generation sequencing platforms (such as metagenomics projects) to large databases such as those maintained at GenBank. The most critical need is for a way to compare hundreds of millions of DNA reads to a protein database using the translated BLAST algorithm BLASTX. Using the existing versions of the BLAST software, searching one read against a complete protein database such as GenBank non-redundant protein (NR) takes many seconds to a few minutes on one CPU. One sequence run from the latest sequencing platforms (e.g. Illumina HiSeq) can produce hundreds of millions of sequence reads. Searching this data against NR using conventional BLAST thus has an unacceptably high computational cost of millions of CPU hours, the cost of analysis vastly exceeding the cost of generating the data. In this project we have developed MulticoreWare BLASTX (MBLASTX). This new-generation software package employs algorithm-level acceleration and the use of the GPU (the powerful parallel graphics processors found on most modern computers) to accelerate sequence searches more than 1,000 times compared to the performance of the latest version of NCBI BLAST on the latest computers. The accuracy of the search results is over 99% compared to NCBI BLAST. This software has accelerated the processing of DNA sequence enough to replace expensive and energy-hungry supercomputers used for this purpose with ordinary, much cheaper computers, lowering the cost and energy consumption of an essential analysis step by more than 1,000 fold.

# Joint USDA-DOE Plant Feedstock Genomics for Bioenergy

## 119

### Systems View of Root Hair Response to Abiotic Stress

Principal Investigator: **Gary Stacey[1]*** (staceyg@missouri.edu), **Co-Principal Investigators:** Jianlin Cheng,[2] Dong Xu,[2] David Koppenaal,[3] and Ljiljana Paša-Tolić[3]

[1]Center for Sustainable Energy, [2]University of Missouri, Columbia; [3]EMSL, Pacific Northwest National Laboratory, Richland, Wash.

http://staceylab.missouri.edu/
http://soyroothair.org/

**Project Goals:**

1. **Analyze the transcriptional response of root hair cells under conditions of abiotic stress (i.e.,heat and drought stress)**
2. **Analyze proteomic and metabolomic changes in root hair cells under conditions of abiotic stress (i.e., heat and drought stress)**
3. **Utilize the tools and information developed in Objectives 1-2 to develop descriptive models of root hair metabolic function.**

Over the last 100 years, the atmospheric concentration of carbon dioxide has dramatically increased, in major part due to the burning of fossil fuels, recent rapid industrialization, and land use changes. The predicted effects of continued climate change are complex but include effects on air and surface temperature, with coincident effects on water availability. Soil temperature can influence root growth, cell elongation, root length and extension, initiation of new lateral roots and root hairs, and root branching. These effects are likely manifestations of the variety of physiological effects brought about by temperature on plant roots; including changes in root respiration, nutrient uptake, as well as physicochemical effects on the soil environment (e.g., changes in nitrogen mineralization). Ambient temperature changes also affects other parts of the plant (e.g., photosynthetic rates), which also affects below ground growth and physiology. When we include in this discussion issues of plant genetic variation, as well as the effects of temperature on water availability, the full complexity of the effects of climate change on the plant root environment becomes clear.

In order to properly understand the effects of climate change, models must be developed that predict impacts across broad temporal (seconds to millennia) and spatial (microns to global) scales. In order to be useful, these models must draw upon accurate experimental data. Systems biology seeks to address these needs by providing a comprehensive, quantitative analysis of the manner in which all the components of a biological system interact functionally over time and space. The recent explosion of interest in systems biology is the result of the development of new tools for system-level analysis of cellular function and the availability of an increasing number of full genome sequences, which enables the full application of these new technologies. The ultimate goal is a new, predictive view of biological function, supplanting the older descriptive understanding. Hence, there is a need to integrate system approaches to understand the effects of climate change on molecules, cells, organisms and ecosystems.

However, the promise of this new 'predictive' science has yet to achieve its full potential. A number of challenges remain. For example, although the new tools do indeed provide for a full systems view of cellular function, integration of dissimilar data (e.g., proteomics, metabolomics, transcriptomics, etc.) remains a formidable challenge. Among the issues compounding the problems of data integration is the issue of "signal dilution", which results from the fact that most studies average the response of whole tissues, obscuring the actual cellular response. Hence, it is impossible to discern the difference, for example, of a gene that is expressed at a low level in all cells from a gene that is expressed at a very high level, but only in a few cells. Approaches are needed to conduct functional genomics on single cells.

This proposal will address the question of signal dilution by focusing, specifically, on soybean root hair cells, which represent a single, differentiated cell type. Over the past 5 years, we have established the soybean root hair cell as an excellent platform for plant systems biology studies. It is now arguably the best characterized cell type in plant biology, as exemplified by our various publications, databases and additional information yet to be published (see Libault et al., 2010).

Our vision is to utilize the soybean root hair system to explore, at a systems level, the biology of a single, differentiated plant cell type, while gaining novel insight into the impacts of temperature and water availability on a crucial root cell necessary for nutrient uptake. The proposed research should provide unambiguous measurements of the impact of these environmental factors on plant cell function, without the compounding effects of tissue dilution. The proposed research will focus on defining the transcriptional, metabolomic and proteomic response of the soybean root hair cell to variations in temperature and water availability. These data, in addition to other data available in our laboratory, will allow the development of computational models to examine regulatory networks that function at a single cell level to control the response to environmental change. The data obtained should provide a better understanding of the impacts of climate change (heat and water limitation) on plant root physiology.

The research team brings a wealth of experience and knowledge to the project, which is a collaboration between scientists at the University of Missouri and at the Environmental Molecular Sciences Laboratory at the DOE Pacific Northwest National Laboratory. The latter facility brings tremendous experience and instrumentation to conduct the metabolomic and proteomic studies described. In addition to the expected research outcomes, the project will also provide training for graduate and postdoctoral students to prepare them for their future careers.

**Reference**

1. Libault, Marc, Brechenmacher, Laurent, Cheng, Jianlin, Xu, Dong, Stacey, Gary (2010) Root hair systems biology. *Trends in Plant Science* 15: 641-650.

# 120

## Night-Time Stomatal Conductance and Transpiration Negatively Impact Biomass Accumulation

C.E. Jahn,[1] L. Derose-Wilson,[1] **J.K. McKay**,[1] **D.R. Bush**,[1] **H. Leung**,[2] and **J.E. Leach**[1]* (Jan.Leach@colostate.edu)

[1]Program in Plant Molecular Biology, Colorado State University, Fort Collins; and [2]Plant Breeding, Genetics and Biotechnology Division, International Rice Research Institute (IRRI), Metro Manila, Philippines

**Project Goals: Our goal is to provide the applied biomass research community and industry with information to allow exploitation of the genes and pathways relevant to biomass accumulation in grasses. The specific objectives for this project are to: Objective 1: Identify genes involved in biomass accumulation. Using genetic populations for rice lines that exhibit extremes in biomass accumulation, we will (a) map QTL using a comprehensive phenotyping/genotyping approach, (b) confirm the QTL location and phenotypic effect and (c) screen a deletion mutant collection to identify large deleted regions corresponding to biomass accumulation. Objective 2: Dissect the QTL using an integrated analysis. First, we will use an expedited approach to generating near isogenic lines containing each QTL. Second, we will integrate genome-wide data to refine the QTL regions including (a) association mapping, (b) expression profiling, (c) mutant analysis, (d) fine scale mapping, and (e) comprehensive sequencing.**

Plant biomass accumulation is the culmination of many processes: carbon assimilation via photosynthesis, carbon losses due to respiration, efficiency of biosynthetic pathways, long-distance transport of assimilates, water balance, and mineral nutrition. Previously, we identified a negative correlation between total biomass and photosynthetic rates in diverse rice varieties (Jahn et al 2011). Although counterintuitive, this result is not unique to rice; strong negative correlations have been observed between photosynthetic rates and biomass accumulation in other crop and weed species. However, to date, there is no mechanistic explanation of this phenomenon. We took detailed physiological measures on a well replicated growth chamber experiment of 20 diverse rice varieties known to differ in biomass. We measured gas exchange in both day and night, chlorophyll fluorescence in light- and dark-adapted plants, and chlorophyll content. These data revealed significant genetic variation among the 20 lines in all traits measured. Trait variation was largely explained by genotype and breeding history (advanced vs landrace) but not by varietal groupings (indica, japonica, aus). Total biomass was negatively correlated with night-time stomatal conductance, transpiration and dark respiration. The 20 lines varied as much as 5.7-fold for night-time stomatal conductance. Among varieties, night-time transpiration rates were between 4 and 26% that of day-time and thus represent a substantial fraction of total daily water loss for some lines. Although little effort to date has been placed on directly improving photosynthesis, new initiatives to improve photosynthetic rates are considered a frontier for grain and biomass yields. However, our work indicates that simply increasing day-time carbon assimilation rates may not be sufficient. A more comprehensive approach will also target carbon metabolism, respiration and water losses due to night-time transpiration.

# 121

## The Hunt for Green Every April: Phenotypic and Metabolomic Analysis of Nutrient Remobilization in Switchgrass

Gautam Sarath[1]* (Gautam.Sarath@ars.usda.gov), Nathan A. Palmer,[1] Aaron J. Saathoff,[1] Christian M. Tobias,[2] Soundararajan Madhavan,[3] Paul Twigg,[4] Kenneth P. Vogel,[1] and Jaehyoung Kim[5]

[1]GFBRU, USDA-ARS, East Campus, UNL, Lincoln; [2]GGDRU, WRRC, USDA-ARS, Albany, Calif.; [3]Department of Biochemistry, University of Nebraska, Lincoln; [4]Biology Department, University of Nebraska, Kearney; and [5]Core for Applied Genomics and Ecology, University of Nebraska, Lincoln, East Campus

**Project Goals: Our interrelated project objectives are: (A) elucidation of the regulation of gene networks, proteins and metabolites for manipulation of plant feedstocks for improved productivity and sustainability, and improved water use efficiency and nutrient utilization; (B) Elucidation of the regulation of gene networks, proteins and metabolites for advanced understanding of carbon partitioning and nutrient cycling in plant feedstocks; and (C) comparative approaches to enhance fundamental knowledge of the structure, function, and organization of plant genomes leading to innovative strategies for feedstock characterization, breeding or manipulation. Our transcriptomic and marker studies will add to existing genomic resources available for this species. The proposed C and N recycling studies in diverse switchgrass genotypes will significantly improve our knowledge in this area, and has direct bearing on plant fitness and sustainability issues of other perennial grasses being developed for use as bioenergy. Genomic and physiological insights**

**obtained during this project will be utilized to better develop molecular markers using a defined population of switchgrass plants that will be exploited to characterize patterns of nucleotide diversity reflected through the processes of mutation, recombination, genetic drift and-directional selection.**

This DOE-USDA funded research seeks to understand traits controlling winter hardiness in switchgrass (*Panicum virgatum* L.). We have planted five different populations with contrasting winter survival in replicated field plots. During the 2010 growing season, one group of replicates was labeled with $^{13}C$ to follow the label in aerial tissues over the growing season. In other experiments, crowns and rhizomes were harvested at boot, anthesis, mid-seed fill, late senescence, post-frost and early green up (2011). These harvest dates were based on stages of development of the cold-adapted cultivar "Summer." Crowns and rhizomes will be extracted for 454 and Illumina sequencing, metabolite analyses and other related physiological studies. Other objectives of this grant are (1) to use next-generation sequencing to query transcript abundance (levels of gene expression) in specific populations of switchgrass plants during regreening and dormancy; and (2) to study the genetic variation (extent of linkage disequilibrium in populations) using over 2000 plants from various genetic backgrounds that have been planted in the field for these analyses. In order to build a bioinformatic pipeline, we have recently acquired and analyzed ~1 million sequences obtained from pre-frost Summer crowns using the 454 platform. Data from this and related experiments will be presented.

DOE-USDA funded research (2010-2012; Office of Science (BER), U. S. Department of Energy grant number DE-AI02-09ER64829)

# 122

## Functional Analysis of Regulatory Networks Linking Shoot Maturation, Stem Carbon Partitioning, and Nutrient Utilization in Sorghum

**Stephen Moose**[1]* (University of Illinois), Pat Brown,[1] and Max Moehs[2]

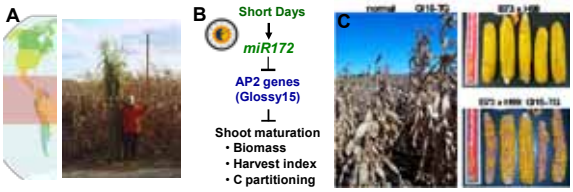
[1]University of Illinois and [2]Arcadia Biosciences

http://cropsci.illinois.edu/faculty/moose/lab/energy.html

**Project Goals: The phenotypes conditioned by the Gl15-OX transgene in maize suggest simple molecular breeding strategies to modulate shoot maturation and enhance sustainable biomass yields of bioenergy grasses. The goals of this project are to increase our understanding of shoot maturation pathway genes in sorghum and test their utility in improving sorghum as a bioenergy feedstock. We are employing three complementary experimental approaches:**

1. **Characterize allelic and expression variation for *miR172*, *AP2*, *SPL* and *miR156* genes within a core set of 12 diverse sorghum genotypes representing grain, sweet, and forage types.**
2. **Allelic variation in shoot maturation pathway genes will be tested for associations with phenotypic variation for flowering time, biomass yields, photoperiod-sensitivity, and stem sugar production in a panel of 500 diverse sorghum genotypes.**
3. **Transgenic sorghum lines with altered expression of *Glossy15* or *miR172* will be generated and evaluated for beneficial changes in flowering time, biomass yields, and nitrogen utilization. These transgenic lines will also be used to identify changes in the expression of down-stream target genes via Illumina RNA tag profiling.**

Phenotypic traits that are important to increasing sustainable biomass production and conversion efficiency from bioenergy crops include growth habit (perennial versus annual), flowering time, vegetative senescence, tillering, carbon partitioning and cell wall composition, and nutrient use efficiency. Each of these traits are impacted by the developmental process of shoot maturation or phase change, where shoot meristems and lateral organs progress through embryonic, vegetative, and reproductive growth. Studies conducted in both *Arabidopsis* and maize have identified a conserved regulatory network of two antagonistically acting microRNAs (*miR156* and *miR172*) and their target transcription factors (SPL and AP2 proteins) that control shoot maturation and phase change. Mutations and transgenic lines that alter the relative activities of *miR156*, *miR172*, *SPL* and *AP2* genes condition changesin growth habit, tillering, cell wall composition, response to photoperiod, flowering time, fertility, and seed size.




| D | Grain hybrids | | Tropical hybrids | | Gl15-OX hybrids | |
|---|---|---|---|---|---|---|
| Yields (Mg/ha) | -N | +N | -N | +N | -N | +N |
| Total biomass | 14.8 | 18.9 | 20.0 | 22.0 | 18.0 | 20.5 |
| Grain | 7.5 | 10.8 | 1.3 | 2.6 | 2.5 | 3.3 |
| Stover | 7.3 | 8.1 | 18.7 | 19.4 | 15.5 | 17.2 |
| Sugar | 0.7 | 0.2 | 3.0 | 3.6 | 2.7 | 3.0 |

Figure 1. Phenotypic effects of *Glossy15* overexpression in transgenic maize. (A) Photograph of tropical-adapted maize hybrid grown in temperate environment (foreground), taller and still green relative to commercial grain hybrid at physiological maturity. (B) Simplified model for photoperiodic regulation of shoot maturation, where short days increase miR172 activity, which downregulates AP2 genes that suppress phase change. (C) Enhanced plant height and reduced seed set in Gl15-OX lines relative to control. (D) Mean yields of total biomass, grain, stover, and sugar pressed from stalk sap from plants harvested at 145 days after sowing. Data from 2008 trial in N-responsive field plots at Champaign, Illinois, and includes 5 elite grain hybrids, 3 high biomass tropical hybrids, and 5 Gl15-OX hybrids where *Gl15* transgene was introgressed into same inbred parents of the 5 elite grain hybrids.

The Moose laboratory has developed and characterized transgenic maize hybrids that overexpress the maize

*Glossy15* gene, which encodes an AP2 protein required for vegetative phase change. The *Glossy15*-overexpression (Gl15-OX) lines delay shoot maturation, leading to prolonged vegetative development, later flowering and vegetative senescence, reduced seed number, and the accumulation of greater amounts of total biomass and sugar in stem tissues relative to current commercial grain hybrids, (Figure 1). It is important to note that when grown without supplemental N fertilizer in field plots depleted for residual N, total biomass yields of Gl15-OX lines are nearly equal to those obtained for the elite grain hybrids when provided N fertilizer. These phenotypic effects mimic those observed when photoperiod-sensitive tropical maize germplasm is grown in temperate environments, in genotypes that are already well-adapted to temperate environments.

# 123

## Role of Histone Modifications in the Regulation of Cell Wall Synthesis in Rice (*Oryza sativa*)

Feng Tan,[1] Hui Wang,[1] Guo-Liang Wang,[2] Pamela Ronald,[3] and **Zhaohua Peng**[1*]

[1]Department of Biochemistry and Molecular Biology, Mississippi State University; [2]Department of Plant Pathology, the Ohio State University; and [3]Department of Plant Pathology, University of California-Davis

**Project Goals: Investigate the regulatory mechanism of cell wall synthesis and cell wall characteristics**

Cell wall synthesis is subjected to precise temporal and spatial regulation. The multiple components, including cellulose, hemicellulose, lignin, pectin, proteins, etc, have to be synthesized and deposited coordinately in the cell wall. Therefore, it is conceivable that some global regulators must be presented in the cell to coordinate different pathways of cell wall synthesis. The goal of our project is to investigate the regulatory mechanism of cell wall synthesis and cell wall characteristics. In rice protoplasts, we find that cell wall removal and regeneration is associated with substantial chromatin reorganization and global histone modification changes. ChIP-Seq studies reveal that substantially more cell wall metabolic pathway genes subjected to the regulation of histone modifications than the average of the genome. Preliminary studies further show that mutations in some key histone modification genes lead to cell wall content change in rice. Our results suggest a critical role of histone modifications in the regulation of cell wall synthesis and the characteristics of cell wall.

# 124

## Characterization of Nitrogen Use Efficiency in Sweet Sorghum

**Ismail Dweikat**[1*] (idweikat2@unl.edu), Tom Clements,[2] Don Weeks,[3] and Jianming Yu[4]

[1]Department of Agronomy and Horticulture, University of Nebraska-Lincoln; [2]Center for Biotechnology, Plant Science Initiative, and Department of Agronomy and Horticulture, University of Nebraska-Lincoln; [3]Dept. of Biochemistry, University of Nebraska-Lincoln; and [4]Dept. of Agronomy, Kansas State University, Manhattan

**Project Goals: Objective 1: Conduct quantitative trait loci (QTLs) analysis and marker identification for nitrogen use efficiency (NUE) in advanced grain sorghum populations. Objective 2: Identify loci and specific alleles that control NUE using whole genome and candidate gene association mapping techniques across a diverse set of grain and sweet sorghum accessions.**

Originated in the dry semi-arid region in Africa, sorghum has developed adaptive traits for stress environments and there is wide genetic variability for those traits including tolerance to low nitrogen supply and efficiency to utilize water and nutrients. These present opportunities for further enhancing stress tolerance and adaptability of the crop through identification of plant characteristics associated with the traits and exploiting them in breeding programs. Nitrogen is one of the essential mineral nutrients required for production of grain and biomass crops. The cost of chemical fertilizer in recent years has increased dramatically and significant amount of fertilizer nitrogen applied are lost in various ways. In sorghum it is estimated that less than fifty percent of nitrogen fertilizer added to the soil is taken up by the plant and converted to grain or biomass. This not only indicates the amount of dollars lost but also shows the yield benefit that would have been obtained if the fertilizer was properly utilized by the plant to produce grain/biomass. It is very important that this loss is reduced through developing nitrogen efficient genotypes. The objectives of this study are to identify quantitative trait loci associated with nitrogen use efficiency (NUE) using two developed recombinant inbred line population developed between a high N-efficient parents (Chin 17 and San Chi San) and the normal genotypes CK60, and to identify loci and specific alleles that control NUE using whole genome and candidate gene association mapping techniques across a diverse set of grain and sweet sorghum accessions. A total of 236 simple sequence repeat (SSR) markers have been mapped on the two populations with an average of 10 cM distance between markers. In addition to the SSR markers, 1530 single nucleotide polymorphisms (SNPs) are being mapped. A nested association mapping (NAM) population is already under development where 100 recombinant lines each were generated of BTx 623 inter-mated with 50 parental lines exhibiting wide variation high biomass producing, early germination and seedling cold tolerant, high NUE, stalk rot resistance, non-flowering, lodging resistance, high

tillering, and stay-green phenotypes. The SNPs and about 2000 SSRs will be used to screen the NAM populations to identify genomic that show association with NUE.

# 125

## Phenomic Analysis of Natural and Induced Variation in *Brachypodium distachyon*

**John Vogel***[1] (john.vogel@ars.usda.gov), Jennifer Bragg,[1] Richard Poiré,[2] Xavier Sirault,[2] Ludmila Tyler,[1,3] Vincent Chochois,[4] Michelle Watt,[4] and Robert Furbank[2]

[1]USDA-ARS, Western Regional Research Center, Albany, Calif.; [2]High Resolution Plant Phenomics Centre, CSIRO Plant Industry, Canberra, Australia; [3]University of California, Berkeley; and [4]CSIRO Plant Industry, Canberra, Australia

**Project Goals:**

1. **Assemble a collection of natural accessions and 2,000 homozygous T-DNA lines.**
2. **Conduct a detailed phenotypic characterization of the collection using a phenomic approach.**
3. **Begin detailed characterization of a select group of mutants and natural accessions.**

Herbaceous energy crops, especially grasses, are poised to become a major source of energy in the United States. Despite their increasing importance, we know little about the basic biology underlying the traits that control the utility of grasses as energy crops. Better knowledge of basic grass biology (e.g. identification of the genes that control cell wall composition, plant architecture, cell size, cell division, reproduction, nutrient uptake, carbon flux, etc.) could be used to design rational strategies for crop improvement and shorten the time required to domesticate these new crops. The model grass *Brachypodium distachyon* (*Brachypodium*) is an ideal system with which to acquire this knowledge. We are conducting high-throughput phenotypic analysis (phenomics) of homozygous T-DNA mutants and natural accessions of the model grass *Brachypodium distachyon* to accelerate the acquisition of this knowledge.

Accurate phenotypic characterization of large numbers of individuals under carefully controlled conditions is a costly and rate limiting step for functional genomic projects. These analyses typically involve destructive measurements, multiple replicate sets of plants and exhaustive manual labor. An additional problem is that measurements made at different times on different sets of plants are not directly comparable due to small variations in environmental conditions. The 'phenomic' approach to these problems is to perform multiple non-destructive phenotypic measurements in an automated high-throughput fashion on a large number of plants at one time and, where possible, to make repeated observations over time to provide more robust data. This approach achieves cost savings due to economies of scale and more reliable data due to standardization of environmental conditions, repeated observations and automation.

We are conducting high-throughput phenotypic analysis of homozygous *Brachypodium* T-DNA mutants and natural accessions. We plan to phenotype 2,000 homozygous T-DNA mutants and >100 inbred lines under defined environmental conditions at the state-of-the-art High Resolution Plant Phenomics Centre, part of the new Australian Plant Phenomics Facility. We have completed pilot experiments designed to identify the optimal growth conditions (light level, photoperiod, light quality, nutrient level, watering and soil type) prior to large-scale phenotyping. In addition, we have completed preliminary experiments to establish methods to examine root architecture. With our conditions and protocols optimized, we have initiated the first full set of experiments to phenotype a diverse collection of 100 natural accessions. These accessions were selected from a larger collection of over 1,000 accessions.

Concurrent with the initial phenotyping of the natural accessions, we are using a PCR-based approach to identify homozygous T-DNA lines containing insertions predicted to disrupt genes. We have optimized and validated our methods and have scaled up our genotyping efforts. To date, we have identified over 200 homozygous T-DNA lines and plan to have sufficient seed to phenotype this initial set beginning in 8 weeks. We anticipate that we will identify the full set of 2,000 homozygous lines within the coming year.

Student Oral Presentation–Monday

# 126

## Beneficial Bacterial Endophyte *Burkholderia phytofirmans* Strain PsJN Significantly Promotes Switchgrass Alamo Growth

Scott Lowman,[1]* Seonhwa Kim,[1] Alejandra Lara-Chavez,[1] Guichuan Hou,[4] Barry Flinn,[1] John Seiler,[3] Jerzy Nowak,[2] and **Chuansheng Mei**[1]

[1]Institute for Sustainable and Renewable Resource, Institute for Advanced Learning and Research, Danville, Va.; [2]Department of Horticulture, [3]Department of Forest Resources and Environmental Conservation, Virginia Polytechnic Institute and State University, Blacksburg; and [4]The Dewel Microscopy Facility in the College of Arts and Sciences, Appalachian State University, Boone, N.C.

http://www.isrr.ialr.org/index.php/faculty/dr-chuansheng-mei

Switchgrass is one of the most promising bionergy crop candidates for the U.S. It gives relatively high biomass yields and can grow on marginal lands. However, the biomass yield varies from year to year and from location to location. Our goal is to develop a low input and sustainable switchgrass feedstock production system utilizing beneficial bacterial endophytes. Beneficial microbial endophytes, generally, promote plant growth, increase nutrient uptake, enhance host tolerance to environmental stresses, and inhibit the growth of plant pathogens and associated diseases. We have demonstrated that one plant growth-promoting bacterial endophyte, *Burkholderia phytofirmans* strain PsJN, is able to colonize and significantly promote the growth of

switchgrass cv. Alamo under *in vitro*, growth chamber, and greenhouse conditions. Using the strain PsJN containing a GFP tag, we were able to visualize bacterial cells inside roots under confocal microscope three days after inoculation (Fig. 1).

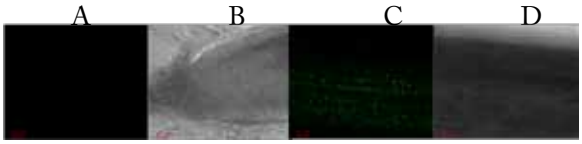



Figure 1. Root confocal images were taken 3 days after inoculation of imbibed switchgrass cv. Alamo seeds with PsJN-GFP (40X magnification). A and B are images of non-infected control root; C and D, images of the PsJN-infected root; A and C were taken under fluorescence; and B and D, under visible light.

In six independent *in vitro* experiments conducted on cv. Alamo, 10-day seedlings derived from seeds inoculated with the strain PsJN had on average 56.5% higher fresh weight than controls derived from seeds inoculated with buffer (PBS) alone. Figure 2 illustrates PsJN growth promotion of *in vitro* Alamo seedlings one month after inoculation.

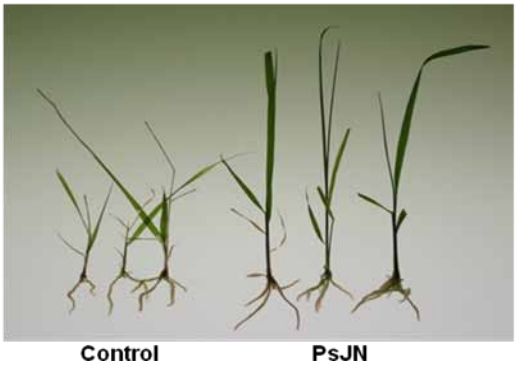



Figure 2. PsJN promoted Alamo growth.

When the one-month-old seedlings were transferred from tissue culture to cavity trays filled with 2/3 Micro Grow and 1/3 *Arabidopsis* soil mix and grown in a growth chamber (28/22˚C day/night temperature, 16-hour photoperiod) for 30 days, the PsJN-inoculated Alamo plants had significantly higher shoot and root biomass than the controls ($p<0.01$). The total dry weight averaged from five experiments was 54.1% higher in the inoculated treatment compared to non-inoculated control. Figure 3 shows one representative experiment indicating that fresh weight and dry weight in Alamo inoculated with the strain PsJN were significantly increased compared with non-inoculated control plants, with a p-value of 7.57E-08 for total dry weight.

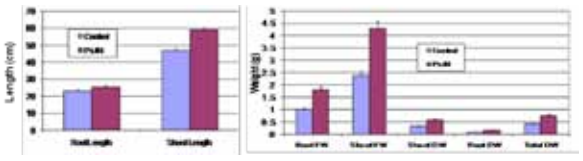


Figure 3. Strain PsJN promoted Alamo shoot and root growth in growth chamber conditions.

Similar results were obtained in greenhouse experiments with transplants grown in 6-gallon pots for two months. The total dry weight of the inoculated plants was 48.6% higher than controls. The inoculated plants formed more and earlier tillers than controls. When the PsJN-inoculated Alamo seedlings were grown *in vitro* for 25 days and transferred to 6-gallon pots containing the field soil with no fertilizer application and grown for 3 months in glasshouse under ambient conditions, they produced twice higher total dry weight than controls. The data indicate the potential benefit of switchgrass seed inoculation with PsJN for the production on marginal lands. In our preliminary data also show that the PsJN-inoculated plants had elevated lignin content compared to controls. We have also begun measuring leaf gas exchange in both PsJN-inoculated and control plants using a Li-COR 6400 photosynthesis system. Plans are to measure leaf physiological properties over the first 12 weeks of growth, starting with newly inoculated plants, to determine the physiological processes involved in the bacterium induced growth enhancement. We will also measure changes that occur during a simulated drought. In contrast to the observed stimulation of switchgrass growth responses caused by PsJN described for cv. Alamo, no beneficial responses were recorded with the upland cultivar Cave-in-Rock. In order to explore this genotype influence further, we are currently conducting a comparative global gene expression profiling in both cultivars following PsJN inoculation, using EST microarrays, in collaboration with Dr. Yuhong Tang of the Noble Foundation, Ardmore, OK. The generated information will be utilized in switchgrass breeding for low input production systems based on genetic compatibility between the host plant and inhabiting microflora. We have worked with Lynchburg Grows (Lynchburg, VA), a non-profit urban farm and environmental educational center that provides job training and educational programs for youth-at-risk and the Central Virginia Governor's school for Science and Technology to set up hydroponic tray units to educate and foster interests in the bioenergy field to the next generation of scientists.

# 127

## Biomass Accumulation in Wide Crosses Between Wild and Domesticated Rice

A. Bordeos,[2] C.E. Jahn,[1] **J. McKay**,[1] **J.E. Leach**[1]* (Jan. Leach@colostate.edu), **D. Bush**,[1] and **H. Leung**[2]

[1]Program in Plant Molecular Biology, Colorado State University, Fort Collins; and [2]Plant Breeding, Genetics and Biotechnology Division, International Rice Research Institute (IRRI), Metro Manila, Philippines

**Project Goals: Our goal is to provide the applied biomass research community and industry with information to allow exploitation of the genes and pathways relevant to biomass accumulation in grasses. The specific objectives for this project are to: Objective 1: Identify genes involved in biomass accumulation. Using genetic populations for rice lines that exhibit extremes in biomass accumulation,**

**we will (a) map QTL using a comprehensive phenotyping/genotyping approach, (b) confirm the QTL location and phenotypic effect and (c) screen a deletion mutant collection to identify large deleted regions corresponding to biomass accumulation. Objective 2: Dissect the QTL using an integrated analysis. First, we will use an expedited approach to generating near isogenic lines containing each QTL. Second, we will integrate genome-wide data to refine the QTL regions including (a) association mapping, (b) expression profiling, (c) mutant analysis, (d) fine scale mapping, and (e) comprehensive sequencing.**

Developing a sustainable liquid fuel production from cellulosic feedstock is a major challenge and will require significant breeding efforts to maximize plant biomass production. Rice has an unusual depth of genetic and phenotypic variation due to years of domestication and selection under very diverse environments (e.g. well-watered, flooded, water limited, etc.). Even more variation is available by using wide crosses between very large, perennial wild rice species (*Oryza longistaminata*) and a very small domesticated species (*O. sativa*, cultivar IR64). We are exploiting these wide crosses to identify traits for biomass productivity. The *O. longistaminata* X *O. sativa* (IR64) population was advanced through backcrossing with the recurrent parent IR64 (currently $BC_4F_2$) to reduce negative characteristics including seed shattering, fertility issues, late flowering, and lodging while maintaining increased biomass. Ten independent families of $BC_4F_2$ generation were phenotyped for total biomass, height, number of productive tillers, and days to flowering. These plant families produced approximately 2-3 times more biomass, were 30 cm taller, and produced 15 more productive tillers than the parent *O. sativa* IR64. These plants will are being genotyped to determine the genetic loci donated from the wild parent that contribute to the increases in biomass in these families and may represent novel genes to increase biomass in energy crops.

# 128

## Transcription Regulatory Networks of Cell Wall Biosynthesis Revealed by Protein-DNA Interaction Mapping

Li Lin,[1] Naomi Young,[1] Pubudu Handakumbura,[1] Ghislain Breton,[2] Todd C. Mockler,[3] Steve A. Kay,[2] and Samuel P. Hazen[1] (hazen@bio.umass.edu)

[1]Biology Department, University of Massachusetts, Amherst; [2]Cell and Developmental Biology, University of California, La Jolla; and [3]Department of Botany and Plant Pathology, Oregon State University, Corvallis

**Project Goals:**

1. **Identify regulators of plant cell wall biosynthesis**
2. **Analyze transcription factor genomic binding sites**
3. **Determine effects of regulator perturbation on amenability to deconstruction and cell wall properties of monocots and dicots**

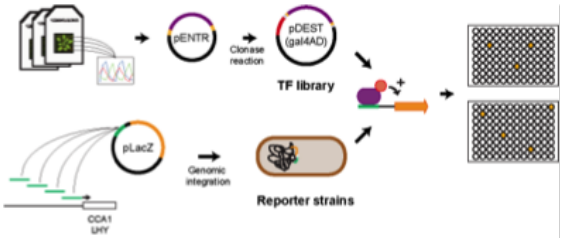



**Figure 1.** Schematic of the promoter hiking approach (Upper half represents the construction of the transcription factor library from identification of the clones using microarray, cloning in Gateway compatible vectors and transfer into gal4-AD vector for fusion protein expression in yeast. The lower half represents the cloning of tilled promoter fragments in fusion to the LacZ reporter and integration into yeast genome. The library and the yeast strain are then combined for B-galactosidase monitoring assays in 96-well plates.

One mechanism regulating cell wall biosynthesis is the activity of transcription factors that control higher order events of growth and differentiation; the likely direct regulation of processive and non-processive glycosyltransferases as well as the phenylpropanoid metabolic grid. We measured interactions among the promoters of *A. thaliana* cell wall genes and a nearly comprehensive transcription factor library using a high throughput yeast one-hybrid assay (Fig 1). In addition to several NAC and numerous MYB transcription factors some of which have been implicated in cell wall regulation, we measured interactions among over twenty other families with cell wall promoters (Fig 2A). While most (60%) bind just one, 48 many bind multiple promoters. Interestingly, cellulose, hemicellulose, and lignin cis-regulatory regions share several interactors.

Figure 2B describes in detail a protein-DNA interaction sub-network of 6 transcription factors: an AS2, bZIP, MYB, AP2, and two NAC proteins that bind cellulose, hemicellulose, and lignin promoters. Many of these interactions have been confirmed with other *in vitro* methods and *in planta*. Of partuclar interest is a NAC proteins that interacts only with cellulose genes; namely *CESA4/7/8*, *KORIGGAN*, and *COBRA-LIKE1*.

We further characterized this interaction in vivo by electrophoretic mobility shift assay using the fragments. Ultimately, we hope to determine the effects of regulator perturbation on amenability to deconstruction and cell wall properties.

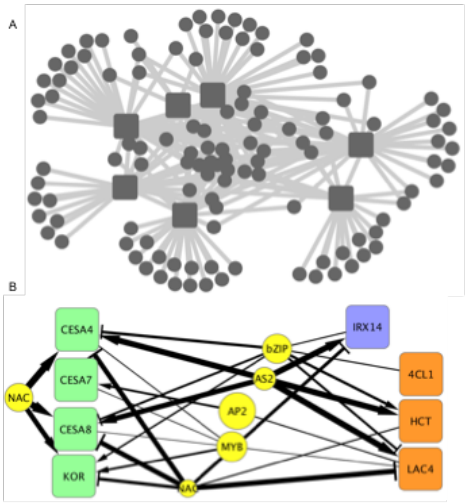



Figure 2. **(A)** An *Arabidopsis thaliana* protein-DNA interaction network of seven cell wall genes and 117 transcription factors. **(B)** A sub-network of three lignin (orange), four cellulose (green), and one hemicellulose (purple) promoter. Squares represent promoter regions and circles the transcription factors (yellow). The size of the circle is proportional to expression level in stem and the thickness of the connector (a.k.a. edge) is proportional to the correlation with the target gene in various microarray experiments. Arrows indicate a positive correlation and a circle a negative correlation.

# 129

## Accelerating the Domestication of *Miscanthus* for Biofuel Production

Changsoo Kim[1]* (changkim@uga.edu), Dong Zhang,[1] Susan Auckland,[1] Lisa Rainville,[1] Katrin Jakob,[2] Jeff Klingenberg,[2] Erik Sacks,[3] Jack Juvik,[3] Cora Muennich,[4] Martin Deuter,[4] Zili Yi,[5] Jianxiong Jiang,[5] and **Andrew H. Paterson**[1]

[1]Plant Genome Mapping Laboratory, University of Georgia, Athens; [2]Mendel Biotechnology, Hayward, Calif.; [3]Department of Crop Sciences, University of Illinois at Urbana-Champaign; [4]Tinplant, Klein-Wanzleben, Germany; and [5]Department of Biotechnology, Hunan Agricultural University, Hunan, Changsha, China

**Project Goals: The objectives of this project are to provide genomic tools and resources for an efficient biomass accumulator, *Miscanthus*. A well-saturated molecular linkage map is a prerequisite for enabling effective breeding of high yielding, locally adapted *Miscanthus* varieties for biomass production. In this project, EST-derived SSRs and SNPs are being implemented to construct a genetic map of *Miscanthus*, facilitating comparison between *Miscanthus* and the fully-sequenced sorghum genome by virtue of marker sequence information. The comparative maps are expected to be useful for transferring information from model species (such as sorghum), and for efficiently increasing marker density near genetic loci of specific interest. The comparative map will permit us to infer the locations in *Miscanthus* corresponding to a variety of previously mapped domestication or biomass-determining genes/QTLs from sorghum and other cereals. Detailed map and sequence information for *Miscanthus* will also provide clues to answer practical and fundamental questions about *Miscanthus* genome structure and organization, such as the discrepancy of the basal chromosome numbers between *Miscanthus* and much of the Saccharinae, the levels and patterns of homoeologous gene duplication, and the types and frequencies of genetic polymorphism in both diploids and polyploids that are important to *Miscanthus* improvement.**

**Research specialization:** Tropical grasses included in the Saccharinae have gained attention as biofuel feedstocks because of their high biomass production in part due to their C4 photosynthetic system. However, their complex genomes with large DNA contents and high ploidies complicate genetics and genomics research. *Miscanthus*, a promising cellulosic feedstock owing to perenniality, a longer growing season, greater leaf area, and higher carbon storage per unit of leaf area than some Saccharinae grasses, currently has only limited genetic maps, which are largely based on dominant-type semi-arbitrary markers such as random amplified polymorphic DNA (RAPD). The current research is focusing on building a framework to leverage sequence information from other taxa in *Miscanthus* improvement by integrating simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs). The outcome of this project will clarify genome structure in *Miscanthus* and its evolution since divergence from a common ancestor shared with sorghum (a botanical model for the Saccharinae), as well as provide breeding resources for sustainable biomass production by identifying specific QTL regions and diagnostic markers.

Funding source: DOE-USDA Plant Feedstock Program, Project grant number: 112786

# 130

## Genetic Variation Among Sorghum and *B. distachyon* Accessions for Biological Conversion Quality

Scott J. Lee,[1] Thomas A. Warnick,[2] Naomi F. Young,[1] Jeffrey F. Pedersen,[3] Susan B. Leschine,[2] and Samuel P. Hazen[1]

[1]Biology Department, [2]Department of Microbiology, University of Massachusetts, Amherst; and [3]USDA-ARS, Department of Agronomy, University of Nebraska-Lincoln

**Project Goals:**

1. **Development of a high throughput translational bioassay for plant biofuel properties**
2. **Assay variation for biological conversion rates among energy crop germplasm**

We developed an assay that provides the ability to measure the impact of pretreatment, conversion processes, and microbial and plant genetic diversity of digestibility,

and thereby determine the potential effects of numerous variables in biofuel production. In contrast to other established methods, the *C. phytofermentans* bioassay provides a direct and quantitative means of assessing feedstock quality, both in terms of digestibility and conversion. The use of *C. phytofermentans* takes into consideration specific organismal interactions, which will be critical in single stage fermentation or consolidated bioprocessing. In this assay, stem tissue from completely senesced plants is pulverized, inoculated with *C. phytofermentans*, and allowed to grow anaerobically. Total sample requirements are as low as 25mg, making multiple measurements of stems from very small plants such as *B. distachyon* and *A. thaliana* possible. Supernatant ethanol concentration measured by HPLC (high performance liquid chromatography) is used as the metric to estimate feedstock quality and conversion efficiency. Ethanol concentration was directly proportional to the concentration of energy crop sorghum feedstock in the culture (Fig panel A) and increased in a linear manner over time (Fig panel B). The assay is capable of detecting significant differences in ethanol production between wild-type sorghum and *brown midrib* (*bmr*) mutants with one or two mutations in the lignin biosynthesis pathway (Fig panel C). We sieved ground biomass and assayed only feedstock ranging from 53 to 62.4 μm and confirmed that differences in ethanol yield are not the result differential grinding among genotypes. We have also detected significant genetic diversity among *B. distachyon* accessions (Fig panel D) sorghum landraces and *A. thaliana* accessions (data not shown). The sorghum accession most amenable to conversion yielded 30% more ethanol than the most recalcitrant, a range we observed among *B. distachyon* and *Arabidopsis thaliana* accessions as well. The genetic differences measured among accessions within several species are similar to the range measured between the *bmr* lignin mutants and wild-type sorghum. While mutations effecting lignin have a deleterious effect on plant architecture, vigor, and yield in many crop species, the accessions we observed significant variation for ethanol yield exhibit no such differences in overall plant architecture. This suggests that gain from selection and transgenic modification for feedstock quality need not be pleiotropic for low biomass yield.

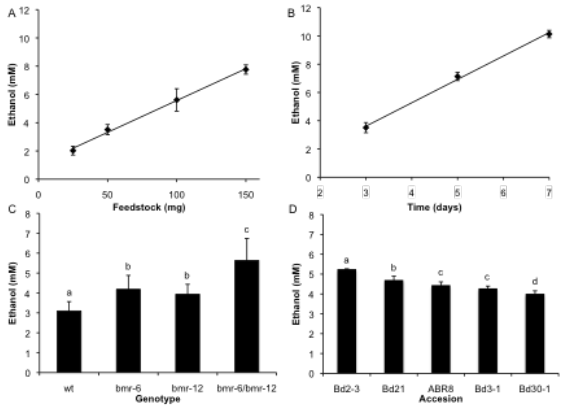



**Figure.** Bioassay for digestibility using *C. phytofermentans* ethanol production as a measure of feedstock quality. Ethanol production on sorghum is proportional to (A) feedstock concentration and (B) time after inoculation. (C) Diversity of ethanol production by *C. phytofermentans* grown on wild type and single (*bmr-6* and *bmr-12*) and double *brown midrib* (*bmr-6/bmr-12*) lignin biosynthesis sorghum mutant feedstocks. (D) Diversity of ethanol production by *C. phytofermentans* grown on five accessions of *B. distachyon* as feedstock. Values followed by the same letter are not significantly different at $P < 0.05$ based on Duncan's Multiple Range test.

# 131

## Regulation of Root Development in *Populus* in Response to Nitrogen Deficiency and Drought

**Victor Busov**[1]* (vbusov@mtu.edu), Hairong Wei,[1] Erik Lilleskov,[2] and Yordan Yordanov[1]

[1]Michigan Technological University; and [2]Forest Service

**Project Goals: see below**

To avoid competition with food production and curb greenhouse gas emissions, lignocellulosic bioenergy crops will be primarily grown on marginal lands. Intermittent, patchy, and limited availabilities of resources like nitrogen (N) and water will severely compromise productivity under such environments. Therefore, sustainable production of bioenergy crops will entail developing varieties that can maintain high levels of biomass productivity under sub-optimal N and water conditions. Root architecture is essential in sensing, absorbing, and transporting both nitrogen and water. The genetic bases of these root traits are poorly understood and thus difficult to manipulate. We use a systems biology approach to understand how nitrogen and water shape root architecture in the lignocellulosic bioenergy crop *Populus*.

The goal of this project is to generate a systems-level knowledge and identify key regulators of root architecture in relation to nitrogen and water use in the bioenergy crop *Populus*. Objectives are to:

1. Perform comprehensive microarray profiling on poplar roots during response to N and water limitations.
2. Construct genetic networks inferred from microarray analysis and generate a systems-level knowledge of the underlying mechanisms.
3. Generate transgenic modifications of key regulators identified in Objective 2 and characterize their response to nitrogen and water.
4. Identify activation-tagged mutants affecting water and nitrogen response, isolate the candidate tagged genes, and recapitulate phenotypes via retransformation in a subset.
5. Use advanced stereo x-ray imaging to characterize root architecture under soil-like conditions in transgenic plants generated in Objectives 3 and 4.
6. Develop web portal for access to, analysis, and dissemination of project data.

To date we have produced a comprehensive transcriptional profile of roots' response to N and water deficiencies. Building on this resource we have used advanced genetic network analyses to generate systems-level knowledge of the underlying molecular mechanisms. We have identified key regulators and have begun modifying their expression in transgenic plants to test the biological significance of the identified network mechanisms. In a parallel and complementary approach taking advantage of the poplar genome

sequence and efficient transformation system, we use activation tagging as a forward genetics approach to discover novel genes or corroborate the effect of genes identified via the genetic network analyses. To this end we have generated 2,000 activation tagged lines and have screened 1,000 for modified response to nitrogen deficiency or drought stress. We have identified and validated 53 mutants which display enhanced growth characteristics under the nitrogen and water stress conditions. For approximately half of these we have already positioned the tag, validated the behavior of the tagged genes and for a limited number initiated recapitulation experiments. A web site which will facilitate access and dissemination of data and germplasm generated through the project is under construction.

# 132

## Genomics of Energy Sorghum Biomass Accumulation

**John Mullet**[1]* (jmullet@tamu.edu), **Patricia Klein**,[2] and **William Rooney**[3]

[1]Department of Biochemistry and Biophysics, Texas A&M University, College Station; [2]Institute for Plant Biotechnology and Genomics, Texas A&M University, College Station; and [3]Department of Soil and Crop Science, Texas A&M University, College Station

**Project Goals: The overall goal of the proposed research is to identify the genetic and biochemical basis for increasing the yield and improving the composition of high biomass cellulosic energy sorghum. The specific objectives of the proposed research are to; (1) characterize the molecular diversity of 750 photoperiod sensitive (late flowering) energy sorghum germplasm accessions and use this information to select 250 diverse accessions for analysis of variation in stem biomass yield, structure, and composition, (2) map QTL for stem biomass yield, structure, and composition in three energy sorghum populations derived from diverse sorghum parental genotypes, and (3) develop information and biological resources that will enable positional cloning of QTL/genes and analysis of gene regulatory networks that modulate energy sorghum biomass yield, stem structure, and composition.**

High biomass energy sorghum (*Sorghum bicolor* L. Moench) has excellent potential as a bioenergy crop and research on this species will also provide fundamental information about the genomics of C4 grass energy crop design for the following reasons; (1) sorghum, like *Miscanthus* and energy cane, is a highly productive C4 grass that under optimum conditions can produce ~20dT of biomass per acre making it one of the most productive bioenergy crops currently under development, (2) energy sorghum has excellent drought tolerance and high water use efficiency, critical attributes for production of bioenergy crops in marginal environments and where irrigation is either too expensive or would deplete water reserves, (3) energy sorghum has wide adaptation and is highly amenable to production and cultivation systems currently used in the U.S. facilitating rapid adoption by producers at low risk, (4) energy sorghum is a resilient low risk annual hybrid crop that can be used in normal crop rotations to maintain soil fertility, reduce pest pressures, and meet annual variation in demand for biomass, (5) the extensive and diverse sorghum germplasm collection (~40,000 accessions) contains useful genetic variation for an array of bioenergy traits including biomass yield, composition, and drought tolerance that can be mined and exploited for further improvement of energy sorghum, (6) sorghum's good genetics, relatively small genome size (~800Mbp) and complete genome sequence provides an excellent technology platform for conducting genome-scale research into pathways that influence biomass yield and composition, and (7) information gained through analysis of energy sorghum will be useful for the design of perennial bioenergy C4 grass species such as switchgrass, *Miscanthus* and energy cane that are more complex in genetics and breeding. Information and biological resources generated by this project will be used to create improved versions of high biomass energy sorghum and other C4 bioenergy grasses in order to minimize acreage used for biomass production, reduce food vs. biofuels competition, and reduce the cost of feedstock, while increasing the carbon balance of biofuels and creating a sustainable source of biomass feedstock for large scale biofuels production in the U.S.

# 133

## Genome and Developmental Variation in DNA Methylation in Poplar

Kelly Vining,[1] Kyle R. Pomraning,[2,3] Larry Wilhelm,[4] Henry D. Priest,[4] Cathleen Ma,[1] Ruoqing Zhu,[1] Elizabeth Etherington,[1] Matteo Pellegrini,[5] **Todd Mockler**,[4] **Michael Freitag**,[3] and **Steven Strauss**[1]* (steve.strauss@oregonstate.edu)

[1]Department of Forest Ecosystems and Society, Center for Genome Research and Biocomputing, [2]Molecular and Cellular Biology Program, [3]Dept. of Biochemistry and Biophysics, [4]Department of Botany and Plant Pathology, Oregon State University, Corvallis; and [5]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles

**Project Goals:**

1. **Produce and characterize DNA methylation-deficient poplars generated by RNA-induced silencing of the poplar homologs of the *Arabidopsis* Decreased DNA Methylation 1 (DDM1) gene, PtDDM1-1 and PtDDM1-2.**
2. **Prepare genomic DNA from tissues of the sequenced *Populus* trichocarpa clone and use commercially-available antibodies against 5-methylcytidine to isolate methylated DNA segments.**
3. **Prepare chromatin from tissues of the sequenced *P. trichocarpa* clone and use commercially available antibodies to methylated lysine residues K9 and K27**

**of histone H3 to produce DNA fractions enriched for these modifications.**

4. **Sequence the methylated and enriched DNA on an Illumina GAII analyzer at the OSU CGRB and conduct bioinformatics and statistical studies to test the following hypotheses: a. DNA and/or histone methylation differs between differentiation states. b. DNA and/or histone modifications are associated with RNA transcription levels. c. Chromatin states and gene expression are associated.**

We are conducting research to determine the role of epigenetic modifications during tree development using poplar (*Populus trichocarpa*), a reference woody feedstock species. Using methylated DNA immunoprecipitation followed by high-throughput sequencing (MeDIP-seq), we have analyzed DNA methylation patterns in the *P. trichocarpa* genome in relation to four biological processes: bud dormancy and release, mature organ maintenance, *in vitro* organogenesis, and methylation suppression.

We sequenced methylated DNA from eleven target tissues in wildtype *P. trichocarpa*: leaves, roots, xylem, phloem, fall buds, winter buds, spring buds, stem explants, callus and *in vitro* regenerated plants. A total of 64 Illumina sequencing lanes represented 1 to 3 biological replicates for each sampled tissue, 16M – 125M sequencing reads per tissue type. Reads aligning to unique positions in the reference genome covered ~30% of genome space. Average sequence depth within covered regions varied by tissue type, ranging from 4 to 12 reads/nucleotide. Numbers of significantly methylated tiled 1Kb genome windows called by RPKM calculations at a 1% false discovery rate varied by tissue type, ranging from approximately 2,000 (xylem) to 40,000 (pooled bud data). In all tissues, transposons and other repeat elements were enriched relative to their overall representation in the genome, with LTR-gypsy retroelements being the most highly enriched transposable element type. Gene methylation exhibited a pattern of higher methylation at promoters, middle of coding region, and 3' UTRs relative to 5' and 3' ends of coding regions. Numbers of methylated genes varied by tissue type and gene region considered, and represented 3-5% of the genes in the genome. We performed bisulfite sequencing of nine selected target regions with varying MeDIP-seq signal. Results confirmed MeDIP-seq results, and allowed a higher-resolution view of methylation at selected genes.

We have produced summary data for genome methylation in *P. trichocarpa*, including distribution of methylation across chromosomes and in and around genes. This process has been driven by the development and adaptation of bioinformatic and statistical methods. Further, we have analyzed similarities and differences in methylation patterns among tissue types from four biological processes. We have developed a customized genome browser (Gbrowse version 1.69), compatible with the most recent (v2*) P. trichocarpa* genome assembly, at which our data can be explored: http://poplar-dev.cgrb.oregonstate.edu/cgi-bin/gbrowse/poplar_v2/.

# 134

## Poplar Biomass Interactome

**Eric Beers**,[1] **Amy Brunner**[2]* (abrunner@vt.edu), **Allan Dickerman**,[3] Mingzhe Zhao,[1] Xiaoyan Jia,[1,2] Chengsong Zhao,[1] and Xiaoyan Sheng[2]

[1]Virginia Tech Departments of Horticulture and [2]Forest Resources and Environmental Conservation; and [3]Virginia Bioinformatics Institute at Virginia Tech

http://xylome.vbi.vt.edu/

**Project Goals:**

1. **Clone approximately 400 Gateway-compatible ORFs corresponding to the poplar xylem gene set (PxORF)**
2. **Identify poplar proteins that co-purify with selected TAP-tagged PxORF proteins expressed in poplar**
3. **Identify high-confidence Y2H interactions for a subset of approximately 60 PxORFs comprising putative regulators of lignocellulose synthesis screened against a poplar xylem cDNA library**
4. **Identify Y2H interactions resulting from a matrix of pair-wise assays between all PxORF proteins**
5. **Produce a protein-protein interaction map that incorporates interactions identified from the three screens**
6. **Maintain a web site to make results available and facilitate distribution of clones**

Proteins are molecular machines that are required for nearly all biological functions based on interactions with other molecules such as carbohydrates, lipids, other low molecular weight molecules, nucleic acids and other proteins. We are mapping protein-protein interactions relevant to biomass production by focusing on proteins coexpressed in poplar secondary xylem. In combination with transcriptomic and metabolomic data, this high-confidence wood interactome will provide a solid foundation for identifying key regulators of wood formation and biomass accumulation.

To date, 374 PxORFs were cloned into Gateway-compatible pENTR vectors, of which 323 were subsequently cloned into pDBdest and 335 into pADdest vectors respectively (http://xylome.vbi.vt.edu/ORF_List). In addition, we prepared transgenic poplar overexpressing 11 PxORFs as TAPa-tagged fusions for co-purification of interacting proteins. We established a replicated field trial of all TAPa-tagged overexpression lines, which will permit phenotypic analysis of the effects of overexpression of these TAPa-tagged proteins and ensure production of sufficient wood for extraction and identification of co-purified proteins. A 108,205 (323DB x 335AD) Y2H binary screen identified 11 PxORF interacting pairs. We are also screening a xylem cDNA prey library for interactors with a subset of PxORFs. For the 26 PxORFs that are completely through the library screen, we have identified 44 unique interacting sequences. Selected interactions have been or will be confirmed by other methods including bimolecular fluorescence complementation and co-immunoprecipitation using plant transient expression systems. The proportional yield from our

binary screen is similar to that represented by the current preliminary binary screen data from the *Arabidopsis* interactome project. In contrast, the proportional yield from our library screen is much higher. Additionally, in most cases, the proportional yield for enzymatic/structural proteins catalyzing metabolic reactions (such as cellulose synthase PB138) is much lower than that of regulatory proteins (such as NIMA kinase PB223). We have begun to integrate our findings for poplar xylem protein interactions with other protein-protein interaction data to produce a preliminary network by whereby poplar proteins are represented by their putative *Arabidopsis* orthologs.

Functional analyses of selected interacting proteins should provide valuable insight regarding new strategies for regulating woody biomass production. Hence, we have begun to functionally characterize select interacting pairs in both poplar and *Arabidopsis* by ectopically expressing or suppressing genes singly and in combination. For example, one interacting pair we are studying is PB15 (ROP-GTPase) and PB129 (DUF620). Co-overexpression of PB15 and PB129 in *Arabidopsis* resulted in expanded interfascicular regions containing enlarged fibers compared to fibers in normal interfascicular regions of the inflorescence stem. However, this phenotype was not observed in transgenics overexpressing just one of these genes, showing the potential of interactome data to be translated into alteration of wood phenotypes.

# 135

## Strategies for Using Molecular Markers to Simultaneously Improve Corn Grain Yield and Stover Quality for Ethanol Production

**Rex Bernardo,[1] Hans-Joachim G. Jung,[2]** Jon M. Massman[1]* (massm016@umn.edu), Magan F. Lewis,[3] and Robenzon E. Lorenzana[4]

[1]Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul; [2]USDA-ARS, Plant Science Research Unit, St. Paul, Minn.; [3]Department of Plant Sciences, North Dakota State University, Fargo; and [4]AgReliant Genetics, Ivesdale, Ill.

**Project Goals: Our objective was to optimize the use of DNA markers to simultaneously breed for high corn grain yield (for non-energy and energy uses) and high stover quality for ethanol production. Specifically, we aimed to (1) determine the prospects of and identify challenges in marker-assisted breeding for both corn grain yield and stover quality traits important for ethanol production; and (2) determine if genomewide selection (which does not require finding markers with significant effects) is superior to the usual approach of selecting only for significant markers, with the goal of simultaneously improving corn grain yield and stover quality.**

About 235 million metric tons of corn (*Zea mays* L.) stover (i.e., stalks, leaves, cobs, husks, and tassels) are left unharvested in U.S. corn fields each year. This stover represents a most abundant source of lignocellulosic substrate that can be converted to ethanol biofuel. But while today's corn hybrids have been aggressively bred for grain yield, they have not been bred for stover-quality traits important for ethanol production. Here, our objective was to optimize the use of DNA markers to simultaneously breed for high corn grain yield (for non-energy and energy uses) and high stover quality for ethanol production.

We used SNP markers combined with classical quantitative-trait analyses to study the extent to which grain yield, agronomic traits, and stover quality can be simultaneously improved in the B73 x Mo17 corn population. Three stover-quality traits were measured: concentration of cell wall glucose in dry stover ("Glucose"); cell wall glucose released from the stover by thermochemical pretreatment and enzymatic saccharification ("Glucose Release"); and concentration of lignin on a cell-wall basis ("Lignin"). Genetic variances were significant for grain yield, moisture, stalk and root lodging, plant height, and all three stover-quality traits. Heritabilities of the stover quality traits were 0.57 for Glucose, 0.63 for Glucose Release, and 0.68 for Lignin. Genetic and phenotypic correlations among traits were generally favorable but also reflected the complexity of corn stover cell wall composition. We found 152 QTL, mostly with small effects, for Glucose Release and cell wall components on both a dry matter and cell wall basis. Because no major QTL were found, we expected that methods that predict performance based on markers, such as genomewide selection, would be appropriate in marker-assisted breeding for these traits. Responses to three cycles of selection for Glucose, Glucose Release, and Lignin were higher with genomewide selection (which utilized all markers rather than only those with significant effects) than with selection based only on significant markers. These responses were determined from NIRS predictions, and we are conducting wet chemistry tests to measure Glucose, Glucose Release, and Lignin in the populations. To our knowledge, this work represents the first report of the usefulness of genomewide selection based on empirical data in plants.

We conclude that current corn-breeding programs should be able to incorporate stover quality for cellulosic ethanol as a breeding objective, without having to use unadapted or exotic germplasm. Given the absence of major QTL and the complexity of the traits, we recommend genomewide selection for the improvement of stover-quality traits for cellulosic ethanol in corn.

# 136

## Alfalfa Transcript Sequencing and SNP Discovery to Improve Biomass Composition

Xuehui Li,[1] Jiqing Guo,[1] John A. Crow,[3] Andrew D. Farmer,[3] Yuanhong Han,[1] Ji He,[1] Ryan Kim,[3] Mingyi Wang,[1] Yanling Wei,[1] **Gregory D. May**,[3] **Maria J. Monteros**,[1,2] and **E. Charles Brummer**[1,2*] (ecbrummer@noble.org)

[1]The Samuel Roberts Noble Foundation, Ardmore, Okla.; [2]Institute of Plant Breeding, University of Georgia, Athens; and [3]National Center for Genome Resources, Santa Fe, N.M.

**Project Goals:**

1. **Sequence alfalfa (*Medicago sativa L.*) transcriptomes to identify common SNP variants and develop a high throughput GoldenGate SNP array, and**
2. **Identify SNP markers associated with biofeedstock-related traits in diploid and tetraploid mapping populations.**

Alfalfa is an important forage crop in temperate and dry tropical regions in the world; it also is a potential biofuel crop. Alfalfa has the advantages of high yield, a high lignocellulose concentration in stems, and low input costs due to high levels of biological nitrogen fixation. Marker assisted selection (MAS) and/or genomic selection (GS) could enhance alfalfa improvement, but large numbers of markers are needed to map important agronomic traits and predict genomic breeding values. The main objective of this study is to identify single nucleotide polymorphism (SNP) for alfalfa using Illumina transcriptome sequencing and to develop SNP assays in candidate genes for biomass yield and composition. We have sequenced 27 alfalfa transcriptomes, including elite genotypes from four major alfalfa breeding companies in the U.S. A pilot sequencing of three genotypes resulted in a total of 80.7 million reads, assembling of which generated 155,484 contigs with a total length of 51.3 Mbp and an average length of 330 bp, giving an average read depth of 36-fold for each genotype. The realignment of reads to the contigs enabled the detection of 260,848 putative SNPs (one SNP per 197 bp) and 11,090 InDels among the three genotypes. Over 95% of the SNPs have coverage of 5 or more reads for each of the three genotypes. Of all contigs, 55.6% were aligned to *M. truncatula* coding sequence, build 3.0, which carry about 190,000 SNPs. The distribution of these SNPs along eight chromosomes is roughly even. We are developing marker assays for selected SNPs in order to map important agronomic traits and assess linkage disequilibrium (LD) and population structure in breeding populations.

We have evaluated the LightScanner high resolution melting (HRM) technology from Idaho Technologies for use with autotetraploid alfalfa. The technology enables clear assessment of allele dosage and is reasonably high throughput. However, assessment of hundreds or thousands of loci is beyond the routine capacity of this technology. Therefore, we are discussing array design with Illumina. Illumina now offers the ability to distinguish among multiple heterozygote classes (AAAT, AATT, and ATTT, for example) making this platform will be able to generate useful data for alfalfa breeding applications. Using short read sequence data generated in this and other experiments, we have identified SNP in several genes in the lignin biosynthetic pathway and mapped them on the alfalfa genetic map. We will preferentially identify SNP in other cell-wall related genes as we analyze the transcriptomes of the 27 genotypes to include on the array.

We have developed genetic maps in three diploid alfalfa populations in addition to maps developed previously in two tetraploid populations. We are in the process of mapping QTL for cell wall composition on the tetraploid populations and are growing two diploid populations in the field, which will be analyzed for yield and composition. SNP markers developed as part of this project will be used to locate candidate genes on the linkage maps for comparison to QTL locations.

# 137

## The Regeneration and Transformation of Foxtail Millet (*Setaria italica*), A Model Biofuel Crop

Xinlu Chen* (xchen24@utk.edu), **Janice Zale,** and **Feng Chen**

Department of Plant Sciences, University of Tennessee, Knoxville

**Project Goals: The goal of this project is to establish genetic transformation systems for foxtail millet (*Setaria italica*), a newly established model plant for genetic improvement of other biofuel crops, such as switchgrass (*Panicum virgatum*).**

Foxtail millet (*Setaria italica* L.) is a warm-season C4 annual crop commonly grown for grain and forage production worldwide. Recently, foxtail millet was established as a model plant for the genetic improvement of other biofuel crops, especially switchgrass (*Panicum virgatum* L.). A number of genomic tools have been established for foxtail millet, including a fully sequenced genome. The goal of this project is to establish genetic transformation systems for foxtail millet. Seeds and immature inflorescences are used as explants. Protocols for callus initiation, somatic embryo formation, and plantlet regeneration from explants have been developed for the foxtail millet genotype Yugu1. Optimal media for the induction of callus and somatic embryos from immature inflorescence explants is determined to be Murashige and Skoog (MS) medium containing 2.5 mgl-1 2,4-dichlorophenoxyacetic acid (2,4-D), 0.6 mgl-1 6-benzylaminopurine (BAP), and 3% sucrose. Calli induction from explants of immature inflorescences was significantly higher (76%) in comparison to the other two media tested (Table 1), whereas calli induction from seed explants was also relatively high

on this medium compared to six others tested (68.5%; data not shown). Callus induction from this medium is used for Agrobacterium-mediated transformation using AGL1 strain harboring a binary vector pCAMBIA1305.1. Transient expression of the reporter GUS gene is improved with treatments by sonication, vacuum, centrifugation and addition of L-cysteine and DTT. The higher centrifugation speed increased the number of GUS foci (Figure1). The Calli inoculated are cultured in the dark at room temperature. After three days, the calli are transferred onto callus induction medium containing 150 mg l-1 cefotaxime and 150 mg l-1 timentin. After two weeks, the calli are subcultured onto callus induction medium containing 150 mg l-1 cefotaxime, 150 mg l-1 timentin and 1.5 mg l-1 hygromycin. After two to three months, the calli are transferred to regeneration medium (MS medium, 0.2 mg l-1 kinetin, 3 % sucrose, pH 5.8) containing 150 mg l-1 cefotaxime, 150 mg l-1 timentin and 1.5 mg l-1 hygromycin. Finally Putative transgenic foxtail millet plants have been regenerated from callis.

Table 1. Media Comparison for Induction of Calli From Immature Inflorescences

| Media | Explants Plated | Calli Induced | Percentage of Calli Induced | Mean Percentage[a] ± S.E.[b] |
|---|---|---|---|---|
| MS, 2.5 $mgl^{-1}$ 2,4-D, 0.6 $mgl^{-1}$ BAP | 560 | 426 | 76.1 | 76.1 ± 5.3 a |
| MS, 5.0 $mgl^{-1}$ 2,4-D, 1.1 $mgl^{-1}$ BAP | 210 | 0 | 0.0 | 0.0 ± 0.0 b |
| N6E | 210 | 8 | 3.8 | 4.0 ± 0.0 b |

[a]Mean percentages followed by the same letter are not significantly different at the 5% level as determined by Tukey's Multiple Comparison; there were four replicates of each treatment. [b]S.E. = standard error.

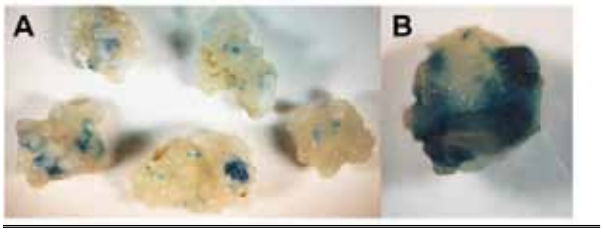



Figure 1. Transient GUSPlus™ expression in foxtail millet calli after treatments with sonication, vacuum and centrifugation. Gus foci showed after centrifugation for one minute at (A) 7,000 rpm and (B) 13,000 rpm.

# 138

## Ploidy Variation and Reproductive Pathways in Upland Switchgrass

**Edward S. Buckler**,[1] **Michael D. Casler**,[2] **Jerome H. Cherney**,[3] Patrick Redmond,[4] and Denise Costich[1]* (dc58@cornell.edu)

[1]USDA-ARS; [2]USDA-ARS U.S. Dairy Forage Research Center, Madison, Wis.; [3]Department of Crop and Soil Sciences, Cornell University; and [4]Cornell University

http://www.maizegenetics.net

**Project Goals: (1) Assemble association panels of diverse populations and linkage populations for switchgrass and reed canarygrass. (2) Trait evaluation for key biofeedstock characteristics in these panels. (3) Develop high density SNP markers in switchgrass. (4) Genotype association panels and linkage populations in switchgrass. (5) Evaluate population structure and germplasm diversity in switchgrass. (6) Establish association mapping and estimate marker based breeding values in switchgrass.**

Many of the plant species that have been targeted for biofeedstock development have complex, polyploid genomes with limited prior research devoted to them. Switchgrass is a prime example of this. We employed a combination of approaches, including flow cytometry, classic cytology and molecular cytogenetics, to gain a better understanding of the extreme variation in chromosome numbers found in this species. Knowledge of chromosome-number variation in our key germplasm will be critical for the interpretation of genetic marker data, genetic mapping and breeding efforts that rely on marker-assisted selection.

In a recently published study (Costich et al. 2010. Plant Genome), our flow cytometric survey of a core set of 11 primarily upland polyploidy switchgrass accessions indicated that there was considerable variation in genome size within each accession, particularly at the octoploid ($2n$ = 8X = 72 chromosome) ploidy level. Highly variable chromosome counts in mitotic cell preparations indicated that aneuploidy was more common in octoploids (86.3%) than tetraploids (23.2%). The incidence of hyper- versus hypoaneuploidy is equivalent in tetraploids, however, this is clearly not the case in octoploids where close to 90% of the aneuploid counts are lower than the euploid number. Fluorescent in situ hybridization (FISH) revealed an unexpected degree of variation in chromosome structure underlying the apparent genomic instability at the octoploid level.

Polyploidy and reproductive biology are linked by the underlying mechanism of unreduced gamete formation, that is, the production of eggs and/or sperm with the somatic chromosome number. This one alteration in the outcome of meiosis can have profound effects on the reproductive success of the individual plant and on the overall population structure, affecting gene flow and the distribution of genetic diversity. As a follow-up to the research described above on

switchgrass ploidy and aneuploidy, we have initiated a study of the reproductive pathways in tetraploids and octoploids, examining the ploidies and genetic relatedness of maternal parent plants (both 4X and 8X) and their offspring (seeds). A flow-cytometric seed screen (FCSS; Matzke et al. 2000. Plant Journal) was carried out to compare the ploidies of the embryo and endosperm cell populations in seeds with the ploidy of the parent. Sets of seed were germinated and grown up to confirm the FCSS analysis and will be genotyped to examine the apomictic versus sexual nature of their origin. A better understanding of the reproductive biology of this species will provide the foundation for more efficient breeding programs, as well as, improved analysis and interpretation of the sequence data being generated by ongoing genomics projects.

# 139

## Genomics of Wood Formation and Cellulosic Biomass Traits in Sunflower

Jessica G. Barb[1]* (jgbarb@uga.edu), Sukhpreet Sandhu,[1] John E. Bowers,[1,2] Angela Ziebell,[3] Crissa Doeppke,[3] Robert Sykes,[3] **Jeffrey F. Dean**,[4] **C. Joseph Nairn**,[4] **Mark Davis**,[3] **Laura F. Marek**,[5] **Steven J. Knapp**,[1] and **John M. Burke**[2] (jmburke@uga.edu)

[1]Center for Applied Genetic Technologies, Univ of Georgia, Athens; [2]Dept of Plant Biology, Univ of Georgia, Athens; [3]US Dept of Energy, National Renewable Energy Laboratory, Golden, Colo.; [4]School of Forestry and Natural Resources, Univ of Georgia, Athens; [5]USDA-ARS and Dept of Agronomy, Iowa State Univ, Ames

**Project Goals: see below**

The long-term goals of this project are to: (a) develop woody, high biomass cultivars of common sunflower (*Helianthus annuus* L.) for biofuel production; (b) gain insights into genetic and non-genetic factors affecting xylogenesis, secondary cell wall differentiation, lignocellulose accumulation, wood formation, biomass yield, and other cellulosic biomass traits in sunflower; and (c) gain an understanding of the chemical, physical, and biofuel properties of sunflower 'wood'. The oilseed sunflower hybrids grown around the world today produce pithy stems, and wood-forming ecotypes have not been discovered in *H. annuus*. In contrast, several wild sunflower species produce woody stems, are interfertile with *H. annuus*, and have the potential to supply genetic diversity for developing woody cultivars and enhancing cellulosic biomass yield in sunflower. Xylogenesis, wood formation, and lignocellulose accumulation have not been studied, selection for cellulosic biomass traits has not been done, and wood-producing cultivars have not been developed in sunflower. This project thus focuses on the development of the resources and knowledge needed for manipulating cellulosic biomass traits in hybrid sunflower breeding programs, and for tapping genetic diversity for wood formation, cellulosic biomass yield, and other traits in two drought-tolerant, wood-forming, wild species: silverleaf sunflower (*H. argophyllus* L.) and Algodones dune sunflower (*H. niveus* subsp. *tephrodes* (Gray) Heiser).

The specific objectives of this project are to:

1. Identify anatomical, developmental, physical, and chemical differences in the stems of woody and non-woody ecotypes of sunflower.
2. Develop an EST database from relevant tissues for use in SNP discovery and microarray development; perform comparative transcriptomic analyses of woody and non-woody sunflower ecotypes to identify genes responsible for wood development and cellulosic biomass traits in sunflower.
3. Characterize the variation present within *H. annuus* and *H. argophyllus* in terms of genetic diversity, wood chemistry properties, biomass traits, and other agronomic traits.
4. Investigate the genetic architecture of biomass traits and wood chemistry in sunflower using a genetic map-based approach to identify genomic regions harboring QTL for wood formation and other cellulosic biomass traits; develop QTL-NILs for wood formation, cellulosic biomass, and other traits.
5. Construct comparative genetic maps of *H. annuus*, *H. argophyllus*, and *H. niveus* ssp. *tephrodes* to characterize chromosomal differences that might limit gene introgression from the wild species into the cultivated sunflower gene pool.

# 140

## Developing Genomic Selection (GS) and Genome-Wide Association Studies (GWAS) for Upland Switchgrass

PIs: **Edward S. Buckler**[1]* (esb33@cornell.edu), **Michael D. Casler**,[2] and **Jerome H. Cherney,**[3]
**Co-Authors:** Denise Costich,[1] Elhan Ersoz,[6] Moria Sheehan,[4] Fei Lu,[5] and Alexandra Lipka[5]

[1]USDA-ARS; [2]USDA-ARS U.S. Dairy Forage Research Center, Madison, Wis.; [3]Department of Crop and Soil Sciences, Cornell University; [4]Nature Source Genetics; [5]Cornell University; and [6]Sygenta

http://www.maizegenetics.net

**Project Goals: (1) Assemble association panels of diverse populations and linkage populations for switchgrass and reed canarygrass. (2) Trait evaluation for key biofeedstock characteristics in these panels. (3) Develop high density SNP markers in switchgrass. (4) Genotype association panels and linkage populations in switchgrass. (5) Evaluate population structure and germplasm diversity in switchgrass. (6) Establish association mapping and estimate marker based breeding values in switchgrass.**

The ability to predict and model trait variation with genomic markers has tremendous opportunities for the breeding of perennial crops. In the standard breeding cycle of perennials, it can be five or more years between cycles of advancement, while genomic selection permits advancement every time a cross can be made, for example, every 1-2 years. To enable genomic selection and genome wide association analysis in switchgrass, we have developed germplasm resources from upland switchgrass, which includes bi-parental mapping populations and association mapping populations. Replicated phenotypic trials in both Wisconsin and New York have shown that while there is tremendous phenotypic variation for a wide range of traits, this species has substantial problems with establishment, which is exacerbated by cold winters and soil conditions. We have combined these trials with two studies on the genome of switchgrass—chromosome biology and high throughput genotyping. Analysis of the chromosome biology found substantial instability in chromosome number, with tetraploids, hexaploids, and octoploids all found in our germplasm. There were also many accessions showing a gain or loss of a few chromosomes in mitotic cells (aneuploidy). While the chromosome variation is tractable, it must be considered in any breeding effort, and it strongly favors using the tetraploids. Finally, genotyping-by-sequencing approaches have been applied to switchgrass in order to identify thousands of variable regions of the genome. These studies are enabling genomic selection models, which will be evaluated in the coming months. Overall, upland switchgrass has tremendous variation, but whether it is bred directly using molecular markers, or traits from upland switchgrass are introgressed with the aid of markers into lowland switchgrass, this study provides a foundation for advancement and identifies the challenges.

# 141

## Transcriptional Genomics in Maize for Improvement of Bioenergy Grasses

Bryan W. Penning,[1,2] Christopher K. Dugard,[1] Nicholas Babcock,[3] John F. Klimek,[1] Robert Sykes,[4] Mark Davis,[4] Clifford Weil,[3] Maureen C. McCann,[2] and **Nicholas C. Carpita**[1]* (carpita@purdue.edu)

[1]Departments of Botany and Plant Pathology, [2]Biological Sciences, and [3]Agronomy, Purdue University, West Lafayette, Ind.; and [4]National Bioenergy Center, National Renewable Energy Laboratory, Golden, Colo.

**Project Goals: The goals of our project are to use bioinformatics and high throughput sequencing technologies to identify and classify the genes involved in cell wall formation in maize for translation to other bioenergy grasses. Targets for genetic modification will be identified and tested for enhanced processing ability.**

The backbone genome sequences of *Arabidopsis* and rice provided functional annotations for an estimated 1500 genes in maize that are implicated in cell wall development in grass species [1; http://cellwall.genomics.purdue.edu/]. Highly parallel pyrosequencing was used to profile gene expression at defined stages of development for the maize internode characteristic of cell division, cell enlargement, fiber differentiation, secondary wall cellulose deposition, and lignification. Algorithms were developed to determine co-expression of suites of genes associated with these stages of development and specifically for primary and secondary cell wall cellulose synthase isoforms, mixed-linkage b-glucan synthases, xylan synthases, and enzymes of the phenylpropanoid and monolignol synthesis pathways. In parallel, we have used Pyrolysis Molecular Beam MS as a high throughput means to classify cell wall architectures of maize stover from the Intermated B73 x Mo17 (IBM) recombinant inbred population. High-density mapping of these diverse populations permitted determination of several quantitative trait loci (QTL) for traits of hexose and pentose, *p*-coumaric acid, guaiacyl and syringyl lignin abundance. We are currently employing comparative expression analysis in B73, Mo17 and target IBM lines to refine candidate genes in the B73/Mo17 genomes responsible for biochemical contributions to diverse architectures. Another line of experiments is defining small RNA populations during different developmental stages. From these populations, potential regulatory sequences derived from naturally occurring antisense transcripts of candidate genes are being elucidated by strand-specific PCR.

This work is supported by the Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under Contract No. DE–FG02–08ER64702

**Reference**

1. Penning et al. (2009) Genetic resources for maize cell wall biology. *Plant Physiol.* 151, 1703-1727

# 142

## Genomic Analysis of miRNAs and Target RNAs of *Brachypodium*

Skye A. Schmidt*[1] (sschmidt@udel.edu), Dong-Hoon Jeong,[1] Linda Rymarquis,[1] Monica Accerbi,[1] Sunhee Park,[1] Marcelo A. German,[2] Matthias Ganssmann,[1] Jixian Zhai,[1] Emanuele DePaoli,[3] Sai Guna Ranjan Gurazada,[4] Blake C. Meyers,[1] and **Pamela J. Green**[1]

[1]Delaware Biotechnology Institute, University of Delaware, Newark; [2]Present address: Dow AgroSciences LLC, Portand, Ore.; [3]Present address: Istituto Agrario San Michele all'Adige, Trento, Italy; [4]Present address: DuPont Crop Genetics, Wilmington, Del.

**Project Goals: In this study, we use next-generation sequencing technology to identify miRNAs in different tissues and following abiotic stress treatments of *Brachypodium distachyon*, a rapidly developing model system for temperate grasses and bioenergy crops.**

miRNAs are small, endogenous RNAs that post-transcriptionally regulate gene expression in nearly all eukaryotic systems. In plants, miRNAs can serve as major regulators of development, stress responses, metabolism, and other processes through the miRNA-guided cleavage of specific target RNAs. While miRNAs and their target interactions are well-characterized in systems such as *Arabidopsis* and rice, little is known about their roles in others including temperate grasses and potential bioenergy crops. In this study, we use next-generation sequencing technology to identify miRNAs in different tissues and following abiotic stress treatments of *Brachypodium distachyon,* a rapidly developing model system for temperate grasses and bioenergy crops. A total of more than 64 million reads were obtained from 12 small RNA libraries, resulting in an average of more than 1.4 million distinct genome-matched small RNA sequences per library, from which both conserved and new miRNAs have been identified. To identify the targets of these miRNA on a global scale, we use an approach called Parallel Analysis of RNA Ends (PARE) that facilitates the sequencing of 3' products of miRNA-guided target RNA cleavage. Because miRNAs and their targets can form missing links in many important gene regulatory networks, the identification of miRNA and target RNA pairs in *Brachypodium* will help to better understand how small RNAs contribute to the regulation of genes and genomes.

This work was funded by grant # DE-FG02-07ER64450 of the Plant Feedstock Genomics for Bioenergy Program of the DOE.

# 143

## Insertional Mutagenesis of *Brachypodium distachyon*

**John Vogel***[1] (john.vogel@ars.usda.gov), Jennifer Bragg,[1] Jiajie Wu,[1,2] Yong Gu,[1] Gerard Lazo,[1] and Olin Anderson[1]

[1]USDA-ARS, Western Regional Research Center, Albany, Calif.; and [2]Shandong Agriculture University, Taian, Shandong, China

**Project Goals:**
**(renewal project).**

1. **Generate 30,000 insertional mutants.**
2. **Sequence DNA flanking the insertion sites of all mutants to place the insertions in a genomic context.**
3. **Collaborate with other groups creating T-DNA lines to integrate their lines into our database.**

Herbaceous energy crops, especially grasses, are poised to become a major source of energy in the United States. Despite their increasing importance, we know little about the basic biology underlying the traits that control the utility of grasses as energy crops. Better knowledge of basic grass biology (e.g. identification of the genes that control cell wall composition, plant architecture, cell size, cell division, reproduction, nutrient uptake, carbon flux, etc.) could be used to design rational strategies for crop improvement and shorten the time required to domesticate these species. The use of an appropriate model system is an efficient way to gain this knowledge. Unfortunately, due to its distant relationship to monocots, *Arabidopsis* is not suitable to study biological features unique to the grasses (e.g. cell wall composition). *Brachypodium distachyon* (*Brachypodium*) is a small annual grass with all the attributes needed to be a modern model organism including simple growth requirements, fast generation time, small stature, small genome size and self-fertility. These attributes led to the recommendation in the DOE's "Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda" report to propose developing and using *Brachypodium* as a model for energy crops to accelerate their domestication. Strategic investments in *Brachypodium* by the DOE are now bearing fruit and *Brachypodium* is rapidly being adopted as the grass model of choice by hundreds of laboratories worldwide. The DOE-JGI has recently sequenced the entire *Brachypodium* genome. The genome sequence and annotation are of unprecedented quality for a draft plant genome and will serve as a firm foundation for a host of functional genomic tools. *Brachypodium* is also readily transformed by *Agrobacterium tumefaciens.* Indeed, with average transformation efficiencies around 50% *Brachypodium* is now, arguably, the easiest grass to transform. Other resources available to *Brachypodium* researchers include a large germplasm collection, resequenced genomes, microarrays, molecular markers, a high-density genetic map, BAC libraries and physical maps. In addition key protocols have been optimized for *Brachypodium* including: efficient crossing, chemical mutagenesis, radiation mutagenesis and live root imaging.

Sequence indexed insertional mutants are an extremely powerful tool for both forward and reverse genetics. We have begun to create a collection of *Brachypodium* T-DNA mutants. Our T-DNA tagging project was initially funded in 2007 and was renewed in 2010. Using our high-efficiency *Agrobacterium tumefaciens*-mediated transformation method, we have generated 8,700 *Brachypodium* $T_0$ lines. We sequenced the DNA flanking the insertion sites in 7,111 lines and assigned 8,107 of the resulting flanking sequence tags (FSTs) in 4,393 insertional mutant lines to 5,348 unique locations in the *Brachypodium* genome. These include 1,559 insertions in genes and 1,039 insertions close to genes (1,000 bp upstream or 500 bp downstream). Information about the WRRC *Brachypodium* insertional mutant population is available in a searchable website designed to allow researchers to order T-DNA lines with mutations in genes of interest. Protocols for working with *Brachypodium*, information about the T-DNA project, and instructions for ordering T-DNA lines are available at http://brachypodium.pw.usda.gov. We have just completed the first cycle of funding and significantly exceeded our key objectives. We created 1,200 more lines and sequenced the flanking DNA in 1,111 more lines than planned. The two postdocs working on this project have both moved to new positions: one is doing research in China and the other is now working on our related phenomics project. A new postdoc and technician have just been hired to continue the project. The goal of the renewal project is to generate another 30,000 T-DNA lines. We have also established collaborations with eight laborato-

ries from five countries to create an international *Brachypodium* T-DNA collection and, together with our collection, we have plans to create a collection of 65,000 T-DNA lines.

# 144

## Characterization of Novel Cell Wall Mutants in Maize

F.J. Kraemer,[1] T. Thomik,[1] S. Hake,[3] and **M. Pauly**[1,2]

[1]Energy Biosciences Institute, University of California-Berkeley; [2]Department of Plant and Microbial Biology, University of California-Berkeley; and [3]Plant Gene Expression Center, USDA-ARS, Albany, Calif.

In recent years plant cell walls and the polymers that constitute them have received increased attention as a potential highly abundant renewable resource for biofuel production. The feasibility of using plant feedstocks will in part be dependent on the optimization of plant wall composition, as it can directly influence conversion yield. Therefore the understanding of plant cell wall polymer biosynthesis and metabolism and the genes involved therein will be essential. Multiple members of the *Poales* order have been proposed as potential bioenergy feedstocks as they combine multiple desirable traits such as C4 photosynthesis, large biomass yield and fast growth. Examples include switchgrass, *Miscanthus* or sugar cane. Also crop residues like corn stover or wheat straw could be utilized.

We performed a forward genetic screen to identify mutants with alterations in their cell wall monosaccharide composition. Mutaganized lines of *Zea mays* (chemical mutagenesis) were analyzed and multiple lines with altered monosaccharide composition have been identified. A summary of the screen and detailed data on promising maize candidates will be presented and discussed.

# 145

## Identification of Candidate Genes Using Rice Mutants for Biomass Engineering In Switchgrass

B. Broeckling,[1] M. Baroidan,[2] C.E. Jahn,[1] **J.K. McKay**,[1] **J.E. Leach**[1]* (Jan.Leach@colostate.edu), **D.R. Bush**,[1] and **H. Leung**[1]

[1]Program in Plant Molecular Biology, Colorado State University, Fort Collins; and [2]Plant Breeding, Genetics and Biotechnology Division, International Rice Research Institute (IRRI), Metro Manila, Philippines

**Project Goals: Our goal is to provide the applied biomass research community and industry with information to allow exploitation of the genes and pathways relevant to biomass accumulation in grasses. The specific objectives for this project are to: Objective 1: Identify genes involved in biomass accumulation. Using genetic populations for rice lines that exhibit extremes in biomass accumulation, we will (a) map QTL using a comprehensive phenotyping/genotyping approach, (b) confirm the QTL location and phenotypic effect and (c) screen a deletion mutant collection to identify large deleted regions corresponding to biomass accumulation. Objective 2: Dissect the QTL using an integrated analysis. First, we will use an expedited approach to generating near isogenic lines containing each QTL. Second, we will integrate genome-wide data to refine the QTL regions including (a) association mapping, (b) expression profiling, (c) mutant analysis, (d) fine scale mapping, and (e) comprehensive sequencing.**

Developing a sustainable biofuels program that makes significant contributions to the national energy budget requires unprecedented inputs of biomass for energy conversion. Our specific objectives are to identify genes and pathways in rice (*Oryza sativa*) that increase plant biomass in order to translate this information into cultivar improvement for new energy crops, such as switchgrass. We have screened over 12,000 chemical and irradiation-induced mutants of rice and identified 13 mutants with increased biomass. Our initial effort focused on one diepoxybutane mutant (DEB1) that has reproducibly shown a two-fold increase in biomass under both field and greenhouse conditions. Diepoxybutane is predicted to cause small deletions in the genome and to identify the deletions in DEB1, we performed a comparative genomic hybridization experiment using an *Oryza sativa* whole genome array. We identified 25 deletions in DEB1, ranging in size from 90 to 5,721 bps. Identification of the deletion responsible for the high biomass phenotype is currently in progress. Comparative genomic hybridization, plus the incorporation of new approaches such as short read whole genome sequencing, will be used to identify candidate genes from the remaining 12 mutants. The contribution of candidate genes to biomass improvement will be validated in both rice and switchgrass using a transgenic approach. Plants with perturbations in these candidate genes will be comprehensively phenotyped by transcript and physiological profiling. This systems biology approach will enable identification of key regulatory networks relevant to bioenergy traits for further rational engineering of switchgrass.

# 146

## Identification of Genes That Regulate Phosphate Acquisition and Plant Growth During Arbuscular Mycorrhizal Symbiosis in *Brachypodium distachyon*

Jeon J. Hong,[1] S. Karen Gomez,[1] Quan Zhang,[1] Gopala Battu,[2] Liudmila S. Yafremava,[2] Kranthi Varala,[2] Matthew Hudson,[2] and Maria J. Harrison[1]* (mjh78@cornell.edu)

[1]Boyce Thompson Institute for Plant Research, Ithaca, N.Y.; and [2]Department of Crop Science, University of Illinois, Urbana

**Project Goals: The goals of this part of our project are to establish *Brachypodium distachyon* as a model to study AM symbiosis in grasses of relevance as bioenergy crops and to use this system to dissect the molecular basis of differences in functionality in different AM symbioses.**

Most vascular flowering plants are able to form symbiotic associations with arbuscular mycorrhizal (AM) fungi. The symbiosis develops in the roots and has a profound effect on plant productivity, largely through improvements in plant mineral nutrition (Smith and Read, 2008). All proposed bioenergy crops including legumes, grasses and trees, are capable of forming AM symbioses, and therefore have the potential to benefit from phosphorus and nitrogen acquisition through the symbiosis. This is significant because phosphorous and nitrogen are the two mineral nutrients whose availability most frequently limits plant growth (Vance, 2001; Vance et al., 2003; Tilman et al., 2006)

*Brachypodium distachyon* is a wild grass species that serves as a model for the temperate grasses, including those proposed as bioenergy crops. There are several detailed ecological studies of AM symbioses of grasses, including studies of the genus *Brachypodium* (van der Heijden et al., 2003; van der Heijden et al., 2006), and these indicate the potential of the AM symbiosis for increasing plant growth in low phosphate soils. However, these studies also illustrate significant differences in plant performance depending on the AM fungal species involved. Variation in plant performance during symbiosis with different AM fungal symbionts is a well documented phenomenon; however, the molecular basis is not understood.

The goals of this part of our project are to establish *Brachypodium distachyon* as a model to study AM symbiosis in grasses of relevance as bioenergy crops and to use this system to dissect the molecular basis of differences in functionality in different AM symbioses. Sequence-based transcript profiling transcript will be used to document transcript profiles in *B. distachyon* AM symbioses that differ with respect to plant performance. From these datasets, transcript profiles and potentially plant processes associated with maximal plant performance will be identified.

Towards these goals, growth systems that enable development of *B. distachyon*-AM symbioses have been established and the interactions of *Brachypodium distachyon* with *Glomus versiforme, Glomus intraradices, Glomus white, Gigaspora gigantea* and *Gigaspora decipiens* have been assessed. These AM fungi all colonize *B. distachyon* roots and establish symbiosis, but their effects on growth of *B. distachyon* and on mineral nutrition, vary significantly.

In *Medicago truncatula* and rice, array-based transcript profiling has been used to document the transcriptional responses to development of AM symbiosis and sets of AM symbiosis-induced genes have been identified (Liu et al., 2003; Güimil et al., 2005; Liu et al., 2007; Gomez et al., 2009). We identified *B. distachyon* orthologs of rice and *M. truncatula* AM-symbiosis induced genes and then examined their expression in *B. distachyon* during symbiosis with different AM fungi. While the expression patterns of the 'AM-specific genes' are conserved in *M. truncatula*, rice and *B. distachyon*, there were some surprising differences in AM-symbiosis induced gene expression patterns in *B. distachyon* relative to rice. For example, expression of a rice peroxidase, OSAM1, is highly induced in rice-AM symbiosis but the predicted ortholog is not induced in *B. distachyon* AM symbiosis. To extend these profiling analyses, Illumina-based transcript profiling of *B. distachyon* during AM symbiosis with three AM fungi is in progress.

**References**


1. Gomez, S.K., Javot, H., Deewatthanawong, P., Torrez-Jerez, I., Tang, Y., Blancaflor, E.B., Udvardi, M.K., and Harrison M, J. (2009). *Medicago truncatula* and *Glomus intraradices* gene expression in cortical cells harboring arbuscules in the arbuscular mycorrhizal symbiosis. *BMC Plant Biology* 9, 1-19.
2. Güimil, S., Chang, H.-S., Zhu, T., Sesma, A., Osbourn, A., Roux, C., Ioannidis, V., Oakeley, E.J., Docquier, M., Descombes, P., Briggs, S.P., and Paszkowski, U. (2005). Comparative transcriptomics of rice reveals an ancient pattern of response to microbial colonization. *Proceedings of the National Academy of Sciences of the United States of America* 102, 8066-8070.
3. Liu, J., Maldonado-Mendoza, I.E., Lopez-Meyer, M., Cheung, F., Town, C.D., and Harrison M, J. (2007). The arbuscular mycorrhizal symbiosis is accompanied by local and systemic alterations in gene expression and an increase in disease resistance in the shoots. *The Plant Journal* 50, 529-544.
4. Liu, J., Blaylock, L., Endre, G., Cho, J., Town, C.D., VandenBosch, K., and Harrison, M.J. (2003). Transcript profiling coupled with spatial expression analyses reveals genes involved in distinct developmental stages of the arbuscular mycorrhizal symbiosis. *Plant Cell* 15, 2106-2123.
5. Smith, S.E., and Read, D.J. (2008). *Mycorrhizal Symbiosis.* (San Diego, CA: Academic Press, Inc.).
6. Tilman, D., Hill, J., and Lehman, C. (2006). Carbon-negative biofuels from low-input high-diversity grassland biomass. *Science* 314, 1598-1600.
7. van der Heijden, M.G.A., Wiemken, A., and Sanders, I.R. (2003). Different arbuscular mycorrhizal fungi alter coexistence and resource distribution between co-occurring plant. *New Phytol.* 157, 569-578.
8. van der Heijden, M.G.A., Streitwolf-Engel, R., Riedl, R., Siegrist, S., Neudecker, A., Ineichen, K., Boller, T., Wiemken, A., and Sanders, I.R. (2006). The mycorrhizal contribution to plant productivity, plant nutrition and soil structure in experimental grassland. *New Phytol.* 172, 739-752.
9. Vance, C. (2001). Symbiotic Nitrogen Fixation and Phosphorus Acquisition. Plant Nutrition in a World of Declining Renewable Resources. *Plant Physiol.* 127, 390-397.
10. Vance, C.P., Uhde-Stone, C., and Allan, D.L. (2003). Phosphorus acquisition and use: critical adaptations by plants for securing a nonrenewable resource. *New Phytol.* 157, 423-447.

# 147

## Sucrose Transporter Genes of *Populus*: An Investigation of Their Importance as Regulators of Biomass and Carbon Partitioning in Trees

**Scott A. Harding*** (sharding@uga.edu), **C.J. Tsai**, Christopher J. Frost, Duo Peng, Liangjiao Xue, and Kate H.C. Tay

Warnell School of Forestry and Natural Resources, and Department of Genetics, University of Georgia

http://aspendb.uga.edu

**Project Goals: see below**

Sucrose is the long-distance transport form of carbon in most temperate zone plants. Its export from source organs and subsequent trafficking among various sink organs, including sites of lignocellulosic biomass accrual, depend on the activity of trans-membrane sucrose transporters (SUT).

**Project Goals: We seek to learn how SUT proteins function, individually and in cooperation with one another, to facilitate:**

1. **the export of sucrose from source leaves.**
2. **the retention, trafficking and utilization of sucrose for lignocellulose production in wood tissues.**

Gene profiling and network analysis as well as metabolite profiling will be used to investigate SUT function with regard to water use, constitutive defense and biomass growth/composition in *Populus*.

**Basic Approach:**
Transgenic manipulation of the relative expression of three *Populus SUT* genes in a tissue-specific manner will be used in combination with drought and defoliation/stem girdling treatments. The overarching plan is to vary the ratio of SUT-mediated supply to SUT-mediated trafficking and sink retrieval processes within the plant .

**Rationale:**
The use of wood as a lignocellulosic feedstock for bioenergy is expected to be accompanied by expanded cultivation of woody species in agriculturally marginal sites. This means that both the yield and composition of harvestable feedstocks will be subjected to environmental constraints that affect the biosynthesis and subsequent utilization of photoassimilate (sucrose). The importance of SUT as an index or marker of tree productivity remains under-investigated. One benefit of proposed research will be to obtain a thorough understanding of SUT function in the context of other more examined gene markers (*e.g.*, sucrose, starch and lignin biosynthetic pathway enzymes). We also expect to learn about the degree to which the distinctive chemical milieu of *Populus* may factor into SUT protein function. Leaves and stems of *Populus* and *Salix* are rich in defensive metabolites, the salicin-derived phenolic glycosides (PGs) that do not occur in other agriculturally important species. Salicin is a substrate and potential inhibitor of SUT proteins. Environmental constraints brought on by limiting nutritional and water status expected to be characteristic of potential bioenergy plantations are therefore likely to perturb PG homeostasis, and possibly sucrose utilization in *Populus*.

**Progress:**
Phylogenetic analysis of the amino acid sequences classified the *Populus* SUT family into the three major groups characteristic of other dicots. Group-1 *PtaSUT3* gene transcripts were localized to leaf vascular traces and stem developing xylem; Group-4 *PtaSUT 4* to leaf spongy mesophylls, stem cambium, developing xylem and phloem; Group-2 *PtaSUT5/6* to all leaf cells, stem developing xylem and phloem fibers. Based on these data, PtaSUT4 is a much more important regulator of sucrose transport in *Populus* than it is in herbaceous annuals. Subcellular localization has been carried out to confirm that in contrast to other SUT proteins, PtaSUT4 is tonoplast-localized. Additional GFP-SUT fusion constructs have been assembled to investigate the subcellular localization of the other SUT proteins. SUT4-RNAi transgenic plants demonstrated a shift of biomass allocation from stem to leaf in both nitrogen (N)-replete and N-limited plants. In those plants, sucrose exhibited a complex pattern of hyper-accumulation in exporting leaves and vascular tissues of the stem, with a slight decrease in the shoot tip and sink leaves. RNAi silencing of SUT4 reduced water uptake from root tissues during drought simulation. Significant RNAi effects on secondary metabolite accumulation and on the transcript levels of carbohydrate-related genes were observed in exporting source leaves.

# 148

## Identifying Genes Controlling Feruloylation in Grass Cell Walls

**Marcia Maria de Oliveira Buanafina**[1]* (mmb26@psu.edu), and Mandeep Sharma[1]

[1]The Pennsylvania State University, Biology Department, University Park

**Project Goals: The focus of this proposal is to identify and characterize new genes controlling feruloylation in grasses, as well as new genes that are responsible for the assembly of lignin into the cell wall and for biomass conversion. This will provide fundamental knowledge concerning the most crucial factors that influence grass cell wall degradability.**

Cellulosic biomass—particularly from grasses—is expected to supply a large and renewable source of biomass for the production of renewable biofuels. However, utilization of grass cell walls for biofuel production is impeded greatly by ferulic acid residues which are ester linked to arabinoxylans (AX) and have the ability to form ferulate dimers functioning in cell wall cross-linking. Such cross-linking is among

the main factors inhibiting the release of fermentable carbohydrates from grasses for biofuel production.

We have shown previously that the expression of ferulic acid esterase (FAEA) in different grass species resulted in a substantial reduction in cell-wall-esterified ferulates and diferulates and increased cell wall hydrolysis. Controlling the level of total feruloylation should have a direct impact on the level of cross-linking and thereby on cell wall degradation.

Currently, the genes underlying AX feruloylation have not been identified and the isolation of such genes could be of great importance in manipulating ferulates accretion to the wall. Mutation of the feruloyl transferase gene(s) should lead to less ferulates secreted to the cell wall and reduced ferulate cross-linking.

We have developed EMS mutagenized populations of model grass species *Brachypodium distachyon* with EMS to be used as a resource for identification of the genes involved in feruloylation, synthesis of the xylan backbone and new genes that are responsible for the assembly of lignin into the cell wall. We have used spectrophotometric, microscopy, HPLC and HPIEC screening techniques to select for new genetic variation in *Brachypodium*.

EMS populations were developed from over 28,000 mutagenized seeds generating over 5,000 M2 families. A total of 12,793 plants have been screened and 1,233 have been selected. Here we report on the potential mutants with *–altered levels of cell wall ferulates, lignification and cell wall AX* – that have been selected.

We also report on the considerable variation on the level of cell wall ferulates, AX, ligninification and cellulase mediated release of sugars, among different *Brachypodium* accessions to be selected to generate our mapping population.

# 149

## Genetic Dissection of Bioenergy Traits in Sorghum

**Wilfred Vermerris**[1,*] (wev@ufl.edu), Ana Saballos,[1] Terry Felderhoff,[2] Sharon Mitchell,[3] **William Rooney**,[2] **Seth Murray**,[2] **Stephen Kresovich**,[3,4] **Jeffrey F. Pedersen**,[5] **Scott Sattler**,[5] and Zhanguo Xin[6]

[1]Genetics Institute and Agronomy Department, University of Florida, Gainesville; [2]Dept. of Soil and Crop Sciences, Texas A&M University, College Station, Tex.; [3]Institute for Genomic Diversity, Cornell University, Ithaca, N.Y.; [4]Office of Research, University of South Carolina, Columbia; [5]Grain, Forage, and Bioenergy Research, USDA-ARS, Lincoln, Nebr.; and [6]USDA-ARS, Lubbock, Tex.

http://www.sorghum.genome.ufl.edu

**Project Goals: The goals of this project are to gain better understanding of the genetic basis of both sugar accumulation and cell wall biosynthesis in (sweet) sorghum, in order to facilitate development of improved germplasm. This is accomplished by identifying the gene(s) underlying a quantitative trait locus (QTL) for stem sugar concentration, by identification of QTL for juice volume, by characterizing novel brown midrib mutants, and by cloning Brown midrib genes.**

Sorghum is an attractive biomass crop for ethanol production because of its low water and fertilizer requirements, tolerance to heat and drought, and high biomass yield. Because of the species' great genetic diversity (Murray et al. 2009), and the fact that sorghum is a diploid, seed-propagated crop, development of cultivars and hybrids adapted to a wide range of environments is feasible. Sweet sorghums are sorghums that can reach heights of up to 6 m and that accumulate soluble sugars in their stems. After squeezing the stalks, these sugars can be fermented directly and conveniently to ethanol or other biofuels. The crushed stems (bagasse) can then be processed as lignocellulosic biomass. Sweet sorghum thus represents an ideal bridge between sugar-based and cellulosic fuels, and, given the rapid establishment of sweet sorghum, this species is expected to be of particular value in extending the processing window of sugarcane-based biorefineries (Vermerris, 2011).

In order to expand the area where sweet sorghum can be produced, both in terms of geographic location (daylength, temperature, pests and diseases) and local conditions (soil quality, water and nutrient availability), regionally adapted cultivars and hybrids need to be developed. **The goals of this project are to gain better understanding of the genetic basis of both sugar accumulation and cell wall biosynthesis, in order to facilitate development of improved germplasm.**

Quantitative trait loci (QTL) associated with sugar concentration of the juice were identified in a recombinant inbred line population derived from the sweet sorghum 'Rio' and the grain sorghum BTx623 (Murray et al. 2008). A major QTL for sugar concentration is located on chromosome 3. We are employing high-throughput transcriptome profiling using the Solexa next-generation sequencing platform to identify the gene(s) underlying this QTL. This approach relies on comparing gene expression profiles of heterogeneous inbred families that are genetically highly similar except for the region containing the QTL. RNA was extracted from several different tissues and developmental stages and expression data are in the process of being analyzed. In addition, we have mapped QTL for juice volume, using a population of recombinant inbred lines derived from the dry-stem, non-sweet grain sorghum BTx3197 and the sweet sorghum 'Rio'. Novel germplasm with an overall higher sugar yield can be developed by combining QTL (and ultimately loci) controlling juice volume and juice concentration.

In order to improve the biomass-to-fuel conversion, we are focusing on *brown midrib* (*bmr*) mutants. The *bmr* mutations change the color and the chemical composition of the vascular tissue. Four independent loci were identified by Saballos et al. (2008) in a collection of mutants first described by

Porter et al. (1978). Additional *bmr* mutants were identified in the TILLING population of Xin et al. (2008). Several *bmr* mutants from both populations have been shown to result in enhanced yields of fermentable sugars following enzymatic saccharification of sorghum biomass, even after thermochemical pretreatment (Saballos et al. 2008; Dien et al., 2009; Pedersen et al.; *in preparation*). As part of this project we have also cloned the *Bmr6* and *Bmr2* genes. The *Bmr6* gene encodes the monolignol biosynthetic gene cinnamyl alcohol dehydrogenase (CAD) (Saballos et al. 2009; Sattler et al. 2009). The *Bmr2* gene also encodes a cell wall biosynthetic enzyme (Saballos et al.; *in preparation*). Knowing the identity of the *Bmr* genes and the nature of the mutations in these genes has enabled the development of allele-specific markers that will allow more efficient use of these mutations in commercial breeding programs.

Funding from the Office of Science (BER), U.S. Department of Energy grant DE-FG02-07ER64458 is gratefully acknowledged.

**References**

1. Dien BS, Sarath G, Pedersen JF, Sattler SE, Chen H, Funnell-Harris DL, Nichols NN, and MA Cotta (2009) Improved sugar conversion and ethanol yield for forage sorghum (*Sorghum bicolor* (L.) Moench) lines with reduced lignin contents. Bioenerg. Res. 2: 153-164
2. Murray SC, Rooney WL, Hamblin MT, Mitchell SE and S Kresovich. 2009. Sweet sorghum diversity and association mapping for brix and height. Plant Genome 2:48-62.
3. Murray SC, Sharma A, Roone WL, Klein PE, Mullet JE, Mitchell SE and S Kresovich. 2008. Genetic improvement of sorghum as a biofuel feedstock I: quantitative loci for stem sugar and grain nonstructural carbohydrates. Crop Sci. 48:2165-2179.
4. Porter KS, Axtell JD, Lechtenberg VL, and VF Colenbrander (1978) Phenotype, fiber composition, and in vitro dry matter disappearance of chemically induced *brown midrib* (*bmr*) mutants of sorghum. Crop Sci. 18: 205-209
5. Saballos A, Ejeta G, Sanchez E, Kang CH, and W Vermerris. 2009. A genomewide analysis of the cinnamyl alcohol dehydrogenase family in sorghum [*Sorghum bicolor* (L.) Moench] identifies *SbCAD2* as the *Brown midrib6* gene. Genetics 181: 783-795.
6. Saballos A, Vermerris W, Rivera L, and G Ejeta. 2008. Allelic association, chemical characterization and saccharification properties of *brown midrib* mutants of sorghum (*Sorghum bicolor* (L.) Moench). BioEnerg. Res. 2: 193-204
7. Sattler S, Saathoff AJ, Haas EJ, Palmer NA, Funnell-Harris DL, Sarath G, and JF Pedersen. 2009 A nonsense mutation in a *cinnamyl alcohol dehydrogenase* gene is responsible for the sorghum *brown midrib 6* phenotype. Plant Physiol. 150:584-595.
8. Vermerris W (2011) Survey of genomics approaches to improve bioenergy traits in maize, sorghum and sugarcane. J. Integr. Plant Biol. 53: 105-116
9. Xin Z., Wang ML., Barkley NA, Burow G, Franks C, Pederson G, and J Burke. 2008 Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. BMC Biol. 8: 103

# 150

## Linkage and QTL Mapping of Switchgrass (*Panicum virgatum* L.) Using EST-Microsatellites and Their Use for Comparison Within the Poaceae

Christian Tobias[1]* (christian.tobias@ars.usda.gov), Miki Okada,[1] Hugh Young,[1] Christina Lanzatella-Craig,[1] Brindha Narasimhamoorthy,[2] Malay Saha,[2] Guifen Li,[2] Rongling Wu,[3] and Joe Bouton[2]

[1]USDA-ARS Western Regional Research Center, Albany, Calif.; [2]Samuel Roberts Noble Foundation, Inc., Forage Improvement Division, Ardmore, Okla.; and [3]Department of Public Health Sciences, Pennsylvania State University, Hershey

**Project Goals: 1. Develop a comprehensive molecular marker system for switchgrass. 2. Develop a saturated microsatellite linkage map of switchgrass. 3. Identification of markers associated with biomass yield, digestibility, maturity, and other traits and identification of associated markers.**

Switchgrass is widely viewed as a promising crop for bioenergy production. However, development of improved cultivars optimized for bioenergy through breeding involves improving yields and altering feedstock composition so that competition for limited arable land is minimized and process efficiencies are fully realized. Fundamental to any advanced breeding program are availability of molecular markers and genetic linkage maps that facilitate modern cultivar development through marker assisted selection (MAS). New crops such as switchgrass stand to benefit from the application of MAS techniques and through comparative approaches with other grasses that will provide a multitude of candidate gene-loci for traits considered important for bioenergy. Difficulties with these approaches are encountered due to polyploidy and outcrossing in switchgrass and must be accounted for during the mapping process.

To determine marker/QTL linkage phase with certainty during mapping, a four grandparent three-generation population using parents and grandparents from both upland and lowland ecotypes has been created consisting of 188 individuals. These were evaluated over two seasons at two locations in Oklahoma following an R-256 honeycomb design. Continuous variations for biomass and related traits were observed in the population. Plant height varied between 129 – 268 cm with the mean of 201 cm. The most productive plant produced 5.23 kg of fresh biomass at the second year harvest. Striking variations were observed for both Ca and K content. Dry matter digestibility ranged between 28.2 - 48.8% with the mean of 37.6%. QTLs associated with the traits of interest will be detected through genotyping with EST-SSR, STS, and genomic SSR markers.

As part of the larger goals of the funded project we used a full-sib mapping population derived from one pair of grandparents in a pseudo-backcross mapping strategy. This

resulted in linkage maps from each individual grandparent that were found to be highly collinear. The male and female framework map lengths and number of both framework and accessory markers were 1376 cM and 563 in the female map and 1645 cM and 542 in the male map, with 97% of the genome estimated to be within 10 cM of a mapped marker in both maps. Consistent with previous cytological data and existing initial linkage studies, there was nearly exclusive preferential pairing. The map resulting from the pollen donor was more affected by transmission ratio distortion than the female parent. Extensive collinearity with sorghum was apparent with the differences in base chromosome number appearing to result from the fusion of sorghum chromosomes 8 and 9. Sub-genome comparisons within the nine switchgrass homology groups as well as interspecific comparisons are now possible using CMAP (www.gramene.org) and all raw mapping data has been published as supplemental data.

A point that remains unclear is the origin of polyploidy in switchgrass. Autopolyploids originate through whole-genome duplication events while allopolyploids arise through hybridization of distinctly different species. Though we have thus far seen evidence for disomic inheritance this does not provide conclusive evidence supporting either mode of origin. To provide further support for one or the other modes of origin, we have performed fluorescence *in-situ* hybridization (FISH) using centromere-specific repeats, and 45S rDNA sequences in dihaploid individuals. These individuals appear to have lost one copy of each subgenome and at meiosis 18 univalents were observed. The plants are functionally sterile, though partially fertile tillers have arisen that have undergone at least partial chromosome doubling. These dihaploids, represent simplified genetic systems for karyotyping and will provide a useful reference genome allowing direct comparisons of subgenome differentiation. Our attempts to unambiguously discriminate individual chromosomes in these dihaploids will be presented.

# 151

## Molecular Mechanisms of Carbon Partitioning by *cpg13* in Poplar

**Matias Kirst*** (mkirst@ufl.edu) and **Gary F. Peter**

School of Forest Resources and Conservation, University of Florida Genetics Institute, Gainesville

**Project Goals: We have identified the *cpg13* gene (carbon partitioning and growth in LG13) as a key regulator of carbon partitioning to lignin and cellulose, and whole plant biomass productivity in a segregating *Populus* hybrid population. We hypothesize that *cpg13* regulates the metabolic competition for carbon, affecting growth, cellulose biosynthesis and lignification. To verify this hypothesis and begin addressing the molecular function of *cpg13*, our goals are to: a. Characterize *cpg13* function in the regulation of gene expression, metabolites, and cell wall chemistry and structure, and b. Determine the spatial and temporal expression and subcellular localization of *cpg13* to advance our understanding of its molecular role.**

Lignin is one of the primary obstacles for efficient bioconversion of wood cellulosics to renewable fuels, and its content is directly and inversely proportional to fermentable sugar yield. In woody species, lignin content is often negatively correlated with biomass productivity. This relationship has been observed in segregating populations of *Populus* and *Eucalyptus*, and in breeding populations and a natural mutant of loblolly pine. In a previous project funded by the Department of Energy ("Genomic Mechanisms of Carbon Allocation and Partitioning in Poplar" – ER64114-1026645-0011741) we identified a major gene involved in the regulation of carbon partitioning to lignin and cellulose, and whole plant biomass productivity. The gene, referred hereafter as *cpg13* (carbon partitioning and growth locus in chromosome XIII), was identified by applying a genetical genomic approach that combines fine-mapping of quantitative trait loci and transcriptome analysis of a *Populus deltoides* and *Populus trichocarpa* hybrid population. However, the functional role of *cpg13* is unknown.

**The primary goal of this project is to dissect the molecular mechanism by which *cpg13* regulates carbon partitioning and growth.** To uncover the mechanism of carbon regulation we are: (1) characterizing *cpg13* function in the regulation of gene expression, metabolites, and cell wall chemistry and structure, and (2) determining the spatial and temporal expression and subcellular localization of *cpg13* to advance our understanding of its molecular role.

To achieve these goals we have generated 20 transgenic lines, over-expressing (OE) and down-regulating the expression of *cpg13*. For down-regulating *cpg13* we developed two RNAi constructs, that target (1) the short N-terminus and (2) the C-terminus, which contains a domain of unknown function (DUF). No transgenic lines could be recovered for the second construct, suggesting that it interferes with expression of other genes that contain the DUF domain. Three to five biological replicates were grown in a greenhouse for all viable transgenic lines during 2010. Instead of proceeding to immediately characterize the transcriptome, metabolome, and wood growth and properties, as originally planned, we chose to narrow the selection to 5 OE and RNAi expressing lines. These lines are now being cloned to obtain 20+ biological replicates, for precise estimates of plant biomass growth properties.

In addition to the analysis of plants over-expressing and down-regulating the expression of *cpg13* we are taking two complementary approaches to determine what cell types express *cpg13* and its subcellular localization. In the first approach, green fluorescent protein (GFP) has been fused in frame at its carboxy terminus to *cpg13* and its native promoter, and the construct has been transformed into poplar. The transgenic plants are being used to analyze expression in different organs, cell types and ontogenetic stages. Overall, initial observations indicate the expression of *cpg13* in lignifying vascular tissues (Figure 1). In the second approach, we are developing antibodies that specifically recognize *cpg13* and use them to confirm protein abundance,

subcellular localization by transmission electron microscopy and biochemical fractionation studies. *Cpg13* has been expressed and efforts are currently under way to purify it.

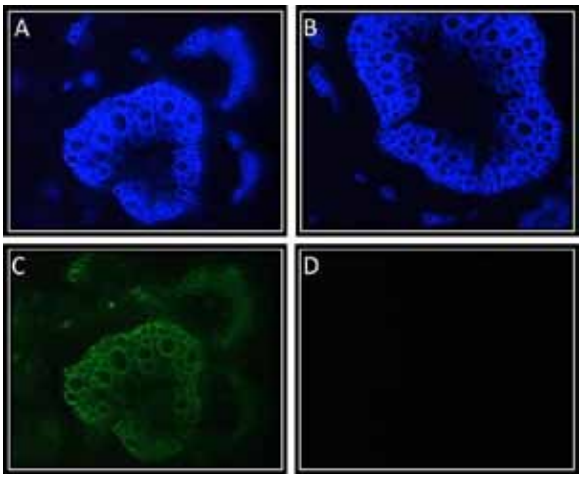



Figure 1. Lignin localization and expression of cpg*13* in poplar petioles (30 micrometer) . Lignin UV-Auto-Fluorescence in (A) cpg13::cpg13:GFP (B) wild-type. GFP fluorescence in (C) cpg13::cpg13:GFP (D) wild-type (negative control). Magnification at 20× and exposure time of 3.54s.

# 152

## The Role of Small RNA in Biomass Deposition and Perenniality in Andropogoneae Feedstocks

**Matthew Hudson*** (mhudson@illinois.edu)

University of Illinois

**Project Goals: We have already shown that a large number of new small RNA species are expressed in *Miscanthus*. We will investigate the extent to which these species target genes known to be involved in biomass production, and how much they contribute to development and variability of biomass. In particular, we will identify small RNA elements that regulate deposition of the cell wall of cells in the plant stem, a critical process in biomass production. We will also investigate the involvement of small RNA in processes that underlie traits critical to regional biomass production, for example flowering and maturity time and overwintering.**

The small RNA (sRNA) transcriptome is a complex network that regulates gene expression, antiviral responses and chromatin remodeling, and controls the self-spread of mobile elements. We employed Illumina deep sequencing of sRNAs, mRNAs, and sRNA-mediated cleavage products of mRNA (the acmRNA or degradome) to probe the role of sRNA in growth, development and biomass deposition in *Miscanthus X giganteus* (Mxg). This perennial allo-triploid grass species has significant potential as a bioenergy crop. Using the genomes of *Sorghum*, maize, and rice as references, we predicted 120 new miRNA families from Mxg that are also conserved in other grasses, and identified 61 Mxg orthologs of known miRNA from other species. For 261 mRNAs computationally predicted to be targeted by miRNAs, miRNA-directed cleavage was observed experimentally. This result validates targets of many the newly discovered miRNAs and shows some mRNAs are cleaved by multiple miRNAs. The majority of miRNAs were differentially expressed to a statistically significant degree among the sampled tissues or organs. In several cases strong negative correlation was observed between miRNA and target mRNA expression patterns, likely indicating a predominant role for the miRNA in tissue-specific mRNA expression. The majority of new Mxg miRNAs were present at highest levels in the inflorescence and rhizome tissues that contribute to a perennial growth habit. Similarly, many miRNAs identified in stem tissues are associated with the regulation of cell wall biology, targeting members of the expansin, β-tubulin, and callose synthase gene families in a tissue-specific manner. Our results suggest the existence of several previously-undescribed miRNAs that control expression of genes likely contributing to perenniality and biomass deposition in grass species.

# 153

## Development of Genomic Tools to Improve Prairie Cordgrass (*Spartina pectinata*), a Highly Productive Bioenergy Feedstock Crop

Jose L. Gonzalez, Kristene Gedye, Vance Owens, and Arvid Boe

South Dakota State University

**Project Goals: Development of SSR markers in prairie cordgrass and initiate linkage map development**

Prairie cordgrass (*Spartina pectinata*) is a tall (180-250 cm) robust rhizomatous perennial grass native to the prairies of North America, grows well in a wide range of conditions, including wet and dry marginal lands, as well as salty soils. Natural populations of PCG can be found as far north as 60°N, making this species ideal for cultivation in the Great Plains of North America. Prairie cordgrass is a C4 species with a wide ecological amplitude especially acclimated to low temperatures allowing early growth in the spring. The breeding efforts in this species have been only recently developed. With the goal of developing a molecular breeding tools we have identified over 1,000 SSR loci derived from both genomic and EST sequences.

EST sequencing has yielded the first transcriptome for this species consisting of 26,302 contigs derived from 454 reads. Selected SSR markers have been used to develop an initial linkage map using a population of 94 individuals. Other genetic and genomic resources are being developed including a clonal germplasm collection, BAC library, additional ESTs sequencing and tissue culture and genetic transformation protocols.

# 154

## Transcriptome Profiling in *Brachypodium distachyon*

**Todd Mockler** * (tmockler@cgrb.oregonstate.edu)

Oregon State University

**Project Goals: The project goals were to design a whole genome DNA microarray for the model grass *Brachypodium distachyon* and to use the array platform for expression profiling in *Brachypodium* under a variety of environmental conditions and in various tissues and developmental stages.**

*Brachypodium distachyon* is the premier experimental model grass platform and is related to candidate feedstock crops for bioethanol production. Based on the recent DOE-JGI *Brachypodium* Bd21 genome sequence and annotation we designed a whole genome DNA microarray platform. The quality of this array platform is unprecedented due to the exceptional quality of the *Brachypodium* genome assembly and annotation and the stringent probe selection criteria employed in the design. We worked with members of the international community and the bioinformatics/design team at Affymetrix at all stages in the development of the array. We have used the *Brachypodium* arrays to interrogate the transcriptomes of plants grown in a variety of environmental conditions including diurnal and circadian light/temperature conditions and a variety of light qualities. We have also examined the transciptional responses of *Brachypodium* seedlings subjected to various abiotic stresses including heat, cold, salt, and high intensity light. We are also generating a gene expression atlas representing various organs and developmental stages. The results of these efforts including all microarray datasets are available through the BrachyBase.org community genome and annotation database.

# 155

## The Biofuel Feedstock Genomics Resource—A Comparative Gene and Transcript Annotation Database for Lignocellulosic Feedstock Species

Kevin L. Childs* (kchilds@plantbiology.msu.edu) and **C. Robin Buell**

Plant Biology Department, Michigan State University, East Lansing, Mich.

**Project Goals: Complete genome sequences are only available for two lignocellulose biofuel species: poplar and maize. Limited amounts of transcript sequence are available for many biofuel species, but those resources are also incomplete. Researchers would benefit from comparative analyses between their genes-of-interest in their species-of-interest with sequence resources from many other species. Such a resource is not available for biofuel species. The goals of this project are to provide a comparative resource where gene and transcript sequences from model species and biofuel species are thoroughly and consistently annotated and where those sequences are related to each other in numerous ways. The success of the project depends on the ability of a researcher to move easily between sequences from multiple species that are related to each other based on functional annotation and/or sequence similarity.**

Candidate plant species to be used for lignocellulosic ethanol production include a large number of species within the Poaceae, Pinaceae, and Salicaceae families. For these biofuel feedstock species, there are variable amounts of genome sequence resources available, ranging from complete genome sequences (e.g. *Populus trichocarpa*, *Zea mays*) to transcriptome data sets in the form of Expressed Sequence Tags (e.g. *Pinus glauca*, *Panicum virgatum*). Some species with importance to biofuel feedstock research have negligible sequence resources. While obtaining genome or transcript sequence is the initial step in a genomics-based approach to biological research, the more challenging step in genomics is the process of understanding gene function and how genes and their products confer the underlying processes/traits in plant biology. This challenge is mostly attributable to two issues: a large percentage of genes within genomes have no known function and experimental approaches to determining gene function on a per gene basis are fiscally prohibitive. One method to improve our understanding of gene function is through comparative approaches in which sequence similarity is used to cross-annotate orthologs and paralogs thereby leveraging all available functional annotation data to improve the annotation of genes in species with limited annotation data. For the current release of the Biofuel Feedstock Genomics Resource (http://bfgr.plantbiology.msu.edu/), we have created a comprehensive, uniform, well annotated resource for data-mining genomic data for biofuel feedstock species. To augment comparative analyses, the predicted genes from seven sequenced plant genomes and the predicted transcriptomes from 44 species, including biofuel feedstock target species and their phylogenetic relatives, are annotated within the database. All sequences in the resource have been aligned to Uniref proteins, InterPro domains, KEGG Orthologs, as well as the predicted gene sequences from fully sequence plant species. Orthologous gene groups have been identified in order to allow users to easily identify orthologous and paralogous relationships between their genes of interest and other sequences in the resource and in order to aid in cross-referencing to sequences from other species. Numerous search functions are provided to allow users to find sequences based on sequence alignments, functional annotation, sequence identifiers, homology to KEGG orthologs, InterPro domain names and GO terms. Additionally, sequences can be searched for SNP and SSR markers.

# Systems Biology for Environmental and Subsurface Microbiology

## 156

### Tracking Carbon Flows in a Model Microbial Community Using Genome-Enabled Methods and Stable Isotope Probing

Nicholas Justice[1]* (njustice@berkeley.edu), Chongle Pan[2]* (panc@ornl.gov), Chris Miller[1]* (csmiller@gmail.com), Curt Fischer,[1] Ben Bowen,[3] Itai Sharon,[1] Paul Wilmes,[1] Alexis Yelton,[1] Brett Baker,[1] Daniela Aliaga Goltsman,[1] Steven Singer,[3] Brian Thomas,[1] Robert Hettich,[2] Trent Northen,[3] and **Jill Banfield**[1]

[1]University of California, Berkeley; [2]Oak Ridge National Laboratory; and [3]Lawrence Berkeley National Laboratory

http://geomicrobiology.berkeley.edu/pages/amd.html

**Project Goals: The primary goal is to develop integrated omics methods, including stable isotope probing, for tracking carbon flows in microbial communities.**

Microbial communities both contribute to carbon fixation and drive the degradative portion of the carbon cycle. However, existing methods are inadequate to systematically examine carbon flows in complex natural communities at the molecular level. The primary goals of our project are to develop an integrated "omics" toolset enabled by community metagenomics and stable isotope probing and to apply this toolset to model microbial communities of tractable complexity. We aim to demonstrate that a comprehensive understanding of the carbon and energy flows in this model community can be achieved. The developed toolset will be broadly applicable to many microbial communities relevant to DOE missions.

The acid mine drainage (AMD) microbial community has been established as a model system for ecological and evolutionary studies and "omics" method development (Denef et al. 2010). Notably, a full carbon cycle is represented in this ecosystem. Organisms in the floating AMD biofilms fix $CO_2$ using chemical energy from iron oxidation. The biofilm undergoes successional changes, and eventually sinks to the sediment-water interface where it is anaerobically degraded. These communities have been extensively studied using a proteogenomic approach via 5 years of collaborative research funded by DOE. We have reconstructed near-complete genomes of essentially all the bacterial (8 lineages) and archaeal (12 lineages) natural populations consistently detected in the system, as well as one fungal genome (see below) and the genomes of numerous viruses/phage and plasmids. In addition, we have mapped changes in population structure and protein abundance over space, time, and developmental stages.

A specific new target for ongoing research is the degradative portion of the AMD carbon cycle. We evaluated *in situ* change in community structure of sunken biofilms using fluorescent *in situ* hybridization (FISH) and 454 pyrosequencing, and analyzed metabolic activities with proteomics. FISH measurments indicate that the sunken decaying community transitioned from dominance by bacteria to archaea. Intriguingly, the sunken community proteome showed more bacterial proteins than expected based on FISH. We attribute this to persistence of bacterial proteins from dead cells in the degrading biomass. Indeed, many bacterial proteins have suffered amine hydrolysis, likely due to exposure to the acidic solution after cell lysis. Twelve archaeal species identified in the sunken biofilms belong predominantly to the order *Thermoplasmatales* (A through I-plasma), but also include the deeply branched ARMAN Euryarchaeota. Genomic analyses of the archaea indicated that they are likely obligate heterotrophs, but that inorganic electron donors may also play an important role in their metabolism. Genotypic differences indicate varied roles for different archaea in community functioning (Justice *et al.,* in prep. Yelton *et al.,* in prep). To understand the roles of different community members, anaerobic communities were cultivated in the laboratory and analyzed using stable isotope probing, proteomics, and metabolomics.

Fungi often colonize late-developmental-stage biofilms at low relative abundance, and may play an important role in recycling carbon in the community. To test this hypothesis, we have targeted "streamer" biofilms dominated by fungal hyphae for genomics, proteomics, and transcriptomics. By developing methods for assembly of genomes from short-read Illumina community DNA sequencing data, we have reconstructed the near-complete genome (27 Mbp) of the dominant fungal community member, *Acidomyces richmondensis*. Predicting proteins from this genome allowed for identification of thousands of fungal proteins, covering many important metabolic pathways (Miller et al., in prep.). Comparative transcriptomic analysis is also being performed to better characterize the potential differences in fungal metabolism between streamer and floating biofilms.

We have also developed a novel method for characterizing microbial community structure from Illumina metagenomic data. Our approach reconstructs full-length ribosomal small subunit genes from short sequencing reads (a previously intractable problem), and provides accurate estimates of relative taxon abundances (Miller et al., in review). Together with FISH, we have used the method to characterize the fungal community. Fungal streamer communities do not share the same community structure as floating biofilms: the typical primary producer is only found at low abundance. Yet neither do they share identical community structure with heterotrophic sunken biofilms; while *Thermoplasmatales*, ARMAN, and *Sulfobacillus* all are abundant, fungi which are absent in sunken biofilms still dominate the community.

Further proteomic and transcriptomics data should reveal if fungal-dominated biofilms represent a distinct degradative community in the carbon cycle.

Carbon flows in the community involve exchange and metabolism of small molecules that can be characterized using metabolomics. Identification of the chemical formula and structure for a detected-spectral feature in an untargeted metabolomics experiment is facilitated by the high resolution, sensitivity, and mass accuracy achieved by liquid chromatography-electrospray ionization quantitative time of flight mass spectrometry (LC-ESI-qTOF-MS). However, often the chemical identity is difficult to discern from mass-information alone. As an assist, heuristically filtered chemical formula generators and isotopic pattern fitting is utilized. Formulas and biological origin of the metabolites was then confirmed using stable isotope labeling. However, this approach does not provide structural information. A novel proton labeling technique was developed and applied to define the location of exchangeable (H/D) hydrogens and elucidate the most abundant metabolic features detected in a model mixed-species microbial community revealed the importance of a new class of lipids (Fischer et al., in review; Bowen et al. in preparation, and in press).

Stable isotopic labeling is also used to link metabolic activities to community structure. Nitrogen or carbon flows from $^{15}N$- or $^{13}C$-enriched substrates into the biomass of a microbial community can be traced by measuring the incorporation of these stable isotopes into biomarkers such as lipids, nucleic acids, and proteins. Existing stable isotope probing (SIP) methods can only measure a few selected lipids, DNA sequences or proteins. We developed a proteomic SIP method that can determine $^{13}C$ or $^{15}N$ atom% of thousands of identified proteins from multiple strains and species in the AMD community. The proteomic SIP method was validated using the AMD communities grown in laboratory at three known $^{15}N$ atom%: 0.4% $^{15}N$ (natural abundance), ~50% $^{15}N$, or ~98% $^{15}N$. 1408~2326 proteins were identified from each sample with false discovery rates of 0.4%~2%. The median $^{15}N$ atom% from each sample has less than 1% deviation from the experimentally imposed $^{15}N$ in the validation samples. The $^{15}N$ atom% was also measured with high precision. 95%~99% of identified proteins have a $^{15}N$ atom% estimated within the 5% range of the expected atom%. This indicates that our method can automatically identify the sequence of, and quantify the degree of heavy atom enrichment in, thousands of proteins from microbial community proteome samples. The method was used to monitor incorporation of $^{15}N$ into established and regrowing microbial biofilms. 1814~2407 non-redundant proteins were identified from four samples in a time course. Most proteins were either unlabeled (<5% $^{15}N$ atom %) or highly labeled (>80% $^{15}N$ atom %). The abundance ratio between two isotopologues of a protein was estimated using the ProRata algorithm. The different labeling patterns among microbial members of the community revealed distinct colonization behavior and metabolic activities among the microorganisms (Pan et al. in press).

Overall, in the first year of the project we have made considerable progress toward deploying metagenomic analyses coupled to isotope-enabled proteomic and metabolomic studies to track carbon and nitrogen flows in microbial communities.

# 157

## Wiring Microbes to the Sun: Mechanisms for Energy Conservation During Microbial Electrosynthesis

Kelly P. Nevin* (knevin@microbio.umass.edu), Trevor L. Woodard, Sarah A. Hensley, Ashley E. Franks, Zarath M. Summers, Jianhong Ou, Oona L. Snoeyenbos-West, Muktak Aklujkar, Pier-Luc Tremblay, Ching Leang, and **Derek R. Lovley**

Department of Microbiology, University of Massachusetts, Amherst

http://www.geobacter.org

**Project Goals: The primary goals of this project are to develop genome-based experimental approaches to elucidate the function of anaerobic microbial communities involved in groundwater bioremediation and novel bioenergy processes and to develop genome-scale models of these communities in order to better understand and optimize bioremediation and bioenergy applications.**

Microbial electrosynthesis is a novel bioenergy approach in which microorganisms consume electrons derived from electrodes to convert carbon dioxide to multi-carbon fuels or other desirable chemicals. When powered with solar technology microbial electrosynthesis has the same net reaction as plant-based photosynthesis, using solar energy to convert carbon dioxide and water to organic compounds and oxygen. However, microbial electrosynthesis has a number of potential benefits over biomass-based strategies for fuel production. These include: 1) solar technology is 100-fold more effective than photosynthesis in harvesting solar energy; 2) microbial electrosynthesis produces fuels directly whereas only a fraction of the energy harvested in biomass is converted to fuel; 3) microbial electrosynthesis eliminates the need for arable land for converting solar energy to fuel; 4) microbial electrosynthesis conserves water and avoids the pollution of water resources associated with biomass production. Thus, microbial electrosynthesis represents a new approach to photosynthesis that can convert solar energy to fuels more efficiently and with less environmental degradation than biomass-based strategies.

Previous studies in our laboratory demonstrated the concept of microbial electrosynthesis with the acetogenic microorganism *Sporomusa ovata*. *S. ovata* colonized graphite cathodes and used electrons derived from the cathodes to reduce carbon dioxide to acetate. Electron consumption and appearance of electrons in acetate were consistent with

the reduction of carbon dioxide to acetate with the cathode serving as the sole electron donor.

In order to better understand the mechanisms for microbial electrosynthesis, the possibility of catalyzing this process with other acetogenic microorganisms was evaluated, including several strains for which the complete genome sequence is available and/or strains that are genetically tractable. Additional *Sporomusa* strains were capable of acetate electrosynthesis, as were *Clostridium ljungdahlii*, *Clostridium aceticum*, and *Moorella thermoacetica*. The acetogen *Acetobacterium woodii* was unable to consume current.

Comparison of predicted mechanisms of energy conservation in the various acetogens suggested a mechanism for microorganisms to conserve energy for cell maintenance during electrosynthesis. Those acetogens capable of electrosynthesis all have proton-dependent ATPases. When carbon dioxide is reduced to acetate with electrons derived from extracellular electron transfer there is a net consumption of protons within the cytoplasm generating a proton-gradient across the inner membrane. Proton flux through proton-dependent ATPases to alleviate the proton imbalance can generate ATP. This mechanism of energy conservation would not be possible in *A. woodii*, which has a sodium-dependent ATPase. Thus, the model is consistent with experimental results.

Surprisingly, *Geobacter metallireducens* consumed current as well as the most active acetogenic microorganisms with carbon dioxide serving as the sole electron donor. *G. sulfurreducens*, which readily reduces a variety of other electron acceptors with an electrode as the sole electron donor, could not reduce carbon dioxide. Analysis of the *G. metallireducens* genome suggested two potential pathways by which *G. metallireducens* might reduce carbon dioxide to acetate. One of these is a reductive TCA cycle and the other is the recently discovered dicarboxylate/4-hydroxybutyrate cycle. The key enzyme of this pathway, 4-hydroxybutyryl-CoA dehydratase, is not found in *Geobacteraceae* other than *G. metallireducens*. The existence of two potential pathways for carbon dioxide reduction in *G. metallireducens* is promising for the development of strains engineered to produce a variety of organic compounds via electrosynthesis. The potential role of the two pathways in electrosynthesis is currently being evaluated with genetic approaches.

The genome of *G. metallireducens* contains genes for a wide diversity of redox-active proteins that could function as electrical contacts between cells and electrodes and for electron transfer into the cytoplasm. Effective strategies for gene deletion and expression of heterologous genes have recently been developed for *G. metallireducens*, making investigations of electron transfer mechanisms feasible. Examination of the genomes of *C. ljungdahlii* and *M. thermoacetica*, the two genetically tractable acetogens capable of electrosynthesis, has demonstrated that acetogens are likely to have much different strategies for electron transfer for electrosynthesis than *G. metallireducens*. We have recently significantly improved techniques for the genetic manipulation of *C. ljungdahlii* to facilitate investigation of electron transfer mechanisms in this gram-positive organism. Elucidation of these pathways is key to optimization of this potentially transformative bioenergy technology.

# 158

## Genomic Patterns of Amino Acid Usage in Fe(III)-Reducing, Sulfur-Reducing and Syntrophic Geobacteraceae

Muktak A. Aklujkar* (muktak@microbio.umass.edu) and **Derek R. Lovley**

Department of Microbiology, University of Massachusetts, Amherst

http://www.geobacter.org

**Project Goals: The primary goals of this project are to develop genome-based experimental approaches to elucidate the function of anaerobic microbial communities involved in groundwater bioremediation and novel bioenergy processes and to develop genome-scale models of these communities in order better to understand and optimize bioremediation and bioenergy applications.**

The family Geobacteraceae within the Deltaproteobacteria includes Fe(III)-respiring Geobacter species, sulfur-reducing Desulfuromonas species, and syntrophic Pelobacter species. Several genomes of Geobacteraceae have now been sequenced, leading to the discovery of intriguing differences not only of physiology and metabolism but even at the level of amino acid usage.

We have already demonstrated that the inhibitory effect of a short nucleotide sequence on histidyl-tRNA synthetase in the *Pelobacter carbinolicus* lineage could account for the loss of ancestral genes of respiratory metabolism, in which multiple closely spaced histidine codons may cause translation to stall. For that study, a histidine demand index, defined as the number of histidine codons divided by the harmonic mean distance between them, was computed for every gene in the genomes of *P. carbinolicus*, *Desulfuromonas acetoxidans*, *Geobacter bemidjiensis*, *Geobacter metallireducens*, and *Geobacter sulfurreducens*. The same analysis has now been performed on the genomes of *Pelobacter propionicus*, *Geobacter daltonii*, *Geobacter lovleyi*, *Geobacter* sp. M18, *Geobacter* sp. M21, and *Geobacter uraniireducens*, revealing that high-histidine-demand coding sequences are fewer in *P. propionicus* as well as *P. carbinolicus*, despite similar frequencies of histidine usage overall.

Similar analyses were performed for each of the other nineteen amino acids, in an attempt to identify patterns that distinguish the *Geobacter* clade that includes *P. propionicus* from the *Desulfuromonas* clade that includes *P. carbinolicus*, or that distinguish the two syntrophic *Pelobacter* species from their relatives that respire Fe(III) or sulfur. One of the most striking results is that genomic glutamine demand is lower in *G. sulfurreducens* and *G. metallireducens* compared

to subsurface *Geobacter* species, higher in *G. lovleyi*, and extraordinarily high in *D. acetoxidans*, suggesting that the common ancestor of *Geobacteraceae* had a high-glutamine-demand proteome that has shifted to ever lower glutamine demand in its Fe(III)-respiring descendants. Both *Pelobacter* species have much lower genomic glutamine demand than their closest relatives. Genomic glutamine demand can be correlated with the frequency of glutamine usage overall. Further investigations are ongoing to determine what high-glutamine-demand proteins have been acquired or lost, which orthologous proteins have variable glutamine content, and whether any protein sequence motifs are linked to mutations of glutamine. Metabolic pathways that involve glutamine are also being compared across the *Geobacteraceae*.

Lower genomic cysteine demand distinguishes *D. acetoxidans*, *P. carbinolicus*, *P. propionicus* and *G. lovleyi* from the majority of *Geobacter* species. Although several of the high-cysteine-demand genes of *Geobacter* species encode multiheme *c*-type cytochromes, this does not fully account for the difference, as the *D. acetoxidans* genome encodes 80 *c*-type cytochromes. Scrutiny of the remaining high-cysteine-demand genes may lead to the identification of redox-active proteins unique to *Geobacter* species and potentially important for respiration of Fe(III).

The median lengths of protein-coding sequences in the genomes of *Geobacter* species are almost all in the range from 283 to 287 codons, with *G. metallireducens* as the sole outlier at 294 codons. Interestingly, *G. bemidjiensis* is at the low end despite having more copies of rRNA and tRNA genes than *G. metallireducens*. *D. acetoxidans* has a median protein-coding sequence length of 289 codons. In contrast, the median lengths for *P. propionicus* (276 codons) and *P. carbinolicus* (275 codons) suggest that shorter proteins may be characteristic of the physiology of syntrophs, perhaps because respiration of extracellular electron acceptors requires more proteins to be made with signal peptides.

In conclusion, this study shows that the genomes of *Geobacteraceae* have unique amino acid clustering patterns that may reflect differences in their environment, physiology and metabolism as well as their phylogeny.

# 159

## Environmental Functional Genomics: Elucidating the Function of Highly Expressed Proteins During Subsurface Uranium Bioremediation

Jiae Yun, Toshiyuki Ueki* (tueki@microbio.umass.edu), and **Derek R. Lovley**

Department of Microbiology, University of Massachusetts, Amherst

http://www.geobacter.org

**Project Goals: The primary goal of this project is to develop genome-based experimental approaches to elucidate the function of anaerobic microbial communities involved in groundwater bioremediation and novel bioenergy processes and to develop genome-scale models of these communities in order to better understand and optimize bioremediation and bioenergy applications.**

One of the common findings in metagenomic, metatranscriptomic, and environmental proteomic studies is that a function can not be assigned to the vast majority of the diverse sequences recovered. This severely limits interpretation of the environmental data. Therefore, tools for experimentally elucidating the function of genes recovered in environmental studies need to be developed.

For example, extensive sequencing of subsurface mRNA collected during *in situ* bioremediation of uranium-contaminated groundwater at the DOE study site in Rifle, CO, as well as proteomic analysis of similar samples, has identified many genes that are actively expressed during the bioremediation process. Most of these genes can be assigned to *Geobacter* species. However, lack of information on the function of these genes has stymied progress in diagnosing the physiological status of the subsurface *Geobacter* species and their activities.

Analysis of proteins in the groundwater during active bioremediation revealed that one of the most abundant proteins stained positively for heme, suggesting that it was a *c*-type cytochrome. A peptide sequence in the protein, identified with mass spectrometry, was identical to that in a putative *c*-type cytochrome in the genome sequence of *Geobacter* strain M18, an isolate from the Rifle site. A homologous sequence was found in the genomes of multiple *Geobacter* species in Subsurface Clade I, a cluster of closely related *Geobacter* species that typically predominate in subsurface environments. Analysis of the subsurface proteome with an antibody designed to recognize the conserved portion of this protein in Subsurface Clade I *Geobacter* species confirmed its presence and also demonstrated that the abundance of the cytochrome was positively correlated with the success of uranium removal from the groundwater.

The lack of a developed genetic system for any of the Subsurface Clade I *Geobacter* species prevented evaluation of the cytochrome function in the available pure cultures and a homologous gene was not found in the genetically tractable *Geobacter sulfurreducens*. However, the function of the cytochrome could be evaluated by testing the ability of the gene to complement strains of *G. sulfurreducens* in which other cytochrome genes were deleted. OmcS is an abundant cytochrome in *G. sulfurreducens* growing on Fe(III) oxide. It is aligned along the conductive pili of *G. sulfurreducens* and is thought to facilitate electron transfer from the pili to Fe(III) oxides. Expression of the subsurface cytochrome gene in a *G. sulfurreducens* strain in which the gene for OmcS was missing restored the capacity for Fe(III) oxide reduction, suggesting that the function of the subsurface cytochrome gene is similar to that previously described for OmcS.

These studies demonstrate that, with the appropriate experimental design, it is possible to elucidate the function of abundant environmental proteins, even when the protein is part of a complex process, such as extracellular electron transfer. Additional functional genomic studies of other genes of unknown function that are highly expressed during bioremediation of uranium-contaminated groundwater are underway.

# 160

## Functional Characterization of Aromatic-Solute Binding Proteins in Environmental Organisms

Ruth Pietri,[1]* Sarah Zerbs,[2] Danielle Corgliano,[2] **Frank Collart,**[2] and Lisa M. Miller[1]

[1]National Synchrotron Light Source, Brookhaven National Laboratory, Upton, N.Y.; and [2]Biosciences Division, Argonne National Laboratory, Lemont, Ill.

**Project Goals: see below**

This project addresses the *Genomic Science Program* mission by identifying the functional characteristics of proteins from key microorganisms and plants that are known to respond to complex environmental conditions. In particular, the program focuses on proteins that facilitate communication between the cell and the environment such as ABC transporters from microbial organisms living in soil. The goal is to characterize these proteins using genomic, biophysical and computational approaches and to develop model systems to address fundamental environmental concerns.

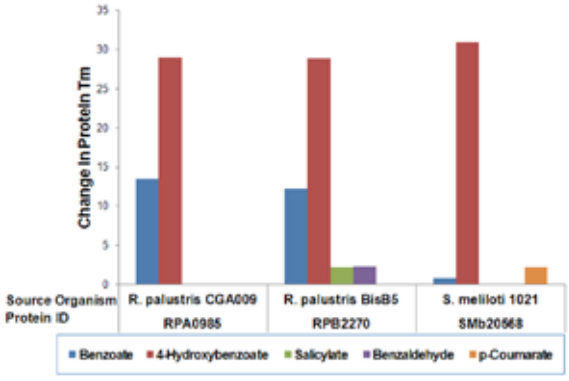



Figure 1. Ligand Specificity in Solute-binding Homologues

In this study we evaluated the ABC transporters from several soil organisms with an emphasis on proteins involved in the utilization of lignin-degradation products. The analysis set includes protein encoded in the genomes of *Rhodopseudomonas palustris, Bradyrhizobium japonicum*, and *Sinorhizobium meliloti*. These microbes share common features in transporter, sensor and metabolic characteristics but the individual strains and organisms were isolated from diverse ecological niches. They metabolize organic compounds derived from lignin-degradation products and have the potential for bioremediation of xenobiotic compounds. We hypothesize the ABC transporters profile for these aromatic compounds are predictive for the metabolic pathways and regulatory networks associated with utilization lignin degradation products. To test this hypothesis, we screened and characterized a set of transporter proteins from soil organisms using fluorescence thermal shift assay (FTS), isothermal calorimetry (ITC), circular dichroism spectroscopy (CD), infrared spectroscopy (FTIR), and small/wide angle x-ray scattering techniques (SAXS/WAXS).The FTS assay is used as a preliminary screen to profile the aromatic compound binding specificity of proteins in soil organisms known to utilize compounds derived from lignin degradation (Figure 1). The kinetic and thermodynamic properties of these proteins, obtained by ITC, showed that the proteins exhibit high affinity for the aromatic substrates with dissociation constants ($K_d$) in the submicromolar range, similar to other transporter binding proteins. Moreover, analysis of the flanking genomic regions reveals the co-localization of these transporter genes with metabolic genes associated with utilization of the transported compounds. The profile and number of transport proteins specific for aromatic compounds is consistent with ecological and laboratory studies, which demonstrate the capabilities of these organisms for the utilization of plant degradation products such as lignin-derived aromatic compounds.

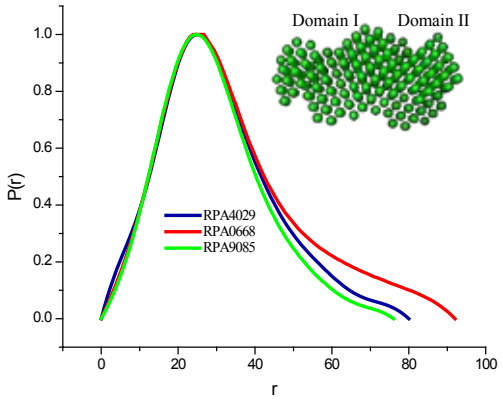



Figure 2. Electron distribution functions of transporter proteins from *R. palustris*. The structure predicted by SAXS/WAXS in shown in the upper right corner.

The specificity of three aromatic-solute binding transporters was further assessed by structural analysis using CD, FTIR, and SAXS/WAXS. CD and FTIR provide complementary information about protein secondary structure, and were therefore employed to determine the secondary motifs of the proteins in the presence and absence of aromatic ligands. SAXS/WAXS on the other hand, are capable of providing information about the size and shape of proteins and were also used to define the three dimensional fold of the transporters in the presence and absence of ligands. CD and FTIR analyses showed that the proteins share similar secondary structures with a mixture of α-helical and β-sheet motifs. The results also demonstrated that the proteins retain their native secondary structures in the presence of aromatic ligands. Similarly, SAXS/WAXS analysis

showed that the proteins have analogous topologies, which remain unaffected in the presence of aromatic ligands. The overall shape of the proteins, derived from their electron distribution functions (Figure 2), suggests that they have ellipsoid-like structure with two distinct domains. Indeed, theoretical structural modeling of the proteins predicted a three dimensional fold consisting of two domains connected by a hinge region, which is conserved in the solute-binding proteins of ABC transporters.

Overall, the FTS assay and ITC results showed that three ABC transporter proteins from *R. palustris* bind aromatic ligands with high affinity. CD, FTIR and SAXS/WAXS data suggested that these proteins have the conserved structure with two mixed α-helix and β-sheet domains connected by a hinge. As with other transporters, binding of the aromatic ligands might induce small movement of amino acid residues in the hinge region. Taken together, the study has shown the capability of indentifying solute-binding proteins that interact with lignin-degradation products using genomic and biophysical approaches, which can be extended to other organisms. These genome-scale studies will enable the identification of other proteins with similar ligand-binding profiles and characteristics and enable refinement of sequence-based methods for extension to other organisms and systems.

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).

# 161

## Molecular Approaches for Elucidation of Sensory and Response Pathways in Cells

Sarah Zerbs,[1]* Michael Endres,[1] Elizabeth Landorf,[1] William Studier,[2] and **Frank R. Collart**[1]

[1]Biosciences Division, Argonne National Laboratory, Lemont, Ill.; and [2]Biology Division, Brookhaven National Lab, Upton, N.Y.

**Project Goals: see below**

Increased knowledge of protein function enhances our understanding of cellular functions and is ultimately required to model biological activities and systems. This 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 long term objective of the program is to define cellular sensory and regulatory pathways that respond to environmental nutrients thereby facilitating a system-level model that predicts the cellular response to environmental conditions or changes. One aspect of this program will develop tools to bridge the gap between genomes and systems biology. In this context we are developing methods for experimental validation of cellular sensory and regulatory networks by experimental validation of the function of two component sensors and transcription factors.

To improve prospects for identification of effectors for bacterial transcriptional regulators, we are evaluating *in vitro* ligand mapping strategies to identify environmental molecules or their metabolic derivatives that bind to transcription regulatory proteins and identify the DNA binding regions for these proteins. One approach maps effectors to TRs by assessment of increased stability attributable to the TR-ligand interaction. Effector binding confers increased thermal stability which can be used as a surrogate to infer binding specificity. This feasibility of this approach was validated with a set of *Escherichia coli* TRs with known effector molecule binding specificity and the method then extended to identify effector ligands for a set of TR from *Shewanella oneidensis*.

In a complimentary approach, we are applying biolayer interferometry to characterize the interaction of TR with DNA binging regions and effector molecules. The *forté*BIO Octet Red instrument is designed to enable a label free approach for the detection of interactions between biological molecules. The instrument uses an array of fiber-optic sensors that detect biomolecular interactions via biolayer interferometry. This instrument can be used for the identification of protein-protein and protein-ligand interactions using relatively small amounts of material in a microwell plate format. Identification of regulatory ligands coupled with knowledge of the DNA-binding regions of the transcription factors allows the association of metabolic pathways with the regulatory network. The functional assignments and ability to define specific sensory and regulatory pathways will increase the predictive capability of current models and support the development of predictive systems-level models. This increased knowledge of the molecular components and control features of cellular sensory and response pathways is essential for our understanding of natural biological processes related to carbon management, sustainability and bioenergy.

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, Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program.

# 162
## Molecular Processes of Mycorrhizal Symbiosis

Peter E. Larsen,[1]* Geetika Trivedi,[2] Avinash Sreedasyam,[2] Leland Cseke,[2] and **Frank R. Collart**[1]

[1]Biosciences Division, Argonne National Laboratory, Lemont, Ill.; and [2]Department of Biological Sciences, University of Alabama, Huntsville

## Project Goals: see below

This fungal-plant symbiosis is widespread and a process of major ecological importance. Establishment of this symbiosis involves a progressive series of complex developmental steps accompanied by radical changes in metabolism and plant/fungal interactions with the environment. The scientific goal of this theme is the delineation of the molecular events associated with alteration of the nutrient assimilation in the plant and fungal organisms.

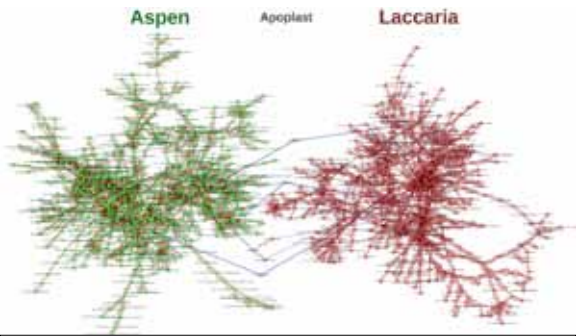



**Figure 1.** This model of mycorrhizal metabolome is predicted from transcriptomic NGS analysis. Edges are specific enzyme activities detected as statistically significantly expressed in the mycorrhizal transcriptome. Nodes are metabolites inferred by detected enzyme activities. Green edges are metabolic activities expressed in aspen root, red edges are metabolic activities expressed in Laccaria, blue edges are the action of metabolite transport predicted from detected, expressed mycorrhizal transporter genes, and blue dashed edges are predicted metabolite diffusion across cellular membranes. Light green nodes are metabolites predicted to be present in aspen roots, pink nodes are metabolites predicted to be present in Laccaria, and grey nodes are metabolites predicted to be present in mycorrhizal apoplast.

The molecular elements in these signaling and metabolic networks are identified by merging Next Generation Sequencing (NGS) transcriptomic analysis of plant roots and associated soil microorganism communities with established databases of metabolic networks, previously archived transcriptomic data, and experimentally-verified protein interactions. The transcriptomic data was used to identify statistically significantly expressed gene models using a bootstrap-style approach, and these expressed genes were mapped to specific metabolic pathways. This mycorrhizal metabolome (Figure 1) is comprised of the expressed metabolic enzymes in the mycorrhizal transcriptome and the transporters required for the exchange of metabolic compounds between these two organisms. Analysis of the models suggests they are consistent with experimental environmental data and provide insight into the molecular exchange processes for organisms in this complex ecosystem. For example, in mycorrhizal symbiosis, aspen exchanges with *L. bicolor* photosynthetic sugars for nutrients. This expectation is validated by the statistically enriched expression of sugar porters by *L. bicolor*. The enrichment of KEGG amino acid metabolism pathways with unique expressed enzyme activities and the enrichment for expressed amino acid transporters for both aspen and *L. bicolor* indicate that, for mycorrhizae formed in WPM, *L. bicolor's* debt to aspen for carbon is paid with organic nitrogen. *L. bicolor* expresses the metabolic capacity to synthesize nitrogenous compounds such as glycine, glutamate and allantoin, via pathways not expressed in aspen roots. In the growth conditions used here, the predicted exchange compounds are the fructose and glucose as well as organic nitrogen compounds, specifically glycine, glutamate and perhaps allantoin. The predictions suggest *L. bicolor* is an active metabolic partner in addition to passively extending the absorptive surface of aspen roots. This role encompasses uptake of ammonium from the medium and synthesis of more complex compounds provided to the plant. Additional experiments in different nutrient environments are expected to uncover additional mechanisms of mycorrhizal metabolic interactions.

These predictions conform closely to prior experimental observations in other plant species and alternate forms of symbiotic relationships between plant roots and soil microorganisms. By merging transcriptomic data with genomic annotation and previously published metabolic pathway information, knowledge of what fraction of an organism's metabolic capacity is being expressed during symbiotic interaction provides insights into the mycorrhizal metabolome that a single source of data could not. The model-based predictions lead directly to specific, testable biological hypotheses and target particular expressed proteins and transporters for molecular characterization. Additional experiments in different nutrient environments will uncover additional mechanisms of mycorrhizal metabolic interactions. Our available and interactive model of the mycorrhizal metabolome will serve as an important resource for other investigators.

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. 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).

# Biological Systems Research on the Role of Microbial Communities in Carbon Cycling

## 163

### With or Without You: Genome-Scale Analysis of Contrasting Strategies for Interspecies Electron Transfer

Zarath M. Summers* (zsummers@microbio.umass.edu), Nikhil S. Malvankar, Masahiko Morita, Mollie E. Murnane, Dan Carlo Flores, Ashley E. Franks, Ludovic Giloteaux, Amelia-Elena Rotaru, and **Derek R. Lovley**

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 in the current research is to determine if the syntrophic associations that are central to the functioning of methanogenic terrestrial wetlands can be predicatively modeled with genome-scale metabolic models.**

Interspecies exchange of electrons is necessary for the proper functioning of the methanogenic ecosystems that are major contributors of atmospheric methane and for successful operation of methanogenic digesters, the most effective strategy yet devised for converting organic wastes to fuel. For over 40 years interspecies hydrogen transfer has served as the paradigm for anaerobic interspecies electron transfer.

In interspecies hydrogen transfer, hydrogen gas serves as the electron carrier, with electron-donating microorganisms reducing protons to hydrogen and hydrogen-oxidizing microorganisms extracting the electrons from hydrogen with the release of protons. Energetic inefficiencies inherent in the production and subsequent oxidation of hydrogen and limitations on the diffusion of hydrogen between partners may constrain the effectiveness of interspecies electron exchange via hydrogen and contribute to instabilities in methanogenic systems.

Recent studies with a model system of two *Geobacter* species revealed an alternative to interspecies hydrogen transfer for interspecies electron exchange (Z.M. Summers *et al.* 2010. Direct exchange of electrons within aggregates of an evolved syntrophic coculture of anaerobic bacteria. *Science* 330(6009): 1413-1415). *Geobacter metallireducens* and *Geobacter sulfurreducens* were cultured on ethanol under conditions in which *G. metallireducens* was required to provide electrons to *G. sulfurreducens* in order for ethanol metabolism to proceed. Strong selective pressure for rapid ethanol metabolism resulted in a co-culture that could metabolize ethanol much faster than the initial co-culture. Adaption for enhanced metabolism was associated with the formation of large (1-2 mm diameter) aggregates of the two species and selection for a mutation that enhanced the production of OmcS, a *c*-type cytochrome aligned along the electrically conductive pili of *G. sulfurreducens*. Introducing the mutation in *G. sulfurreducens* prior to initiating the co-culture accelerated aggregate formation and adaption for rapid ethanol metabolism. Additional genetic analysis and physiological studies ruled out the possibility of interspecies transfer of hydrogen or formate significantly contributing to the electron exchange and demonstrated that both pili and OmcS were essential for effective co-culture function. The aggregates were electrically conductive with sufficient conductance to account for the electron exchange within the aggregates. These studies suggest that the two *Geobacter* species directly exchanged electrons via long-range (μm) conduction through conductive pili with short-range (Å) electron exchange mediated by the OmcS of *G. sulfurreducens*, and potentially one or more cytochromes of *G. metallireducens*.

Aggregation of methanogens and the other organisms involved in the anaerobic conversion of wastes is essential for efficient conversion of wastes to methane in anaerobic digesters. It has previously been suggested that the close proximity of cells promotes interspecies hydrogen transfer. However, measurements of electron conduction through aggregates collected from commercial-scale methanogenic digesters yielded conductivities comparable to those of the *Geobacter* co-culture aggregates. Temperature studies indicated that conduction along organic filaments was the most likely mode of electron flux. Additional physiological studies ruled out hydrogen as a significant electron carrier. Studies on gene expression within the aggregates are underway to better understand the components responsible for the apparent direct electron transfer in these systems.

The direct electron transfer described above contrasts with the many instances of interspecies hydrogen transfer previously reported for a diversity of co-cultures in laboratory studies. It is clear that interspecies hydrogen transfer can take place under some conditions. For example, *Pelobacter carbinolicus,* a close relative of *Geobacter* species, grew syntrophically with the methanogen *Methanospirillum hungatei*, converting ethanol to methane. However, the two cell types did not aggregate as would be expected for direct cell-to-cell electron transfer. Furthermore, analysis of gene expression in both co-culture partners with whole-genome microarrays demonstrated that *P. carbinolicus* expressed genes for hydro-

gen production and that *M. hungatei* continued to express the genes for hydrogen uptake that are expressed during pure culture growth on hydrogen. *P. carbinolicus* did not increase expression of genes for components, such as pili and *c*-type cytochromes, that are hypothesized to be important in direct electron transfer. Co-cultures of *P. carbinolicus* with *G. sulfurreducens*, which is known to be capable of direct electron transfer under similar conditions with *G. metallireducens*, effectively metabolized ethanol, but did not form aggregates, further demonstrating that *P. carbinolicus* exchanges electrons via interspecies hydrogen transfer rather than direct cell-to-cell electron transfer.

It seems likely that the electron-donating microorganisms and methanogens that have routinely been used to study syntrophic metabolism in laboratory incubations were isolated with procedures that selected for organisms that exchange electrons via interspecies hydrogen transfer. It is expected that the study of pure cultures recovered with novel strategies that select for microorganisms that can directly exchange electrons, as well as metatranscriptomic and proteomic studies of natural methane-producing aggregates, will provide additional information on the mechanisms of direct electron exchange within methanogenic aggregates.

# 164

## Systems Biology Approach to Elucidate Electron Transfer in Methanogenic Consortia

Harish Nagarajan*[1] (nh@ucsd.edu), Mallory Embree,[1] Zhongyin Qian,[1] Eric Chu,[1] Ka Wai Lin,[1] Martin Santillan,[1] Adam M. Feist,[1] **Derek R. Lovley,**[2] Bernhard Ø. Palsson,[1] and **Karsten Zengler**[1]

[1]Systems Biology Research Group, Department of Bioengineering, University of California San Diego, La Jolla, Calif.; and [2]Department of Microbiology, University of Massachusetts, Amherst

http://systemsbiology.ucsd.edu
http://geobacter.org

**Project Goals: The long-term objective of this research is to adopt systems biology approaches to predict the activity of microorganisms involved in important aspects of the global carbon cycle under multiple environmental conditions. The short-term goal of this project adresses the development of genome-scale metabolic models for methanogens of interest, couple them with the metabolic model of a syntrophic partner, and elucidate the electron transfer mechanisms of this syntrophy. These syntrophic associations are important in methanogenic terrestrial wetlands, one of the main sources of greenhouse gas.**

Reconstruction of genome-scale metabolic networks has become a common denominator in systems biology. Following an established protocol, we have reconstructed the metabolic networks of two methanogens, *Methanococcus maripaludis* and *Methanospirillum hungatei,* representative species of Methanococcales and Methanomicrobiales, respectively. The reconstructions currently contain 441 and 420 reactions with 422 and 369 genes, respectively. The reconstructions capture all the major central metabolic, biosynthetic, methanogenesis pathways, as well as pathways for carbon dioxide fixation. These genome-scale models will be interrogated using constraint-based modeling approaches and validated based on physiological data. Taken together with the published *Methanosarcina barkeri* reconstruction, these models represent the three different classes of methanogens, thus paving the way for a systems level comparative analysis of methanogenesis.

In this project, we will also couple the generated methanogenic metabolic models with a syntrophic partner to interrogate mechanisms of direct electron transfer or hydrogen transfer in microbial communities. To establish the mathematical framework for modeling direct cell-to-cell electron transfer, we will first use the laboratory evolved syntrophic co-culture of *Geobacter metallireducens* and *Geobacter sulfurreducens* as a model system. For this purpose, we have expanded the existing models for the two *Geobacter* species to account for extracellular electron transfer pathways. We have accounted for all the possible routes of electron transfer in and out of the cell through the various electron carriers such as cytochromes, ferredoxin, quinones, NAD, and FAD. In addition to the stoichiometry associated with the respective redox reactions, these pathways also account for the appropriate gene association and cellular localization of the different electron carriers. This detailed representation of extracellular electron transfer pathways enables us to characterize the role of direct electron transfer in the functioning of the evolved syntrophic consortium. This mechanistic understanding is crucial for the functioning of the microbial community in carbon cycling.

# 165

## Can Microbial Functional Traits Predict the Response and Resilience of Decomposition to Global Change?

**Principal Investigators: Steven D. Allison**[1,2]* (allisons@uci.edu), **Eoin L. Brodie**,[3] **Michael Goulden**,[2] **Adam Martiny**,[1,2] **Jennifer B.H. Martiny**,[1] and **Kathleen K. Treseder,**[1]
**Coauthors:** Ying Lu,[1] Mari J. Nyyssonen,[3] and Claudia Weihe[1]

[1]Department of Ecology and Evolutionary Biology, University of California, Irvine; [2]Department of Earth System Science, University of California, Irvine; and [3]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.

http://allison.bio.uci.edu/research.html#traits

**Project Goals: The main goal of this project is to connect diverse microbial groups with the extracellular enzyme systems that catalyze the decay of organic material. We will also determine whether different groups of microbes**

**and their enzymes respond to environmental changes, and whether they can recover from such changes. Finally, we will develop mathematical models to predict the responses of microbial communities and their associated functions under new environmental conditions.**

Recent technological advances have revealed tremendous genetic and metabolic diversity in microbial communities of bacteria, fungi, and archaea. Microbes play fundamental roles in regulating carbon losses from terrestrial ecosystems by catalyzing the breakdown of dead plant and soil organic material. This process relies on the production of enzymes that act outside of microbial cells to convert complex molecules into available forms. Although these enzymes regulate carbon cycling and sequestration in nearly all terrestrial ecosystems, we do not currently know which microbes produce the diversity of enzymes observed in natural ecosystems. Without this information, we cannot accurately predict how much carbon will be lost from ecosystems under future environmental conditions that may occur with climate change or nutrient pollution.

Our research will take place in a grassland ecosystem in Southern California that hosts an ongoing environmental change experiment funded by DOE. We will assess microbial and enzyme responses to drought and nitrogen addition, two environmental changes likely to affect an increasing number of terrestrial ecosystems locally and globally (Figure 1). High-throughput DNA sequencing will reveal shifts in the composition of the microbial community, and novel gene cloning and expression techniques will link enzymes with specific groups of microbes. We will use this information to construct mathematical models of microbial and enzyme responses to environmental change. Our models will be tested by directly manipulating microbial communities and environmental conditions. The decay rates of specific chemical compounds will be assessed with a new nanotechnological tracer approach. Tracking microbial communities, enzymes, and rates of decay over time will verify if our models are correct and tell us how quickly microbes can recover from environmental perturbations.

The most important scientific impact of this work will be to establish a firm connection between the composition of microbial communities and the enzymatic functions that affect carbon cycling. In addition, our research will generate knowledge and models useful for predicting how ecosystems will store and release plant-derived carbon under future environmental conditions. The enzyme genes and microbes we identify may also have potential industrial applications, such as the processing and synthesis of biofuels.

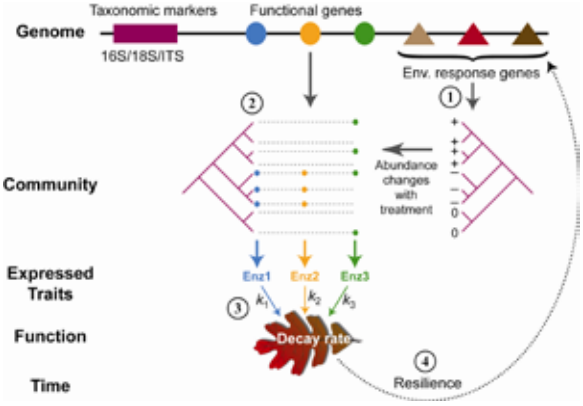



Figure 1. A trait-based framework for predicting the function of microbial communities under environmental change. Genes controlling enzymatic function and environmental responses are linked to taxonomic markers in the genomes of individual cells. In some cases, a particular trait is linked to a taxonomic group (Enz1 and Enz2), whereas others may be widely distributed among taxa (Enz3). Thus, changes in taxon abundance due to environmental drivers can affect the abundance distribution of functional genes in the community. Extracellular enzyme activities represent the expressed traits of the microbial community. Different enzymes control the decay rates (k1, k2, k3) of organic carbon compounds in litter. Circled numbers correspond to proposed research questions:

1. How do microbial taxa respond to environmental changes?
2. How are extracellular enzyme genes distributed among microbial taxa?
3. Can we predict enzyme function and litter decomposition rates by combining enzyme gene distributions with microbial taxa responses to environmental change?
4. Are microbial communities and their functions resilient to environmental change?

# 166

## Phylogenetic Distribution of Functional Traits in Microbes

**Adam C. Martiny**[1,2*] (amartiny@uci.edu), **Kathleen Treseder**,[1] Gordon Pusch,[3] and Ross Overbeek[3]

[1]Department of Ecology and Evolutionary Biology, University of California, Irvine; [2]Department of Earth System Science, University of California, Irvine; and [3]Argonne National Laboratory

**Project Goals: The goal of this project is to determine how microbes respond to global environmental changes. As part of this project, we investigating the phylogenetic distribution of functional traits in microbes and whether this is related to the complexity of the trait**

The goal of this project is to identify at what phylogenetic resolution do we find a specific functional trait. The hypothesis is that some traits are associated with large phylogenetic

clusters (e.g. oxygenic photosynthesis) whereas other traits are associated with small phylogenetic clusters (e.g. growth on small simple organic carbon molecules). To address this hypothesis, we mapped specific functional traits onto a 16S rRNA tree and determined the average phylogenetic cluster size where 90% or more of the organisms have the trait. The traits were defined either by genome annotation as subsystems for a function or the ability to grow on a particular carbon source. We detected only a few traits that are associated with clusters of 3% or more 16S rRNA sequence dissimilarity, whereas most traits are associated with small microdiverse clusters. However, most traits were non-randomly distributed. We next tested if the phylogenetic cluster size was related to the complexity of the trait. Complexity was defined either by the number of genes underlying the function or the structural complexity of a carbon substrate. We found that the average phylogenetic cluster size of a trait was significantly positively related to the complexity of the trait. This study demonstrates that microbial functional traits are related to phylogenetic groups of different sizes. Furthermore, the complexity of a trait influences the phylogenetic distribution of a trait, potentially as a result of different rates of gene gain and loss. This has general implications for our understanding of microbial evolution and biogeography.

# 167

## Systems Level Insights Into Alternate Methane Cycling Modes in a Freshwater Lake Via Community Transcriptomics, Metabolomics and NanoSIMS Analysis

M.G. Kalyuzhnaya[1]* (mkalyuzh@uw.edu), S. Yang,[1] D.A.C. Beck,[1] J. Bickford,[1] M.C. Konopka,[1] L. Chistoserdova,[1] V.J. Orphan,[2] and M.E. Lidstrom[1]

[1]University of Washington, Seattle; and [2]California Institute of Technology, Pasadena

**Project Goals: The major goal of this project is to initiate a systems-level approach to studying natural communities involved in alternate modes of methane oxidation, such as aerobic versus anaerobic and nitrate (other factor)-dependent versus strictly oxygen dependent modes, in order to bridge gaps in understanding the specialized bacterial communities involved in these processes, including discovery of novel species and novel metabolic pathways involved in these processes. This project takes advantage of a large metagenomic dataset enriched in the DNA of the methanotrophic species employing different types of methane metabolism, generated in collaboration with the Joint Genome Institute. The specific objectives of this project are:**

**Objective 1. Identify actively transcribed pathways**. Via next-generation sequencing-based transcriptomic profiling of four types of microcosm communities stimulated with methane (aerobic, anaerobic, aerobic with added nitrate and anaerobic with added nitrate), identify species actively involved in methane oxidation, based on expression patterns, and delineate specific pathways for methane oxidation and electron transfer.

**Objective 2. Identify physiologically active pathways.** Via methane and nitrogen flux analysis and via community metabolomics employing chromatography-tandem LC-MS/MS (quadrupole mass spectrometry), GC x GC-TOFMS (comprehensive two-dimensional Gas Chromatography-Time-Of-Flight Mass Spectrometry), follow dynamics of methane/ nitrate consumption and identify major metabolites indicative of specific metabolic pathways, and correlate gene expression profiles and predicted methane utilization pathways in the four simulated settings, as above.

**Objective 3. Identify activity of individual cells.** Via FISH (fluorescent *in situ* hybridization) or HISH (halogen *in situ* hybridization) -coupled nano-SIMS (Secondary Ion Mass Spectrometry) employing specific DNA probes and labeled methane and nitrate, couple a specific function in methane metabolism to a specific functional guild in the community, following the fates of methane and nitrate (and other characteristic ions) at the single cell resolution level.

In this work we integrate heterogeneous multi-scale genomic, transcriptomic, and metabolomic data to redefine the metabolic framework of $CH_4$ oxidation from single ecosystem-relevant microbial species (such as methanotrophic isolates from Lake Washington sediment) as well as a natural ecosystem (Lake Washington sediment). Draft genome sequences of model methanotrophic cultures were generated and manually annotated. Preliminary reconstruction of the C1-metabolic pathways was performed. In order to underline core methylotrophic functions a set of RNA-seq based transcriptomic experiments were carried out. In addition metabolomic experiments were performed to define C1-metabolic networks of type I and type II methanotrophic bacteria and to identify potential metabolic markers of methanotrophy. Around 60 metabolites including amino acids, carboxylic acids, sugar phosphates, nucleotides and CoA derivatives were quantified by LC-MS/MS and GC-MS. To elucidate the major metabolic flux for methane assimilation, $^{13}C$ –isotopic labeling experiments were performed.

New data reveal an unexpected carbon distribution in model proteobacterial methane oxidizers, providing a new version of the methanotrophy metabolic network. Similar system-level approaches were applied for characterization of the metabolic potential of yet uncultivable microbes from natural environments.

# 168

## Genes, Isotopes, and Ecosystem Biogeochemistry: Dissecting Methane Flux at the Leading Edge of Global Change

**Scott Saleska**[1]* (saleska@email.arizona.edu), Virginia Rich[1]* (vrich@email.arizona.edu), Gene Tyson,[2] Patrick Crill,[3] Changsheng Li,[4] Steve Frolking,[4] and Jeff Chanton[5]

[1]University of Arizona; [2]University of Queensland; [3]Stockholm University; [4]University of New Hampshire; and [5]Florida State University

http://eebweb.arizona.edu/faculty/saleska/research.htm#isogenie
http://isogenie.pbworks.com/w/page/23915054/FrontPage

**Project Goals: Our project focuses on understanding the biological and earth science aspects of $CO_2$ and $CH_4$ cycling at "the leading edge of climate change"—a subarctic wetland system where climate change-induced permafrost melt is transforming methane sinks into sources. Our research goals are: (1) to discover functional relations for scaling microbial community composition and metabolism to the ecosystem biogeochemistry of $CH_4$ and $CO_2$; (2) to learn how these relations are affected by shifting environmental variables, and (3) apply this knowledge to better understand and predict changing carbon budgets in subarctic ecosystems already experiencing substantial climate change. To achieve these, we aim to: (A) Characterize microbial community composition and metabolic function associated with in situ methano -genic and -trophic pathways; (B) Characterize ecosystem biogeochemical cycling of $CH_4$ and $CO_2$, using isotopes to partition contributions from separate methano -genic and -trophic metabolisms.**

**Microbial communities in northern wetlands are central to understanding current and future global carbon cycling.** Northern wetlands are both critical, contributing a tenth of global $CH_4$ emissions and containing one-quarter of global soil carbon, and vulnerable, with permafrost area expected to shrink 50% by 2050. As permafrost thaws, increasing $CH_4$ emissions from northern wetlands are likely to cause positive feedback to atmospheric warming. Wetland $CH_4$ cycling is mediated by microbes, but connecting ecosystem-scale fluxes to underlying microbial population dynamics and genomics has not been achieved. Recent transformative technical advances in both high-throughput investigations of microbial communities and high temporal-resolution biogeochemical isotope measurements together now permit a uniquely comprehensive approach to opening the microbial "black boxes" of wetland methane cycling that impact carbon cycling on global scales.

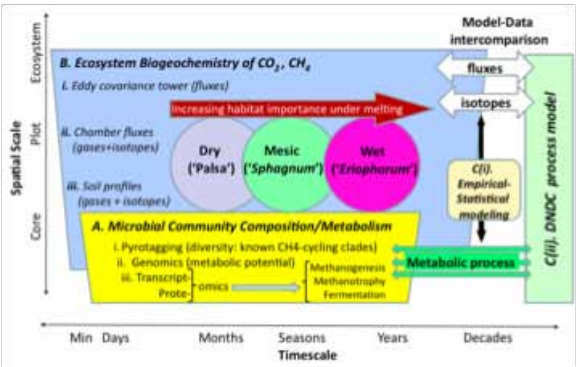
**We are investigating how microbial community composition and function scale to ecosystem biogeochemistry of $CH_4$ and $CO_2$, and how such scaling is affected by climate change.** To accomplish this, we employ a three-pronged interdisciplinary investigation of Sweden's Stordalen Mire, an established wetland field site at the thawing southern edge of the discontinuous permafrost zone:

**(A) Molecular microbial ecology** *to identify the genes and lineages that mediate $CH_4$ cycling through the soil column, across the major wetland habitats, and over the growing season, using:* (i) *pyrotagging*: profiling community diversity using the 16S rRNA gene, (ii) *metagenomics*: community metabolic potential by bulk sequencing of microbial DNA, (iii) *metatranscriptomics*: community expressed genes by sequencing microbial mRNA, and (iv) *metaproteomics*: mass spectrometry analysis of community proteins. Coupled to biogeochemistry and modeling, these methods link microbes, genes, transcripts and proteins with biogeochemical processes and ecosystem fluxes. (*Investigators: Tyson and Rich*)

**(B) Continuous biogeochemical measurements** *of $CO_2$ and $CH_4$ fluxes and isotopic compositions to quantify carbon characteristics and cycling at three spatial scales*: (i) *ecosystem*, through an in-place eddy flux tower, (ii) *site*, through an in-place system of autochambers, and (iii) *soil profile*, through an in-development system of soil gas samplers. C isotopes of $CH_4$ and $CO_2$ at scales (ii) and (iii) will be automatically measured in the field using a recently developed laser spectrometer, and H isotopes of $CH_4$ and $H_2O$ will be analyzed by traditional IRMS. Acetate and dissolved carbon species will also be quantified and isotopically characterized. The average age of the mineralized organic matter will be measured via the $^{14}C$ ratios of $CO_2$ and $CH_4$. (*Investigators: Saleska, Crill, and Chanton*)

**(C) Modeling** *soil processes and ecosystems, to characterize the details of $CH_4$ production, and to test the importance of microbial ecology to ecosystem biogeochemistry.* (i) *Gas diffusion and fractionation modeling*: Stable isotope and flux data will be incorporated into a diffusion model to discriminate between methanogenesis pathways, and quantify $CH_4$ oxidation. This will identify the zones and times of maximum and minimum methanogenesis and methanotrophy, as well as transitions between types of methanogenesis. (ii) *Ecosystem process modeling using the Wetland-DNDC model*, which simulates wetland carbon gas fluxes. We will first test this model against basic flux data from the site; second, develop the model to include isotopes, followed by testing against isotope data; third, compare the model's separately simulated methane production and consumption processes with the corresponding observed microbial functional activity, as recorded in metatranscriptomic and -proteomic data; and fourth, use the refined Wetland-DNDC to project the impacts of continued permafrost thaw on wetland $CH_4$ cycling at this site. (*Investigators: Chanton, Li and Frolking*).

Overview of Technical Approach:




Project Grant DE-SC0004632

# 169

## Plant Stimulation of Soil Microbial Community Succession: How Sequential Expression Mediates Soil Carbon Stabilization and Turnover

Shengjing Shi[1]* (shengjing.shi@berkeley.edu), Jeremy Bougoure,[2] Nicholas Bouskill,[3] Erin Nuccio,[1] Donald Herman,[1] Romain Barnard,[1] Patrik D'Haeseleer,[2] Jennifer Pett-Ridge,[2] Eoin Brodie,[3] Zhili He,[4] Jizhong Zhou,[4] and **Mary Firestone**[1]

[1]Department of Environmental Science, Policy and Management, University of California, Berkeley; [2]Chemical Sciences Division, Lawrence Livermore National Laboratory, Livermore, Calif.; [3]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley; and [4]Institute for Environmental Genomics, Department of Botany and Microbiology, University of Oklahoma, Norman

**Project Goals: The goal of our project is to understand the succession of soil microbial communities involved in degradation and transformation of C released from the roots of *Avena barbata*, a grass common to California. This project involves the identification of metabolic pathways of soil microorganisms responsible for C processing by labeling *A. barbata* with $^{13}CO_2$ and the stable isotope probing technique. Fluorescent in situ hybridization probing and NanoSIMS analysis will be used to identify active community members consuming $^{13}C$ derived from roots and visualize spatial associations between microbes and plant roots. The influence of elevated $CO_2$ on succession and sequential expression of microbial function will also be investigated using transcriptome-enhanced GeoChip, CAZy/FOLy functional microarrays, and second generation sequencing. Gene expression networks will then be identified in order to model the microbial community cascades which enable C stabilization and turnover.**

Soil organic carbon (C) is the largest pool within the terrestrial C cycle and is derived from decomposed plant and soil microbial materials. The fluxes that control the size of this pool are critical to the global C cycle. It is known that plant roots and their exudates exert control over the microorganisms mediating decomposition of complex soil C compounds. Over time, living roots become root debris and undergo decomposition by soil microorganisms, ultimately entering stabilized pools. Therefore, the change over time of the composition and function of the C-degrading and transforming microbial communities associated with living and decomposing roots defines a central biological component of soil C stabilization. Our project focuses on the succession of soil microbial communities involved in degradation and transformation of C released from the roots of *Avena barbata*, a grass common to California. Metabolic pathways of soil microorganisms responsible for C processing will be identified by labeling *A. barbata* with $^{13}CO_2$ and distinguishing the functional C-cycling transcriptome by stable isotope probing of rhizosphere microbial communities. Fluorescent in situ hybridization probing and NanoSIMS analysis will be used to identify active community members consuming $^{13}C$ derived from roots and visualize spatial associations between microbes and plant roots. The influence of elevated $CO_2$ on succession and sequential expression of microbial function will also be investigated using transcriptome-enhanced GeoChip, CAZy/FOLy functional microarrays, and second generation sequencing. Gene expression networks will then be identified in order to model the microbial community cascades which enable C stabilization and turnover.

# 170

## Microbial Food Web Mapping: Linking Carbon Cycling and Community Structure In Soils Through Pyrosequencing Enabled Stable Isotope Probing

**Daniel H. Buckley*** (dbuckley@cornell.edu)

Cornell University

**Project Goals: see below**

The terrestrial biosphere contains a large fraction of global C and nearly 70% of the organic C in these systems is found in soils. The majority of C 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.

In their metabolic diversity and their sheer numbers microbes dominate the soil processes that underlie ecosystem processes in terrestrial environments. Yet, the magnitude of their contribution lies in stark contrast to the little we

know about soil microorganisms and the principles that govern their function. Soil microorganisms do not behave as a homogeneous trophic level, and the composition of the soil microbial community can influence microbial processes in the soil both qualitatively and quantitatively. As a result, in order to understand changes in the terrestrial C-cycle in response to environmental change it is important to examine the internal dynamics of soil microbial communities and the manner in which they influence community function.

While we have made strides in understanding environmental controls on decomposition we still lack a coherent concept of the soil microbial food web. There is a general assumption of functional equivalence for different soil microbial communities with respect to the soil C cycle, but the validity of this hypothesis has been questioned. This deficiency in our knowledge results from the absence of in situ methods for identifying microorganisms involved in the soil C cycle and as a result we have a glaring lack of information about which organisms actually mediate critical soil processes. Our research program will approach this fundamental problem by developing a method for pyrosequencing enabled stable isotope probing (SIP) to dissect the microbial food web. Pyrosequencing enabled 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. We will use this groundbreaking approach to examine whether changes in soil community composition are ecologically significant with respect to their impact on the soil C-cycle. Project objectives include:

- Develop and validate protocols for pyrosequencing enabled $^{13}C$-SIP of nucleic acids.
- Determine whether carbon input parameters (composition, quantity, timing) alter the route of C through the soil community and whether these shifts are driven by microbial community structure.
- Evaluate whether microbial community structure is functionally equivalent with respect to the C-cycle across edaphically related soils that differ in management history.

The method and the results generated by this project will improve our ability to examine the impacts of management decisions, soil history, and environmental change on the behavior of microbial communities in terrestrial ecosystems, revealing the ecological mechanisms by which microbes regulate both C mineralization and C retention in soils, and improving our ability to predict changes in terrestrial ecosystem processes in the face of accelerating global change.

# 171

## Development of Novel Random Network Theory-Based Approaches to Identify Network Interactions in Soil Microbial Communities

Ye Deng,[1,2]* Zhili He,[2] Yujia Qin,[2] Feifei Liu,[1,2] Jason Shi,[1,2] Qichao Tu,[2] Joy D. Van Nostrand,[1,2] and **Jizhong Zhou**[2]

[1]Glomics Inc., Norman, Okla.; and [2]Institute for Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, Norman

**Project Goals: High-throughput metagenomics technologies can rapidly produce massive data sets, but one of the greatest challenges is how to extract, analyze, synthesize, and transform such vast amounts of information to biological knowledge. Thus, the goal of this study is to develop a novel conceptual framework and computational approaches based on a new mathematical approach, random matrix theory (RMT), using large scale, high throughput metagenomics sequencing and hybridization data.**

The interactions among different microbial populations in a community play critical roles in determining ecosystem functioning but very little is known about the network interactions in a microbial community due to the lack of appropriate experimental data and computational analytic tools. High-throughput metagenomics technologies can rapidly produce massive data sets, but one of the greatest challenges is how to extract, analyze, synthesize, and transform such vast amounts of information to biological knowledge. Thus, the goal of this study is to develop a novel conceptual framework and computational approaches based on a new mathematical approach, random matrix theory (RMT), using large scale, high throughput metagenomics sequencing and hybridization data.

We have developed RMT-based conceptual framework for identifying functional molecular ecological networks (fMENs) with the high throughput functional gene array (GeoChip) hybridization data of soil microbial communities in a long-term grassland FACE (Free Air $CO_2$ Enrichment) experiment. Our results indicated that RMT is powerful in identifying fMENs in microbial communities. Both fMENs under elevated $CO_2$ ($eCO_2$) and ambient $CO_2$ ($aCO_2$) possessed the general characteristics of complex systems such as scale-free, small-world, modular and hierarchical. However, the topological structure of fMENs is distinctly different between $eCO_2$ and $aCO_2$, at the levels of the entire communities, individual functional gene categories/groups, and functional genes/sequences, suggesting that $eCO_2$ dramatically altered the network interactions among different microbial functional genes/populations. Such shifts in functional network structure are also significantly correlated with soil geochemical factors.

We have used this new approach to identify phylogenetic molecular ecological networks (pMENs) using metagenomics sequencing data of 16S ribosomal RNA (rRNA)

genes from soil microbial communities in the FACE site as well. Our results demonstrate that the RMT-based network approach is also very useful in delineating pMENs in microbial communities. The structure of the identified networks under $aCO_2$ and $eCO_2$ was substantially different in terms of overall network topology, network composition, node overlap, module preservation, module-based higher order organization (meta-modules), topological roles of individual nodes, and network hubs. These results suggested that $eCO_2$ dramatically altered the network interactions among different phylogenetic groups/populations. In addition, the changes in network structure were significantly correlated with soil carbon and nitrogen content, indicating the potential importance of network interactions in ecosystem functioning. To our knowledge, these are the first studies to demonstrate the changes in network interactions of microbial communities in response to $eCO_2$. In a word, elucidating network interactions in microbial communities and their responses to environmental changes is fundamentally important for research in microbial ecology, systems microbiology, and global change.

Student Oral Presentation–Tuesday

# 172

## From Community Structure to Functions: Metagenomics-Enabled Predictive Understanding of Temperature Sensitivity of Soil Carbon Decomposition to Climate Warming

**Jizhong Zhou**[1,*] (jzhou@ou.edu), Liyou Wu,[1] Kai Xue,[1] Lei Cheng,[1] Jin Zhang,[1] Ye Deng,[1] Joy D. Van Nostrand,[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] Xu Xia,[1] Christina Schädel,[1] Qichao Tu* (qichao@ou.edu), and **Yiqi Luo**[1]

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

**Project Goals: see below**

Understanding the responses, adaptations and feedback mechanisms of biological communities to climate change is critical to project future states of the earth and climate systems. Although a significant amount of knowledge is available on the feedback responses of aboveground communities to climate change, little is known about the response of belowground microbial communities due to the challenges in analyzing soil microbial community structure and functions. Thus, the overall goal of this study is to provide system-level, predictive mechanistic understanding of the temperature sensitivity of soil carbon decomposition to climate warming by using cutting-edge integrated metagenomic technologies. Towards this goal, the following four objectives will be pursued: **(i)** To determine phylogenetic composition and metabolic diversity of microbial communities in the temperate grassland and tundra ecosystems; **(ii)** To delineate the responses of microbial community structure, functions and activities to climate change in the temperate grassland and tundra ecosystems; **(iii)** To determine the temperature sensitivity of microbial respiration in soils with different mixtures of labile versus recalcitrant carbon, and the underlying microbiological basis for temperature sensitivity of these pools; and (**iv**) To synthesize all experimental data for revealing microbial control of ecosystem carbon processes in responses to climate change. We have carried out our studies at two contrasting long-term experimental facilities, the temperate grassland ecosystems in OK and tundra ecosystems in Alaska.

**Feedback responses of microbial communities to warming**. We have used integrated metagenomic technologies to analyze the responses of microbial communities in a long-term experimental warming grassland ecosystem in Oklahoma. Our results revealed at least three lines of evidence for microbial mediation of carbon cycle feedback to climate warming. First, long-term experimental warming induced a decline in temperature sensitivity of heterotrophic soil respiration due to soil microbial community adaptation. Second, warming significantly stimulated functional genes for labile but not recalcitrant carbon decomposition, which may promote long term stability of ecosystem carbon storage. Third, warming stimulated functional genes for nutrient cycling, favoring plant growth and vegetation carbon uptake. Our results indicate that microorganisms critically regulated ecosystem carbon cycle feedback to climate warming, with important implications for the carbon-climate modeling. Similar metagenomics analysis of soil microbial community in tundra ecosystems in Alaska are in progress.

**Temperature sensitivity**. We are conducting constant temperature (15°C) soil incubations from the two research sites. Temperature sensitivity is being measured by exposing soils to cycling temperatures at several time points during the 1-year incubation. In a single temperature cycle, soils are exposed to 6 temperature levels from 5 to 30°C over a 36 hour span; soil temperature is increased for 6 hours, with flux measurements made in the second half of that period after the temperature has equilibrated. In first 90-day period we are conducting this temperature cycle treatment 3 times. Then, we are performing this cycle 1 time in each 90-day period thereafter for a total of 6 temperature cycles over the period of 1-year. These temperature cycle experiments are a critical component of this study and are designed to disentangle the temperature sensitivity of the different microbial communities from differences in substrate availability that can confound comparisons between long-term incubations at different temperatures. From these experiments we can calculate $Q_{10}$ relationships for different soils and different conceptual pools of soils within our modeling framework described later in this proposal. We expect that this experimental design will allow us to explore temperature sensitivity of different components of the organic matter

pool through time, and as it relates to the active microbial community as determined by the genomic analyses.

**Development of amplicon-sequencing approaches for uncovering functional gene diversity**. Functional genes involved in recalcitrant carbon degradation and N fixation were selected for pyrosequencing on the 454 titanium platform. The *ligDFEG* gene cluster is involved in the cleavage of the ß-aryl ether bond, the most abundant linkage in lignin (50-70% of total). Of these, *ligE*, a glutathione S-transferase domain protein, was targeted based on the sequence abundance in GenBank. Primer sequences were developed from a Hidden Markov Model derived compilation deposited into RDP's Fungene database. Second, the aromatic ring hydroxylating dihydrogenase (ARHD) gene, known for biphenyl degradation, was selected as this structure constitutes ~10% of lignin. The *nifH* gene was selected for its important contribution to soil fertility. Prior evidence indicates that *nifH* diversity responds to warming and is correlated with changes in soil carbon. Lastly, pyrosequencing will be performed on the *16S rRNA* gene to investigate changes in microbial community composition with treatment.

**Evaluation of bioinformatic approaches for assembling short read sequences**. We have also begun to explore the next-generation sequencing technologies (NGS) for characterizing our microbial communities. We have performed an initial assessment of the technical shortcomings of NGS such as what fraction of the total diversity in a sample can be recovered by NGS and what are the types and frequencies of errors in assembled genes from complex communities. To this end, we compared the two most frequently used sequencing platforms, the Roche 454 FLX Titanium and the Illumina Genome Analyzer (GA) II, on the same DNA sample obtained from a complex freshwater microbial community. We found that the two platforms provided a very comparable view of the community sampled, e.g., the derived assemblies overlapped in ~90% of their total sequences. Evaluation of base call error, frameshift rate, and contig length suggested that short reads (Illumina) offered equivalent, if not better, assemblies with longer reads (454). We are currently expanding our approaches to enable the use of higher volume data from more complex (soil) microbial communities. Such analyses will provide strategies for shotgun sequencing the entire microbial communities.

**Modeling integration and development**. We use data assimilation techniques to integrate soil incubation data into a model to optimize parameter estimation of different soil organic carbon (SOC) pools (labile to recalcitrant) with different turnover times. Soil carbon dynamics were described by a 1st-order differential equation. Model parameters were optimized using Bayesian probability inversion and a Markov chain Monte Carlo (MCMC) technique, which generates posterior probability density functions of model parameters. Preliminary results show that soil carbon efflux data of less than 200 days of incubation length only constrained parameters of a 1-pool model indicating that daily C decomposition by the end of the experimental period was still dominated by the $CO_2$ efflux of the labile SOC pool. Long-term datasets are required to constrain parameters of multiple carbon pool models in order to determine temperature sensitivities of carbon pools with slower turnover rates.

# 173

## Meta-"omics" Analysis of Microbial Carbon Cycling Responses to Altered Rainfall Inputs in Native Prairie Soils

**David D. Myrold**[1]* (david.myrold@oregonstate.edu), Peter J. Bottomley,[1] Maude M. David,[3] Robert L. Hettich,[4] Janet K. Jansson,[3] Ari Jumpponen,[2] Charles W. Rice,[2] Susannah G. Tringe,[3] Nathan C. VerBerkmoes,[4] Stephanie A. Yarwood,[1] and Lydia H. Zeglin[1]

[1]Oregon State University; [2]Kansas State University; [3]Lawrence Berkeley National Laboratory; [4]Oak Ridge National Laboratory

**Project Goals: see below**

Soils process and store large amounts of C; however, there is considerable uncertainty about the details that influence microbial partitioning of C into soil C pools and what influential forces control the fraction of the C input that is stabilized. It is not clear how the microbial community will respond to climate-induced modification of precipitation and inter-precipitation intervals, and if this response will affect the fate of C deposited into soil by the local plant community. Part of this uncertainty lies with our ignorance of how the microbial community adapts physiologically or in composition to changes in soil moisture brought about by shifts in precipitation.

Our **goal** is to harness the power of multiple meta-omics tools to understand the functioning of whole-soil microbial communities and their role in C cycling. Towards this end, we will:

1. Further develop and optimize a combination of meta-omics approaches to study how environmental factors affect microbially-mediated C cycling processes.
2. Determine the impacts of long-term changes in precipitation timing on microbial C cycling using an existing long-term field manipulation of a tallgrass prairie soil.
3. Conduct laboratory experiments that vary moisture and C inputs to confirm field observations of the linkages between microbial communities and C cycling processes.

We will study the rainfall manipulation plots (RaMPs) established more than a decade ago at the Konza Prairie in Kansas. The RaMPs experiment consists of two treatments: (1) natural precipitation, and (2) extended precipitation interval. The extended precipitation interval consists of storing rainfall and reapplying it at an interval 50% longer than that between natural rainfall events.

We will employ a systems biology approach, considering the complex soil microbial community as a functioning system

and using state-of-the-art metatranscriptomic, metaproteomic, and metabolomic approaches. These omics tools will be refined, applied to field experiments, and confirmed with controlled laboratory studies. Our experiments are designed to specifically identify microbial community members and processes that are instrumental players in processing of C in the prairie soils and how these processes are impacted by wetting and drying events.

At this time, we have made progress towards the development and optimization of the methods we will use for metatranscriptomic and metaproteomic samples. We have set up a short-term laboratory experiment using soil collected from the Konza site. Soil received two treatments: (1) control or inoculated with an *Arthrobacter* strain whose genome has been sequenced, and (2) control or amended with acetate (a generic source of readily available C) or 4-chlorophenol (a xenobiotic C compound that is metabolized by the added *Arthrobacter* stain). After incubation, soils were extracted for either RNA or protein and gene expression and protein abundance have been assessed. See the David et al. poster for a summary of the initial results from this experiment.

Additional laboratory experiments have been planned, which will address the following questions: 1) Is there a temporal lag between peak transcriptional activity and peak biomass/protein production? 2) Do different soil microbial taxa display different strategies to cope with modified wet/dry cycles, both in terms of temporal patterns of response and physiological adaptations that control utilization of C and allocation of C within/without the cell? The designs of these experiments will be presented.

In early summer 2011 we will conduct the first of two samplings of the RaMPs experiment. The first sampling will occur when plant growth is optimal and will be timed to collect soils prior to a major rainfall event (when soils are relatively dry), shortly after wet-up, and at one more time as the soils begin to dry. At each of these times, we will measure soil community properties (pyrotagged analysis of rRNA DNA and rRNA, metatranscriptome, and metaproteome) and soil C cycling processes (soil respiration, extracellular enzyme activities, etc.).

This research is sponsored by the DOE-BER, Biological Systems Research on the Role of Microbial Communities in Carbon Cycling Program, under Contract No. DE-SC0004953.

# 174

## Assessing the Microbial Basis of Carbon Cycling in Prairie Soils with an Integrated Omics Approach

Maude M. David[1]* (MMDavid@lbl.gov), Lydia Zeglin,[2] Manesh Shah,[3] Stephanie Yarwood,[2] Konstantinos Mavromatis,[4] Susannah Tringe,[4] Robert Hettich,[3] Nathan VerBerkmoes,[3] David Myrold,[2] and **Janet K. Jansson**[1,4]

[1]Earth Sciences Division, Lawrence Berkeley National Lab., Berkeley, Calif.; [2]Oregon State Univ., Corvallis; [3]Oak Ridge National Lab., Oak Ridge, Tenn.; and [4]Joint Genome Institute, Walnut Creek, Calif.

**Project Goals: This project aims to evaluate the carbon cycling potential of the soil microbiota in Kansas native prairie soils using different omics approaches.**

Soils of the Great Prairie store more than 30% of the terrestrial organic carbon in the continental U.S. This area is expected to experience dramatic changes in precipitation patterns, with either longer drought periods or more rainfall, as a consequence of climate change. However, the impact of a changing climate on the stored carbon pools in these soils is currently not known. This project aims to evaluate the carbon cycling potential of the soil microbiota in Kansas native prairie soils using different omics approaches.

In order to generate a database for the "omics" analyses, 15GB of DNA sequence were obtained via Illumina sequencing from the Kansas prairie soil. A total of 102 million reads were assembled using a CLC Genomics workbench, and we obtained 1.4 million contigs of at least 300 bp in length. Those contigs were submitted to IMG for gene prediction and annotation.

To develop protocols for extraction of RNA and protein from soil, we used a model bacterium that has been genome sequenced, *Arthrobacter chlorophenolicus*, as an inoculum. The model strain was inoculated into liquid medium and non-sterile Kansas prairie soil, and acetate and 4-chlorophenol were added as general and specific carbon substrates, respectively. The same substrates were added to the soil without inoculum to assess the response of the indigenous microbes in the soil.

The RNA was extracted from soil using a MoBio kit and treated with DNase. Key target genes were quantified by quantitative PCR and RT-QPCR and the metatranscriptomes were obtained by Illumina sequencing. A first set of genes, 16SrRNA and gfp (encoding the green fluorescence protein that was stably inserted in the chromosome of the *A. chlorophenolicus* strain used) were chosen to estimate the *A. chlorophenolicus* cell number. Two other genes, *ICL* (isocitrate lyase, part of the 2-C bypass to the TCA cycle), and *suCAB* (succinyl CoA, part of the main TCA cycle) were used to track the pathways used by the microorganisms with different substrates.

To study the metaproteome, we first optimized a method for extraction of proteins from soil, based on a combination of lysis with SDS, boiling/sonication, followed by TCA precipitation, protein digestion and a size exclusion filtration (modified from Chourey et al, 2010). With this optimized protocol over 1000 proteins were identified via 2d-LC-MS/MS on an LTQVelos per soil microcosm. This protocol allowed us to directly characterize the metaproteome in soils with high levels of humic acids. The resulting peptides were searched against 4 databases: 1) the assembled reads of the metagenome from Kansas prairie soil, 2) this database complemented with the unmapped reads that didn't assemble into contigs, 3) 16S rRNA gene sequences data from the same soil was used to build a selected reference isolate genome database, and 4) the Arthrobacter isolate genome for those samples that had been inoculated with the model strain.

The first RNA results show that gfp transcript could be detected under most conditions and was thus a good estimator of the cell activity and abundance in the samples. Different metabolic pathways were more predominantly expressed depending on the substrates and incubation conditions used. In addition, the initial metaproteome data indicated that several of the enzymes involved in the degradation pathways of the particular organic substrate amended to the microcosms were expressed in soil and in liquid media. In general, several proteins involved in response to stress (thioredoxin, chaperonin, cold-shock proteins, etc.), were expressed in the prairie soil.

These initial studies will define the best techniques for omics applications in soil as well as baseline information about the carbon cycling potential in the Great Prairie of the U.S. The next step will be to determine the impact of altered precipitation on the soil microbial community and carbon cycling functions in these soils.

**Reference**

1. Chourey, K; Jansson, J; Verberkmoes, N; Shah, M; Chavarria, KL; Tom, LM; Brodie, EL; Hettich, RL. "Direct cellular lysis/protein extraction protocol for soil metaproteomics". *Journal of Proteome Research*. 2010 12:6615-22.

# 175

## The Impact of Global Warming on the Carbon Cycle of Arctic Permafrost: An Experimental and Field Based Study

**T.C. Onstott**[1]* (tullis@princeton.edu), **S. Myneni**,[1] B. Stackhouse,[1] **S.M. Pfiffner**,[2] T. Vishnivetskaya,[2] A. Layton,[2] **L.G. Whyte**,[3] N. Mykytczuk,[3] J. Allan,[3] R.C. Wilhem,[3] **R. Hettich**,[4] K. Chourey,[4] **T.J. Phelps**,[4] **S. Elliott**,[5] M.F. Kidd,[5] and P. Hatcher[6]

[1]Princeton Univ.; [2]Univ. of Tennessee; [3]McGill Univ.; [4]Oak Ridge National Laboratory; [5]Los Alamos National Laboratory; and [6]Old Dominion Univ.

http://www.princeton.edu/southafrica/DOEpermafrostproject

### Project Goals:

1. **Perform ~2 years long, heating experiments on well-characterized, intact cores of Arctic active-layer and permafrost from a proposed reference ecosystem site where $CO_2$, $N_2O$ and $CH_4$ fluxes, temperatures, humidity, soil moisture, nutrients, microbial diversity and activities and isotopic analyses are currently being measured in the field.**
2. **Perform phylogenetic, metagenomic, transcriptomics and proteomic analyses of the intact cores.**
3. **Characterize the abundance and composition of the solid and dissolved organic matter and the inorganic geochemistry in the active layer and permafrost.**
4. **Characterize changes in the organic matter composition, 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 during the long term heating experiments as the permafrost thaws under water saturated and water under saturated conditions.**
5. **Compare the results from intact cores and the heating experiments with field measurements.**
6. **Based upon these heating experiments and field measurements 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.**

Permafrost, or perennially frozen ground, underlies ~24% of the Earth's surface and contains ~1/3 of the global soil organic C. It is, therefore, a possible source of extremely potent greenhouse gases, such as $CH_4$, $N_2O$ and $CO_2$. Temperatures in the Arctic may increase 4-8°C over the next 100 years, thereby increasing the depth of the active-layer and thawing the underlying permafrost. Field observations and ice core records suggest that with thawing, the relatively undegraded permafrost organic C will be rapidly metabolized, creating a positive feedback to global warming through increased $CH_4$, $N_2O$ and $CO_2$ emissions. Although many researchers have measured $CO_2$ and $CH_4$ fluxes and characterized the microbial diversity of the

Siberian and Canadian active-layer and permafrost, the relationship between methanogenic, methanotrophic and heterotrophic *in situ* activities within the active-layer and $CO_2$ and $CH_4$ fluxes as a function of temperature has not been delineated either in field or lab experiments. Defining these relationships is essential for determining the extent and rate of this positive feedback in order for these processes to be accurately reflected in global climate models.

To address this paucity of data we will collect 30 intact cores for long term heating experiments from the active layer and permafrost (0-1 meters below surface) at the McGill Arctic Research Station (MARS) on Axel Heiberg Island in the Canadian high Arctic. This site has been the location of climate investigations for the past 50 years. The extensive Arctic wetland area adjacent to the lake at this site will be the source of the intact cores (Fig. 1 top). $CO_2$ flux measurements indicate higher emissions during peak summer months compared to spring time when the ground is completely frozen. The microbial community of the active layer and associated permafrost varies as a function of depth with aerobic phyla dominant near the surface and anaerobic phyla dominant at greater depths. Archaeal phyla, including methanogens, comprise 0.1% of the microbial community. A low diversity fungal component is also present in the active layer and permafrost. Pore water geochemical results also indicate decreasing $O_2$ availability as a function of depth and proximity to the lake. The total organic carbon concentration is uniform with depth, but the dissolved organic carbon varies dramatically with depth with values as high as 100 mM. Aerobic viable cell counts from the active layer are also 100 times those of the permafrost. Anaerobic incubation experiments utilizing organic carbon amendments detect enhanced production of $CO_2$ and $CH_4$ with increasing temperature up to 15˚C, relative to undetectable $CO_2$ and $CH_4$ production at 0˚C.

Based upon last summer's coring campaign a new coring bit has been manufactured that will enable the collection of 1-meter long intact cores within polycarbonate tubes that can be sealed and frozen on site (Fig. 1 bottom). Heating experiments will be performed on these tubes in the lab where the temperature, precipitation and humidity of the headspace will be controlled and the gaseous, aqueous and solid phase constituents analyzed over time. Experiments will be performed under both water-saturated conditions and partially saturated conditions that reflect the observed variations in the water table depth of the site. The ongoing incubation experiments will be utilized to design the timing of sacrificial core analyses and the type of organic substrates that will be added in a subset of the cores. Organic and inorganic geochemical, metagenomic, metatranscriptomic and proteomic profiles will be performed on the core samples prior to and during prolonged heating to address questions of nutrient fluxes, diversity, abundance, activity and spatial relationships between microorganisms, respectively. A high sensitivity $^{14}C$ RNA isotope microarray will be developed based upon the observed 16S rRNA community structure that will map the carbon trophic cascade as permafrost with $^{14}C$ labeled compounds thaws. Cavity ring down spectrometers will be used to monitor C, O and H isotopic analyses of $CH_4$ and $CO_2$ from the heated cores, from cores amended with $^{13}C$ labeled compounds and from permafrost emissions at the MARS field site.

Figure 1. (top) Aerial view of the McGill Arctic Research Station showing polygonal terrain adjacent to Colour Lake. (bottom) Coring of wetland active layer and permafrost during July 2010.

# 176

## Characterization of Active Layer and Permafrost from High Canadian Arctic Soil and Potential Effects of Climate Change on Greenhouse Gas Emissions

**T.C. Onstott**,[1] **S. Myneni**,[1] B. Stackhouse[1]* (bstackho@princeton.edu), **S.M. Pfiffner**,[2] T. Vishnivetskaya,[2] A. Layton,[2] **L.G. Whyte**,[3] N. Mykytczuk,[3] J. Allan,[3] R.C. Wilhem,[3] **R. Hettich**,[4] K. Chourey,[4] **T.J. Phelps**,[4] **S. Elliott**,[5] M.F. Kidd,[5] and P. Hatcher[6]

[1]Princeton Univ.; [2]Univ. of Tennessee; [3]McGill Univ.; [4]Oak Ridge National Laboratory; [5]Los Alamos National Laboratory; and [6]Old Dominion Univ.

http://www.princeton.edu/southafrica/DOEpermafrostproject

**Project Goals: The goal of this project is to characterize changes in the geochemical profile and the microbial community of active layer and permafrost soils from a high Canadian Arctic environment based on long-term heating experiments.**

Thawing of permafrost soils and the carbon stores that they contain has been of concern as a potential source of additional $CO_2$ and $CH_4$ to the atmosphere, providing a positive feedback on a warming climate. To assess the potential for increased release of greenhouse gases to the atmosphere, the active layer and permafrost from a moist, acidic soil in the high Canadian Arctic was characterized and used to perform microcosms with $^{13}C$-labeled substrate additions under in situ and simulated warming conditions.

C/N ratios in this soil were between 15:1-18:1 and increased with depth, lower than the values observed in other Arctic soils (20:1-40:1) and exhibiting an inversion of typical soil C/N profiles. High levels of dissolved organic carbon (DOC; 100 mmol C/L) were observed in the subsurface at 30-35 cm depth and decreased into the permafrost while total dissolved nitrogen increased with depth. Porewater composition was dominated by $SO_4^{2-}$ (~1 mM) and contained soluble Fe and Mn at depth, indicating a gradual transition to an anoxic region.

Microcosms from active layer (35 cm) and the permafrost (75 cm) samples were incubated anaerobically with acetate, $CO_2/H_2$, methanol, or unamended, under different temperature regimes, 4˚C and 20˚C. Production of $CH_4$ reached 2.8 nmol/g(FW)/day in $CO_2/H_2$ enriched samples, and $CO_2$ reached 3.5 nmol/g(FW)/day in acetate enriched samples. With warming from an in situ summer temperature of 4˚C to 20˚C, $CO_2$ production increased by an average of 36% across all samples, whereas $CH_4$ production did not increase on average.

This study suggests that upon warming these soils will likely be a net source of $CO_2$ to the atmosphere and the active layer deepens and additional soil carbon is available for degradation. Emissions of $CH_4$, however, do not seemed poised to substantially increase, likely due to the presence of more favorable electron acceptors that prevent the system from fully progressing to methanogenesis.

# 177

## Multi-Scale Modeling of *Methanosarcina* Species

**William W. Metcalf**[1,4]* (metcalf@illinois.edu), Nathan D. Price,[2,4] Rachel J. Whitaker,[1,4] and Ping Ma[3,4]

[1]Departments of Microbiology, [2]Chemical and Biomolecular Engineering, [3]Statistics, and [4]The Institute for Genomic Biology, University of Illinois at Urbana-Champaign

**Project Goals: The long-term goal of the proposed research will be the creation of integrated, multiscale models that accurately and quantitatively predict the role of methanogenic organisms in the global carbon cycle under dynamic environmental conditions. To achieve this goal the following specific aims will be addressed: (1) we will sequence the genomes of numerous members of the Order Methanosarcinales; (2) genomic sources of phenotypic variation will be identified through comparative genomics and tested experimentally using genetic approaches; (3) the transcriptional networks of two *Methanosarcina* species will be established using a combination of experimental and bioinformatic techniques; and (4) we will develop comprehensive metabolic network models across the order Methanosarcinales.**

Methanogenesis is responsible for a significant fraction of the global carbon cycle and plays an essential role in the biosphere. In many anaerobic environments, turnover of organic matter is completely dependent on methanogenic archaea. Although a great deal is known about the physiology and metabolism of these organisms, our ability to incorporate methanogens into carbon cycle models remains in the "black box" stage. To address this issue we are developing systems level models that capture the metabolic and regulatory networks of *Methanosarcina* species, which are among the most experimentally tractable of the methane-producing archaea. Our initial goal is to sequence the genomes of approximately forty strains that span the taxonomic scale within the Order *Methanosarcinales*. These strains include isolates with thoroughly characterized physiology, biochemistry and genetics from both marine and freshwater environments. New strains from additional well-characterized ecosystems are being isolated and characterized for genome sequencing in later stages of the project. The genomic sources of phenotypic variation are being assessed by comparative genomic analysis and by development of comprehensive constraint-based metabolic network models for each strain. To understand the mechanism of adaptation of these organisms to changing environmental conditions, we are examining the evolutionary changes in global transcriptional responses of each organism to a range of growth conditions using the RNA-seq approach. Our ultimate goal is to incorporate the observed regulatory and metabolic networks into integrated, multi-scale models that accurately and quantitatively predict the role of methanogenic organisms in the global carbon cycle under dynamic environmental conditions. At present, we have completed draft sequences of four *Methanosarcina* genomes, with eight more in progress. Manually curated metabolic models have been created and tested for a model freshwater species (*Methanosarcina barkeri*) and a model saltwater species (*Methanosarcina acetivorans*), which will serve as starting points for modeling of our newly sequenced strains. Preliminary RNA-seq experiments have been conducted for *M. acetivorans*. To fully exploit these data, we have developed penalized linear regression for modeling the RNA-Seq read counts. These models explain more than 60% of the variations and can lead to improved estimates of gene expression.

# 178

## Photoacclimation and Adaptation in the Green Lineage: A Case Study of the Picoeukaryote *Micromonas*

Marie L. Cuvelier,[1] Marijke J van Baren,[1]
**PI: Alexandra Z. Worden**[1,2*] (azworden@mbari.org)
**Co-PIs:** Stephen Callister,[2] Richard Smith,[2] and Joshua Stuart[4]

[1]Monterey Bay Aquarium Research Institute, Moss Landing, Calif.; and [2]Ocean Sciences Department, University of California, Santa Cruz

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

**Project Goals: To develop a systems biology approach to the study of the widespread marine alga *Micromonas* and investigate the impact of environmental perturbations on primary production.**

*Micromonas* is a green alga that thrives from tropical to arctic marine systems and belongs to the picophytoplankton (unicellular algae ≤2 µm in diameter). Algae in this size class mediate a significant portion of $CO_2$ uptake in marine ecosystems. Green algae like *Micromonas* also provide simplified model systems for exploring the evolution of members of the 'green lineage', which includes all land plants. In 2009, we published the complete genome sequences of two isolates of *Micromonas*, namely RCC299 and CCMP1545, in collaboration with the JGI and community members. While the genomes provided a wealth of information, knowledge is lacking about the physiology of these algae and their ability to survive under changing conditions such as rising $CO_2$ levels or increased exposure to ultraviolet (UV) radiation resulting from shallowing mixed layers. Photosynthetic organisms have various acclimation mechanisms to compensate for changes in irradiance. Several gene suites have been identified that appear to be related to light-stress and photoprotection. For example, in higher plants such as *Arabidopsis*, ELIPs (early light induced proteins) accumulate transiently when exposed to stress, including high light (HL).

We tested the response of RCC299 to a rapid shift from the maximal-growth light level to HL or HL+UV. A combination of flow cytometry, quantitative PCR and short read transcriptome sequencing was used to explore cell physiology and putative photoprotective genes under experimental manipulations. Results showed that chlorophyll per cell decreased after short term (1 to 2.5 hours) exposure to HL+UV and was lower than in the control under both types of stress, as might be expected. Two general responses were observed for expression of chloroplast-encoded genes: i) those involved in transcription and translation were down-regulated in all conditions, however in HL and HL+UV this shift was visible earlier in the time course; ii) HL+UV treatment resulted in down-regulation of the photosystem II subunits. This effect was also seen, to a lesser extent, in HL conditions. Several nucleus-encoded genes were up-regulated under HL+UV, and to a lesser extent in HL, but not in controls—these genes are implicated in photoprotection. Moreover, ELIPs were more highly expressed in HL+UV but not HL or controls. Perhaps most exciting was expression of two divergent chloroplast-encoded genes—suggesting they represent novel functional aspects of the photosynthetic apparatus. Together, the results indicate that *Micromonas* is adapted for survival in dynamic light environments.

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 and a Moore Foundation Young Investigator Award to AZW.

# 179

## Connecting Genomes to Physiology and Response: Systems Biology of the Green Alga *Micromonas*

Emily Reistetter,[1] Samuel Purvine,[2] Charles Ansong,[2] Janani Shutthanandan,[2] Zbigniew Kolber,[3] Shuangchun Yan,[1] Marijke J van Baren,[1]
**Co-PIs:** Joshua Stuart,[4] Richard Smith,[2] Stephen Callister[2*]
**PI: Alexandra Z. Worden**[1,3] (azworden@mbari.org)

[1]Monterey Bay Aquarium Research Institute, Moss Landing, Calif.; [2]Pacific Northwest National Laboratory, Richland, Wash.; [3]Ocean Sciences Department, University of California, Santa Cruz; and [4]Biomolecular Engineering Department, University of California, Santa Cruz

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.**

Approximately half of global photosynthetic $CO_2$ uptake is performed by marine algae. Thus far, there are no model systems relevant to both marine carbon cycling and evolution of the eukaryotic supergroup Plantae. Development of such a system is important given the onset of climate change and limited understanding of how earth systems will move forward under current perturbations. Prasinophytes are a group of unicellular marine green algae that are evolutionarily distinct from the model green alga *Chlamydomonas*, but are related to land plants and thought to retain characteristics present in the Plantae ancestor. Here, we describe a prasinophyte model system that will be of broad ecological relevance to marine carbon cycling. Our system uses *Micromonas*, a widespread pico-prasinophyte that is exceptional in having a small genome (21 Mb), low gene redundancy and gene 'fusions', which join together domains typically encoded by separate genes. These features are valuable for investigating and assigning functions to genes and domains by their association with a known pathway or

physiological response. Moreover, given its relationship to plants the *Micromonas* system will enable modeling of more general primary producer responses across multiple ecological and evolutionary scales. The project builds on complete genome sequences from two *Micromonas* strains which share at most 90% of their protein encoding gene complements. Differences between these strains include elements of RNAi and vitamin biosynthesis pathways. We are using high-performance chemostats, transcriptomics and high-throughput proteomics to explore fundamental biological questions and their ecological implications. The data will address the known divide between transcriptional and translational responses. Comparative analyses will also allow investigation of pathways for which particular constituents appear to be derived from other lineages. Computational methods will be used to discover how gene activities are regulated and modulated in response to cellular events and processes. Our overall goal is for this system to provide insights on members of the green lineage—both marine algae and plants—including both novel and conserved genetic mechanisms.

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.

# 180

## Soil Metagenomics and Carbon Cycling in Terrestrial Ecosystems

**Cheryl R. Kuske**[1]* (Kuske@lanl.gov),
**Co-authors:** John Dunbar,[1] La Verne Gallegos-Graves,[1] Stephanie A. Eichorst,[1] Carolyn F. Weber,[1] Blaire T. Steven,[1] Gary Xie,[1] Kuan-Liang Liu,[1] Lawrence O. Ticknor,[1] Jean F. Challacombe,[1] Asli Unal,[1] Monica Moya Balasch,[1] Donald R. Zak,[2] Rytas Vilgalys,[3] Christopher W. Schadt,[4] R. David Evans,[5] J. Patrick Megonigal,[6] Bruce A. Hungate,[7] Robert B. Jackson,[3] Andrea Porras-Alfaro,[8] and Susannah Tringe[9]

[1]Los Alamos National Laboratory; [2]University of Michigan; [3]Duke University; [4]Oak Ridge National Laboratory; [5]Washington State University; [6]Smithsonian Environmental Research Center; [7]Northern Arizona University; [8]Western Illinois University; and [9]Joint Genome Institute, Lawrence Berkeley National Laboratory

**Project Goals: The LANL Science Focus Area (SFA) in Soil Metagenomics and Carbon Cycling in Terrestrial Ecosystems combines field and model system studies to identify patterns of soil microbial community response to elevated $CO_2$ and other climate change factors (altered precipitation, soil warming), and to identify key microbial participants and processes in soil carbon partitioning across multiple terrestrial ecosystems.**

Increased plant growth in response to elevated atmospheric $CO_2$ increases carbon inputs to the soil. The collective activities of complex soil microflora determine whether this additional carbon is sequestered in the soil or released back into the atmosphere and thus have local, regional and global impacts on terrestrial carbon storage and cycling. However, currently we only have a rudimentary understanding of the composition of microbial communities, their involvement in processes that represent control points in carbon flux, and the rates at which these processes occur in terrestrial ecosystems. Accurate climate modeling and carbon management scenarios require an understanding of these microbially-mediated processes.

The LANL Science Focus Area (SFA) in Soil Metagenomics and Carbon Cycling in Terrestrial Ecosystems combines field and model system studies to identify patterns of soil microbial community response to elevated $CO_2$ and other climate change factors (altered precipitation, soil warming), and to identify key microbial participants and processes in soil carbon partitioning across multiple terrestrial ecosystems. This poster highlights elements of the SFA directed toward (a) assessment of soil bacterial and fungal communities, and their responses to long-term elevated $CO_2$ across six of the DOE's FACE and OTC field research experiments (http://public.ornl.gov/face) using a variety of metagenomic approaches, (b) identification of active cellulolytic bacteria and fungi in different soils using stable isotope probing, and (c) development of a fungal large subunit rRNA gene databases and classification tool to enable interpretation of fungal metagenome data. Two associated posters provide more detailed information on metagenome projects within the SFA: (a) response of soil cellulolytic fungal communities, as measured by the cellobiohydrolase I gene (*cbhI*), to long-term elevated $CO_2$ in the DOE FACE and OTC sites, plus seasonal patterns and soil *cbhI* gene expression patterns at the NC pine FACE site, and (b) responses of soil bacterial communities comprising biological soil crusts and shrub root zones to long-term elevated $CO_2$ at the NV desert FACE site.

# 181

## Impact of Ecosystem Type and Season on Responses of Fungal Communities and Their Cellulolytic Constituents to Elevated Atmospheric $CO_2$

**Cheryl R. Kuske**[1]* (Kuske@lanl.gov),
**Co-authors:** Carolyn F. Weber,[1] Monica Moya Balasch,[1] La Verne Gallegos-Graves,[1] Rytas Vilgalys,[2] Donald R. Zak,[3] Bruce A. Hungate,[4] Robert B. Jackson,[5] Christopher W. Schadt,[6] R. David Evans,[7] and J. Patrick Megonigal[8]

[1]Los Alamos National Laboratory; [2]Duke University; [3]University of Michigan; [4]Northern Arizona University; [5]Oak Ridge National Laboratory; [6]Washington State University; and [7]Smithsonian Research Institution

**Project Goals: We investigated the impacts of ecosystem type and elevated atmospheric $CO_2$ on cellulolytic fungal communities by conducting a large-scale Sanger sequencing study of the fungal glycosyl hydrolase 7 family cellobiohydrolase I gene (cbhI) across five U.S. Department of Energy FACE sites. At the NC pine FACE sites, we determined seasonal patterns of diversity and composition for the total fungal community (LSU gene) and the cellulolytic fungal guild (cbhI gene), to determine how seasonal dynamics may affect community responses to elevated $CO_2$ and nitrogen (N)-fertilization at this site. The expression patterns of cbhI genes in soils from different depths and treatments are also being characterized.**

Elevated atmospheric $CO_2$ generally increases plant productivity and subsequently increases the availability of cellulose in soil to microbial decomposers. As key cellulose degraders, soil fungi are likely to be one of the most impacted and responsive microbial groups to elevated atmospheric $CO_2$. However, we do not understand how fungal diversity is distributed *in situ,* which fungi contribute most to cellulose degradation, or how they respond to elevated $CO_2$. We investigated the impacts of ecosystem type and elevated atmospheric $CO_2$ on cellulolytic fungal communities by conducting a large-scale Sanger sequencing study of the fungal glycosyl hydrolase 7 family cellobiohydrolase I gene (*cbhI*) across five U.S. Department of Energy FACE sites. At the NC pine FACE sites, we determined seasonal patterns of diversity and composition for the total fungal community (LSU gene) and the cellulolytic fungal guild (*cbhI* gene), to determine how seasonal dynamics may affect community responses to elevated $CO_2$ and nitrogen (N)-fertilization at this site. The expression patterns of *cbhI* genes in soils from different depths and treatments are also being characterized.

To investigate the impacts of ecosystem type and a decade of elevated atmospheric $CO_2$ treatments on cellulolytic fungal communities, we sequenced 10,677 *cbhI* gene fragments across five distinct terrestrial ecosystem experiments. The composition of fungal *cbhI* genes was unique in each ecosystem. Using a 114-member *cbhI* sequence database compiled from known fungi, less than 1% of the environmental sequences could be classified at the family level indicating that cellulolytic fungi *in situ* are either dominated by novel fungi or by known fungi that are not yet recognized as cellulose degraders. Shifts in fungal *cbhI* composition and richness in response to elevated $CO_2$ exposure varied among the ecosystems. In aspen forest and desert creosote bush soils, *cbhI* gene richness was significantly higher after exposure to elevated $CO_2$ than under ambient $CO_2$. In contrast, richness was not altered in desert soil crusts, but the relative abundance of dominant operational taxonomic units (OTUs) was significantly shifted. Collectively, the data show that responses are complex.

Seasonal surveys of *cbhI* and the fungal LSU gene across treatment and control plots at the loblolly pine FACE site indicated that there is a trend of decreased fungal richness in elevated $CO_2$ plots compared to ambient plots across all seasons. Nitrogen fertilization in both ambient and elevated $CO_2$ plots increased richness in spring and summer sampling points indicating that some taxa are nitrogen limited during these seasons. In contrast, there did not appear to be any seasonal change in the cellulolytic fungal guild. These studies enable us to understand how much of a control seasonal fungal community dynamics exert on responses to climate change parameters.

Gene expression studies at the NC pine FACE site demonstrated that the expressed *cbhI* genes are dominated by Basidiomycota-like sequences that are not closely related to any of the available *cbhI* gene sequences from presently known fungi. In addition, the rank abundance of *cbhI* genes from DNA-based versus RNA-based clone libraries show that the most abundant *cbhI* genes are not always the most highly expressed. The DNA and RNA-based surveys provide complimentary information, revealing the community potentially able to carrying out a specific function and the populations that are most active under particular environmental conditions.

Collectively, these studies show that ecosystem type and seasonal dynamics play important roles in controlling responses to climate change parameters.

# 182

## Response of Soil Microbial Communities in an Arid Ecosystem to Eight Years of Elevated $CO_2$

**Cheryl R. Kuske**[1]* (Kuske@lanl.gov),
**Co-authors:** Blaire T. Steven,[1] La Verne Gallegos-Graves,[1] Jayne Belnap,[2] R. David Evans,[3] Shawn R. Starkenburg,[1] and Patrick S. Chain[1]

[1]Los Alamos National Laboratory; [2]United States Geological Survey; and [3]Washington State University

**Project Goals: The goal of this project is to describe the changes in community structure and ecosystem functioning expected to occur in response to elevated concentrations of atmospheric $CO_2$, in a manipulated and highly replicated field experiment. Our role in the larger context of the project is to characterize changes in the soil microbial communities.**

Arid and semi-arid ecosystems, which cover ~40% of the terrestrial surface are predicted to be especially sensitive to climate change. We have built on previous studies conducted in the Nevada Desert FACE Facility (NDFF), a large scale, replicated Free Air $CO_2$ Enrichment (FACE) experiment to describe changes in an arid ecosystem associated with elevated atmospheric $CO_2$, a primary driver of climate change. The ecosystem is shrub-dominated with large surfaces lacking plant cover. Biological soil crusts exist in the inter-plant spaces and cover from 35 to 65% of the surface area. Soil crusts consist of cyanobacteria, algae, lichens and associated heterotrophic bacteria and fungi. Previous studies have identified several ecosystem responses to elevated $CO_2$. The plant populations at the NDFF responded to elevated $CO_2$ through increased biomass and changes in the commu-

nity structure, while the soil crusts responded to increased $CO_2$ through decreases in moss and lichen cover.

We have been employing a suite of molecular techniques, ranging from targeted gene sequencing to shotgun metagenomics, to characterize the soil microbial communities in the NTS-FACE site. Our preliminary studies have demonstrated that the bacterial populations are patchily distributed with only minimal species overlap between replicate samples. However, there was clear differentiation of the soil microbial populations by habitat. The bacterial populations associated with the creosote bush root zones were significantly different than those in the inter-space biologic soil crusts. Populations of Cyanobacteria, Actinobacteria and Proteobacteria dominated soil crusts whereas creosote root zone bacterial populations were dominated by Actinobacteria and to a lesser extent Proteobacteria. The differences in these communities also translated into differences in the response of these communities to elevated $CO_2$. However, the detected differences in response to elevated $CO_2$ were not consistent between the different metagenome approaches. The phyla found to respond to the $CO_2$ treatment were generally consistent between methods but the direction of their response differed. For example, 16S rRNA sequencing on Sanger or pyrotag platforms suggested a decline in the relative abundance of Cyanobacteria in response to elevated $CO_2$. In contrast, analysis of shotgun metagenomes using the SEED tool of MG-RAST suggested that Cyanobacteria were enriched in response to elevated $CO_2$. This disparity may be due to the presence of cyanobacteria in the soil crusts that are only distantly related to currently described species. The lack of reference genomes for these organisms renders them effectively invisible to metagenome taxonomy assignments and results in a reduction of Cyanobacteria calls compared to 16S rRNA sequencing. Nevertheless, these data suggest particular bacterial phyla were sensitive to elevated $CO_2$. We are now pursuing highly replicated field sampling coupled with pyrotag and shotgun metagenome sequencing to better define the fine scale changes in bacterial and fungal communities in response to elevated $CO_2$.

submitted post-press

# 182A‡

## Bacterial Cheaters and Investors: Isolation of Non-Polymer Degraders and Extracellular Enzyme Producers from the Environment

Yasawantha D. Hiripitiyage* (yhiripit@kent.edu), Shifang Hsu, Ashley Golphin, Laura G. Leff, and Christopher B. Blackwood

Department of Biological Sciences, Kent State University, Kent, Ohio

**Project Goals: During the process of decomposition, extracellular enzymes catalyze breakdown of polymeric compounds into smaller molecules that can be taken up by cells. Little is known about the influence of "cheating" bacteria that do not secrete extracellular enzymes but use monomers resulting from the activity of extracellular enzymes secreted by other organisms ("investors"). The goal of this study was to isolate and identify cheating and investing bacteria from environmental samples.**

We hypothesized that higher levels of fungal biomass would result in greater extracellular enzyme activity and consequently, a greater population of cheating bacteria. We also hypothesized that increased moisture content would increase diffusion of extracellular enzymes and monomers from investors to cheaters, and therefore would favor cheating bacteria. Samples were collected from Jennings Woods, an experimental forest in Northeast Ohio, USA. We sampled decayed leaves from three habitats in and around vernal pools comprising a complete factorial design of two treatments: high or low fungal biomass based on the observation of fungal hyphae, and high or low moisture. Samples were used to inoculate bi-layer gradient agar plates where the lower layer contained monomers and the upper layer contained polymers as the carbon source. After inoculation on the bi-layer gradient agar plates, the numbers of bacterial colonies growing on each plate were counted. Results showed that colony counts were higher on the side of the plate containing the higher concentration of monomers ($P<0.01$). However, differences among treatments (high or low fungal biomass and high or low moisture) were not statistically significant ($P>0.05$). Screening was carried out for the growth of the isolates on the polymer cellulose and its monomer cellobiose and results showed that 40 percent of the total isolates did not grow on cellulose while all of the isolates grew on cellobiose. Fungal biomass and moisture treatments did not differ significantly in the proportion of isolates that could grow on cellulose ($P>0.05$). Non-polymer degrading bacteria were isolated and appear to be cheaters, however, there was no detectable difference among treatments.

# Systems Biology Strategies and Technologies for Understanding Microbes, Plants, and Communities

## Analytical Strategies for the Study of Plants, Microbes, and Microbial Communities

# 183

### Proteomics and Pan-omics Driven Analysis of Microbial Communities, Comparative Biology and Environmental Symbiosis

Mary S. Lipton[1]* (mary.lipton@pnl.gov), Stephen J. Callister,[1] Kristin E. Burnum,[1] Roslyn N. Brown,[1] Kim K. Hixson,[2] Janani Shutthanandan,[1] Samuel O. Purvine,[2] Angela D. Norbeck,[1] Matthew E. Monroe,[1] Carrie D. Nicora,[1] Gordon A. Anderson,[1] and **Richard D. Smith**[1] (PI)

**Collaborators:** Steve Giovannoni,[3] Cameron Currie,[4] Michael Kahn,[5] Donald Bryant,[6] Norman Lewis,[5] and David Kramer[7]

[1]Biological Sciences Division; Pacific Northwest National Laboratory; [2]Environmental Molecular Sciences Laboratory; Pacific Northwest National Laboratory; [3]Oregon State University; [4]University of Wisconsin; [5]Washington State University; [6]Pennsylvania State University; and [7]Michigan State University

**Project Goals: This project employs comprehensive global and directed pan-omics analyses and novel informatics approaches (developed in parallel in this program) of microbes, plants and microbial communities to enhance scientific understanding through elucidation of phenotypic relationships between environmentally important microorganisms, characterization of higher organisms, characterization of the metabolic activities within microbial communities, and identification of post-translationally modified proteins.**

Inherent to exploiting microbial function for carbon cycling, bioremediation or biofuel production or utilizing plants as energy precursors is the detailed understanding of the physiology of the cell. Cellular functions are dictated by the proteins expressed in the cell, their resident lifetime in the cell, their localization and their modification state. Additionally, these processes mitigate the metabolites in the cell that serve as energy and carbon currency. This project exploits the technological and informatics advances of the pan-omics measurement pipeline at PNNL (as described in the poster by Anderson et al and Metz et al) to address organism-specific scientific objectives developed in conjunction with biological experts for a number of different microbes and plants. In our poster, we highlight the ability to use pan-omics data for, characterization of microbial communities, elucidation of phenotypic and genotyptic relationships, advances in the characterization of protein modification state, and the determination of protein localization in stem, root and leaf tissues of *Arabidopsis*.

Microbes do not live in isolation; therefore, understanding the function of a microbe in the environment and the effect of the environment upon the microbe requires the characterization of the community as a whole. Research on individual microbes takes on a larger significance if the findings about an organism in cell culture can be extrapolated to the activities of the organism within the natural community in the environment. SAR11, also known as *Pelagibacter ubique*, is the dominant heterotrophic bacterial clade in the oceans, where roughly 25% of the 16S rRNA gene sequences retrieved from uncultured marine bacteria belong to the SAR11 group. Evolutionary selection to minimize genome size in large, nutrient-limited ocean populations, known as genome streamlining, has been implicated as an important factor in the evolution of SAR11. These cells have dispensed with many pathways and transporters that are typically present in bacteria with more complex genomes. We have investigated how *Pelagibacter* respond to iron limitation by applying differential measurements using our new pan-omics platform to *Pelagibacter* cell cultures.

The fungus-growing ant– microbe symbiosis is a paradigmatic example of organic complexity generated through symbiotic association and, over the last decade, it has become a model system for studying symbiosis. We have demonstrated an in-depth profiling of the fungal garden complete with bacteria (fungus alone, isolated bacteria, and fungal garden intact) to understand the relationship between the fungus and the bacterial protectors. Proteomics and metabolomic studies of the secreted proteins from the bacteria have been characterized in an effort to understand the relationship between the ants and the fungus. These studies demonstrate, the ability to use pan-omics measurements on an ecosystem level, spanning bacteria to multi-cellular organisms.

The genotype of an organism is the full hereditary genome, while the phenotype is the actual observed biochemical characteristics of an organism. Although the genome of an organism influences its phenotype, phylogenetically diverse organisms can share a common phenotype. As such, genome-based comparisons are limited in describing these common mechanisms in diverse organisms. We have developed a number of proteomics databases for each of the six bacterial phyla known to contain chlorophyll-based phototrophs, including the recently discovered Candidatus *Chloroacidobacteria thermophilum*, which is currently the only

known phototroph within the phylum *Acidobacteria*. Using these databases, to investigate diverse bacteria that share a similar photosynthetic phenotype while having vastly different genotypes. With a better understanding of the photosynthetic pathways, and especially the pathways occurring within cyanobacteria and chloroplasts, systems biology approaches will be poised to determine how these organisms can be used to create alternative fuels, as well as their role in the carbon cycle.

Symbiosis is the long-term interaction between different biological species. One sort of symbiosis is nitrogen fixation occurring in specialized symbiotic interactions between plants and bacteria is a major source of useful nitrogen. We have used Pan-omics measurements to elucidate the interaction between *Medicago truncatula* and *Sinorhizobium meliloti*, mapping proteins expressed in both the plant and the symbiont with preliminary understanding of their interaction.

Additional information and supplementary material can be found at the PNNL proteomics website at http://oberproteomics.pnl.gov/

Acknowledgements: This research is supported by the Office of Biological and Environmental Research of the U.S. Department of Energy. Portions of this research were performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the DOE's Office of Biological and Environmental Research Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute through Contract No. DE-AC05-76RLO 1830.

# 184

## Proteomics and Pan-omics Measurements for Comprehensive Systems Characterization of Biological Systems

**Thomas O. Metz*** (thomas.metz@pnl.gov), Scott R. Kronewitter, Qibin Zhang, Aaron T. Wright, Ronald J. Moore, Tao Liu, Weijun Qian, Erin S. Baker, Therese R.W. Clauss, Karl K. Weitz, Carrie D. Nicora, Heather M. Brewer, Daniel J. Orton, Anil K. Shukla, Rui Zhao, Feng Yang, Roslyn Brown, **Joshua N. Adkins, Gordon A. Anderson, Mary S. Lipton,** and **Richard D. Smith (PI)**

Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Wash.

http://www.pnl.gov/

**Project Goals: This project endeavors to implement and apply advanced capabilities that aim at comprehensive molecular characterization of biological systems, including the extension of proteomics to cover cover post-translational protein modifications and the implementation of broad metabolomics, lipidomic and glycomic measurements. Together with more widely available genomics and transcriptomics capabilities, this project will provide the transformative "pan-omics" measurement capabilities needed to elucidate interacting networks of genes, proteins, and biochemical reactions in biological systems.**

The goal of BER's Genome Science Program (GSP) is to achieve a predictive systems level understanding of plants, microbes and biological communities via the integration of fundamental science and technology developments to enable biological solutions to challenges in energy, environment and climate. Achieving this goal requires comprehensive proteomics, metabolomics, lipidomics, and glycomics, i.e. pan-omics, measurement capabilities and the integration of data generated by these approaches. This project aims to facilitate understandings of biological systems by providing pan-omics molecular measurement capabilities that will be applied in biology-driven collaborative projects led by investigators actively engaged in developing systems biology approaches in support of BER's research agenda. Our strategy benefits from advances in high resolution nano-liquid chromatography (LC) separations combined with high mass accuracy mass spectrometry (MS) measurements and other developments that afford large gains in performance and throughput. These efforts also include the automation of key steps in proteomics sample processing; fractionation of protein samples based on surface membrane protein enrichment and subcellular fractionation methods using differential gradient centrifugation; and implementation of novel methods for protein extraction from environmental (e.g. soil) samples. Additional advancements involve the implementation of targeted proteomics methods (e.g. activity-based protein profiling and multiple reaction monitoring) and approaches for the elucidation of protein isoforms (e.g. integrated top-down and bottom-up proteomics) and post-translational modifications (e.g. phosphoproteomics and characterization of protein glycosites).

To facilitate these goals, this project includes efforts to develop and apply new measurement platforms and integrated analytical strategies implemented in concert with the computational advances necessary for handling increased data production rates, improved data processing algorithms, the development of methods to integrate multiple pan-omics data streams, and efforts needed to effectively disseminate results and information to collaborators and the broader scientific community. Developments are driven by and applied in the context of external collaborative projects aimed at garnering the knowledge needed to lay a foundation for predicting behaviors of and manipulating biological systems critical to DOE missions.

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.

# 185

## Pan-omics Measurements Platform and Informatics Analysis Pipeline

**Gordon A. Anderson*** (gordo@pnl.gov), Ronald J. Moore, **Joshua N. Adkins,** David J. Anderson, Kenneth J. Auberry, Mikhail E. Belov, Kevin Crowell, Stephen J. Callister, Therese R.W. Clauss, Kim K. Hixson, Gary R. Kiebel, Brian L. LaMarche, **Mary S. Lipton,** Da Meng, **Thomas O. Metz,** Matthew E. Monroe, Heather M. Brewer, Carrie D. Nicora, Angela D. Norbeck, Daniel Lopez-Ferrer, Daniel J. Orton, Ljiljana Paša-Tolić, David C. Prior, Samuel O. Purvine, John D. Sandoval, Anuj Shah, Yufeng Shen, Anil K. Shukla, Mudita Singhal, Gordon W. Slysz, Aleksey V. Tolmachev, Nikola Tolić, Karl Weitz, Aaron Wright, Rui Zhang, Rui Zhao, and **Richard D. Smith** (rds@pnl.gov)

Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Wash.

This project is developing pan-omics measurement and integrative informatics capabilities to enable comprehensive global molecular characterization to understand, model, and potentially manipulate biological systems. These new analytical capabilities are achieved through the application of advanced separations-MS measurement platforms that greatly increase measurement quality and throughput. This new platform combines fast, multidimensional separations (i.e., fast LC in conjunction with millisecond-scale ion mobility separations) with ultra-fast and accurate mass measurement time-of-flight MS, and provides greatly expanded proteome coverage and greater sensitivity, in addition to at least an order-of-magnitude increase in throughput. Pan-omics measurement capabilities are based on essentially identical separations-MS measurement platforms and similar data processing/informatics pipelines; metabolomics, lipidomics, and glycomics measurements, as well as the expanded proteomics measurements. A key element of pan-omics analysis is informatics methodologies to integrate data from various measurements and incorporate approaches for managing and communicating data, data quality, and ambiguities (e.g., the confidence in peptide and protein identifications, modification sites, abundance levels, etc.).

Advanced measurement, informatics, and computational technologies and approaches are being explored and evaluated for possible broader implementation based on their robustness and suitability for implementation in high-throughput pan-omics, their ability to improve data quality, and their potential to facilitate new biological insights from collaborative applications. For example, we have further developed and applied new IMS-TOF MS measurement platforms that greatly extend our current measurement capabilities by providing data production rates an order of magnitude greater than current (e.g., Orbitrap or FTICR MS-based) platforms in addition to significantly enhancing data quality. Application of the new platforms in conjunction with integrated analytical strategies and increasing automation will provide the throughput needed to more routinely and more extensively cover the range of post-translationally modified proteins, as well as other pan-omics measurements.

Development and application of new measurement capabilities and computational tools are essential for generating, processing, integrating, and disseminating data and information from GSP studies of responses over multiple scales that provide a foundation for manipulating biological systems. We are leveraging the extensive experience and capabilities developed to date within the high throughput proteomics facility at PNNL to extend this capability to multiple omics measurements and provide a framework for effective data integration. Also crucial is the ability to manage, integrate, disseminate, and extract information from increasingly large and complex datasets. The measurement advances noted above require corresponding computational and informatics advances necessary for: managing the resulting increased data production rates; evaluating and controlling data quality; processing and integrating data from the various analysis streams; and disseminating data and information to collaborators, users, and the broader scientific community. Thus, we are developing a suite of data analysis tools, data consolidation applications, and statistical packages, as well as visualization software for data interpretation, and that will support integration of the enhanced proteomics and metabolomics data sets. This framework further supports integration of genomics data from public repositories and aim to provide the needed infrastructure to interoperate with the GTL Knowledgebase.

This poster highlights several developments that enable pan-omics measurements. These developments include; 1) Advances in measurement capabilities, 2) Data management and enhanced informatics analysis capability, and 3) Initial developments of an integrative analysis framework.

This research is supported by the Office of Biological and Environmental Research of the U.S. Department of Energy. Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute through Contract No. DE-AC05-76RLO 1830.

# 186

## Informatics Infrastructure to Enable Pan-omics Measurements of Biological Systems

Matthew E. Monroe* (matthew.monroe@pnl.gov), **Joshua N. Adkins, Gordon A. Anderson,** Kenneth J. Auberry, Kevin Crowell, David A. Clark, Gary R. Kiebel, **Mary S. Lipton, Thomas O. Metz,** Ronald J. Moore, Angela D. Norbeck, Daniel J. Orton, Samuel H. Payne, Samuel O. Purvine, John D. Sandoval, Anuj R. Shah, Gordon W. Slysz, and **Richard D. Smith** (rds@pnl.gov)

Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Wash.

**Overview**: We have established a robust, flexible computational architecture and infrastructure to manage the storage and tracking of raw and processed data associated with pan-omics research. This architecture additionally provides a framework to support data integration that will enable the creation of comprehensive pan-omics measurement datasets. The infrastructure has a web-based interface for accessing and updating information and has mechanisms for exporting and processing the data associated with the identified peptides, proteins, metabolites, etc. The informatics infrastructure will continue to evolve and expand to support the advanced measurement platforms and analysis capabilities needed to enable pan-omics studies.

The Pan-Omics Research Information Storage and Management System (PRISM) provides a flexible and robust infrastructure that serves as the foundation for developing data integration workflows for pan-omics data. The existing informatics infrastructure and analysis tools provide the scalable architecture and base capabilities that allow rapid development to enable integration of pan-omics data. The PRISM infrastructure has evolved and expanded since its inception, and its current design allows for continued expansion, including supporting the increased production rates afforded by a new LC-ion mobility mass spectrometer platform. The PRISM architecture employs a "plug-and-play" paradigm, where individual steps of the informatics pipeline are developed as configurable, cohesive, and independent modules that can be readily chained together in multiple ways to create effectively new informatics pipelines. PRISM incorporates community-developed analysis software, commercial software, and a variety of in-house developed tools for peptide ID, deisotoping, quantitation, and data handling. We anticipate adding additional analysis tools developed for new informatics analysis pipelines related to lipidomics, activity-based proteomics, phosphoproteomics, top-down proteomics, metabolomics, etc. The modular nature of PRISM allows us to offload computationally intensive processing tasks to high-performance computing or cloud computing resources that are becoming available to the scientific community (e.g., the Magellan project, Amazon, or the planned GTL Knowledgebase).

The PRISM infrastructure supports a wide array of functions, including tracking research projects and their associated biological samples, managing the storage and tracking of raw and processed data files, and automated software processing of pan-omics data. As research projects become larger and more diverse, we can further expand LIMS-type capabilities supported by PRISM (e.g., tracking instrument operation and maintenance details). PRISM enables facility staff to better plan and define the sample processing and analysis strategy, including the ability to specify sample run batching and blocking parameters, and to annotate samples with processing factors for use in later data analyses.

PRISM provides several interfaces for accessing and exporting both the raw and processed data. The primary portal for interfacing with PRISM is the DMS website, which allows researchers to browse and search existing information, add new information, and export data. PRISM also includes programmatic interfaces to allow batch export of data using standalone software. The Multi Dataset Analysis and Rollup Tool (MDART) provides a mechanism for collaboration, standardization, and scientific documentation of processed pan-omics data. This tool interfaces with PRISM to allow researchers to export, process, and 'filter' data, and provides the flexibility essential for dealing with a variety of data types, application interests, and data analysis needs. MDART uses workflows to define a systematic and repeatable, yet flexible approach to processing data.

The Mass and Time Tag System (MTS) component of PRISM is responsible for collating peptide search results to form accurate mass and time (AMT) tag databases that can be used for high throughput quantitative studies. MTS is federated across a compute cluster, allowing for ready expansion to support the increasing volumes of data that will be generated in pan-omics studies, including supporting the new measurement platform. VIPER and MultiAlign are used to characterize detected LC-MS or LC-IMS-MS features and match those features to the AMT tag databases. These tools now use the Statistical Tools for AMT tag Confidence (STAC) algorithm to assign confidence values to peptides identified via the peak matching process, thus allowing researchers to filter the results to obtain a specified false discovery rate (FDR).

The PRISM system provides a flexible and robust infrastructure that serves as the foundation for developing data integration workflows for pan-omics data. Workflows are a set of connected operations similar to the work of a researcher (e.g., the integration of multiple time points into a time course dataset). As pan-omics measurements and infrastructure expand, we expect to incorporate new workflows to support pan-omics data integration.

This poster will illustrate the current PRISM capability and developments in progress to enable pan-omics data integration from multiple omics analysis workflows as well as to deal with the significant increase in data volumes generated by the new LC-ion mobility measurement platform.

This research is supported by the Office of Biological and Environmental Research of the U.S. Department of Energy. Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute through Contract No. DE-AC05-76RLO 1830.

# 187

## Fidelity and Dynamics of DNA Methylation in Plants

Qin Yao,[1] Changjun Liu,[1] John Shanklin,[1] Chuan He,[2] and **John Dunn**[1]* (jdunn@bnl.gov)

[1]Department of Biology, Brookhaven National Laboratory, Upton, N.Y.; and [2]Department of Chemistry and Institute for Biophysical Dynamics, University of Chicago, Ill.

**Project Goals: The first goal of this project is to develop defined in vitro systems for determining the mechanisms by which cytosine DNA methylation is normally maintained in plants and how factors leading to DNA oxidative damage impact the fidelity of DNA methylation. A second goal is to develop, demonstrate, and validate a streamlined "DNA target-enrichment" method coupled to bisulfite sequencing for in-depth methylation mapping of specific plant gene sets and their associated control elements.**

Epigenetics is defined as the study of heritable changes to genome structure and function that do not change the nucleotide sequence of the DNA. Methylation of cytosine to form 5-mC in genomic DNA is an important epigenetic marker that plays a critical role in regulation of gene expression, chromatin structure and genome stability. In all organisms, cytosine methylation is a postreplicative process. Newly synthesized DNA strands are unmethylated, thus creating hemi-methylated duplexes at replication forks. In both mammals and plants most methylation occurs at the DNA dinucleotide CpG, where both cytosines in the complementary strands of adjacent base pairs are methylated. In mammals the UHRF1 protein recognizes these hemi-methylated CpG sites via its SET-and RING-associated (SRA) domain. Structural studies have shown that the 5-mC residue in hemimethylated DNA bound to UHRF1 is flipped outside of the DNA helix into a specific 5-mC-binding pocket within the SRA domain. This causes DNA looping and allows the N-terminal region of the SRA domain to interact with the DNA's major and minor grooves. The residue requiring modification is then flipped out of the helix and presented to the DNA methyl transferase DNMT1 for addition of the methyl group.

In *Arabidopsis* VIM1 encodes an SRA domain methylcytosine-binding protein that probably functions similarly to UHRF1 in playing a major role in maintaining DNA methylation patterns following DNA replication. To gain further insight into how VIM1 functions, we have cloned and expressed and purified VIM1 and are using electrophoretic mobility-shift and fluorescence anisotropy titration assays to study its interaction with model duplex DNAs containing cytosine or 5-mC in one or both strands.

A fine-scale mapping tool "Bisulfite Patch PCR" [1] is also being used to discover the potential epigenetic regulation underlying the spatial and temporal expression of 26 genes/transcription factors involved in lignin biosynthesis in *Arabidopsis*. This new approach allows us to process ~100 genes from multiple samples at the same time.

Recent studies of genomic DNA from human brain, neurons and mouse embryonic stem cells have demonstrated that these DNAs contain a sixth base, oxidized 5-mC or 5- hydroxymethylcytosine (5-HmC). Current thinking is that 5-HmC does not merely mimic 5-mC groups but likely plays an independent role in yet unknown epigenetic regulation of various biological processes. We are using bacteriophage T4 β-glucosyltransferase to transfer a glucose moiety containing an azide group onto the hydroxyl group of 5-HmC. The azide group can then be chemically modified with biotin for detection, affinity enrichment and subsequent sequencing of 5-HmC–containing DNA fragments to reveal the genomic locations of 5-HmC in the DNA[2]. Dot blot detection of biotinylated glucose-HmC shows about 0.06% 5-Hmc in plant leaf genomic DNA. Efforts are underway to map these sites. We are also determining if 5-HmC effects VIM1 binding and if 5-HmC residues can be removed by the *Arabidopsis* DNA glycosylase/lyase, Repressor of Silencing, ROS1, thereby allowing the 5-HmC generated by oxidation of 5-mC, to be replaced by C, resulting in active demethylation.

### References

1. Varley, K.E. and Mitra, R.D. (2010) Bisulfite Patch PCR enables multiplexed sequencing of promoter methylation across cancer samples. *Genome Res*, **20**, 1279-1287.
2. Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., Zhang, W., Jian, X. et al. (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol, 29, 68-72.

Early Career Program
Speaking Wednesday 10:30 a.m.

# 188

## A Systems Biology, Whole-Genome Association Analysis of the Molecular Regulation of Biomass Growth and Composition in *Populus deltoides*

**Matias Kirst*** (mkirst@ufl.edu)

School of Forest Resources and Conservation, University of Florida Genetics Institute, Gainesville

**Project Goals: This project main goal is to apply an association genetics approach to unveil the molecular basis of biomass productivity and composition. To comprehensively capture the genetic variants that regulate traits of value for bioenergy production, we are combining sequence-capture and high-throughput sequencing to genotype coding and regulatory sequences in the whole-genome of *P. deltoides*. To achieve this goal we have: (1) optimized sequence-capture for unbiased, high-throughput and low-cost recovery of target coding and regulatory sequences in *P. deltoides*. A set of over 220,000 probes that efficiently capture exon and 500 bp of putative regulatory sequences of 24,000 genes have been**

**developed so far. Next we are (2) genotyping a *P. deltoides* unstructured population for association mapping. Oligonucleotides optimized for recovery of target coding and regulatory sequences are being used for sequence capture in 500 individuals of an association population. Captured fragments will be resequenced and polymorphisms genotyped for association analysis. (3) Upon completion of genotyping, we will identify significant SNP-trait associations with biomass growth and carbon partitioning to define genes and alleles that regulate trait variation. Alternative alleles detected in polymorphic sites will be tested individually and in a combined model for marker-trait association to identify the genes that regulate biomass growth and partitioning of carbon into lignocellulosics.**

Poplars are the principal short rotation woody crop species for providing clean, renewable and sustainable fuels in North America, because of their fast, perennial growth habit and wide natural distribution in a broad range of environments. While poplars provide the benefits of an ideal bioenergy crop, with few exceptions, the genes that regulate productivity and biomass composition are largely unknown, despite their critical relevance for efficient conversion of biomass to biofuels. This gap is the main barrier for efficient molecular breeding and selection of superior poplar germplasm and, consequently, the extensive adoption of this woody crop as a renewable bioenergy source. Association genetics has become the primary approach for identification of genes that regulate complex traits in human genetics, agriculture and forestry because this strategy captures information on a broad range of alleles that control phenotypic variation, at high-resolution. Poplars are particularly suited to unveil the molecular basis of biomass productivity and composition using association genetics because of minimal domestication, large open-pollinated native populations with limited genetic structure, and high levels of genetic and phenotypic variation. However, with few exceptions, association genetic studies in plants have been hampered by limited gene and polymorphism coverage, because of the limited knowledge of the genetic variants and the low multiplexing capacity of genotyping platforms available to plant species. As a consequence, for most traits analyzed to date only a limited fraction of the genetic diversity impacting phenotypic variation has been uncovered.

# 189

## Comparative Gene Expression of the *Caldicellulosiruptor* Genus using RNAseq

Loren J. Hauser[1]* (hauserlj@ornl.gov), Sara Blumer-Schuette,[2] Ira Kataeva,[3] Sung-Jae Yang,[3] Farris Poole,[3] Daniel Quest,[1] Inci Ozdemir,[2] Andrew Frock,[2] Erika Lindquist,[4] Tanya Woyke,[4] Bob Cottingham,[1] **Michael W.W. Adams,[3]** and **Robert M. Kelly[2]**

[1]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]North Carolina State University, Raleigh; [3]University of Georgia, Athens; and [4]Joint Genome Institute, Walnut Creek, Calif.

**Project Goals: This project has two goals. The first is to compare the gene expression profiles, using RNAseq, of a series of related high growth temperature bacteria from the genus *Caldicellulosiruptor* grown using both simple and complex carbon sources. The second is to develop a set of analysis tools to process large amounts of RNAseq data.**

All known members of the *Caldicellulosiruptor* genus grow optimally between 65°C to 80°C and can anaerobically degrade plant biomass using various and complementary strategies. They are prime candidates for use in an industrial consolidated bioprocessing facility to produce second generation biofuels from complex plant material such as switchgrass. In collaboration with the Department of Energy Joint Genome Institute (JGI) we have recently completed sequencing and annotating the genomes of eight members of this genus. In addition, we have generated RNAseq data from four members grown on a variety of carbon sources including, glucose, maltose, cellobiose, starch, crystalline cellulose (Avicel), and dilute acid pre-treated switchgrass. Two of the primary advantages of RNAseq are its dynamic range and sensitivity. Greater than 98.5% of all protein coding genes had some detectable expression in all growth states and varied in expression level up to $10^6$ fold. The expression levels of some genes, when grown on different carbon sources, varied by over $10^3$ fold. As expected, the genes encoding ABC sugar transporters, cellulases and other glycosyl hydrolases were amongst the genes with the greatest changes in expression levels when grown on sugars versus complex carbon sources such as switchgrass. However, there were a number of other genes, such as members of a CRISPR cluster and some genes involved in fatty acid metabolism, that had unexpected changes in expression when grown on different carbon sources. We are developing an analysis pipeline to process and visualize the data and will also compare them with the results from DNA microarray analyses. RNAseq analyses will also include identifying the 5' end of transcription units, defining operons, identifying co-regulated genes and operons, and predicting transcription factor binding sites. Preliminary analysis has identified putative promoters embedded in genes, which allows the definition of unconventional operons and regulons. A thorough analysis will undoubtedly reveal additional unique biological phenomenon.

## Biological Systems Interactions

# 190

## PNNL Foundational Scientific Focus Area—Biological Systems Interactions

**Jim Fredrickson**[1]* (jim.fredrickson@pnl.gov) and **Margie Romine**[1]
**Co-Principal Investigators:** Gordon Anderson,[1] Scott Baker,[1] Alex Beliaev,[1] Bill Cannon,[1] Mary Lipton,[1] Jon Magnuson,[1] Thomas Squier,[1] and H. Steven Wiley[1]
**Laboratory Research Manager:** Harvey Bolton Jr.[1]
**External Collaborators:** Don Bryant,[2] Frank Collart,[3] William Inskeep,[4] Francois Lutzoni,[5] Andrei Osterman,[6] Margrethe Serres,[7] and David Ward[4]
**PNNL Contributors:** Alice Dohnalkova,[1] David Kennedy,[1] Bryan Linggi,[1] and Steve Lindemann[1]

[1]Pacific Northwest National Laboratory, Richland, Wash.; [2]Pennsylvania State University; [3]Argonne National Laboratory; [4]Montana State University; [5]Duke University; [6]Burnham Institute for Medical Research; and [7]Marine Biological Laboratory

**Project Goals: The main scientific objectives of the PNNL FSFA include: development of a mechanistic understanding of interactions among key members of microbial autotroph-heterotroph associations (AHA) using the tools of genomics and systems biology; understanding the collective energy, carbon, and nutrient processing in AHAs that contributes to their stability and efficient utilization of resources; probing interspecies co-adaptations and functional innovations that contribute to robustness and functional efficiency and exploring the types of microbe-microbe and microbe-environment interactions that control genome evolution; and understanding cellular strategies that permit a system of interacting organisms to control the excess generation of reactive oxygen species to promote adaptive responses that enhance their survival.**

The PNNL Genomic Science Foundational Scientific Focus Area (FSFA) is addressing critical scientific issues on microbial interactions, investigating how microorganisms interact to carry out, in a coordinated manner, complex biogeochemical processes such as the capture and transfer of light and chemical energy. The primary research emphasis is on associations between autotrophic and heterotrophic microorganisms with the additional objective of obtaining a predictive understanding of how interactions impart stability and resistance to stress, environmental fitness, and functional efficiency. The main scientific objectives of the FSFA include: development of a mechanistic understanding of interactions among key members of microbial autotroph-heterotroph associations (AHA) using the tools of genomics and systems biology; understanding the collective energy, carbon, and nutrient processing in AHAs that contributes to their stability and efficient utilization of resources; probing interspecies co-adaptations and functional innovations that contribute to robustness and functional efficiency and exploring the types of microbe-microbe and microbe-environment interactions that control genome evolution; and understanding cellular strategies that permit a system of interacting organisms to control the excess generation of reactive oxygen species to promote adaptive responses that enhance their survival (see PNNL FSFA posters for additional detail). Biological systems under investigation range from co-cultures consisting of well-characterized model organisms, to highly evolved lichen-forming microeukaryotes, to microbial mats associated with geothermal and hypersaline environments.

Microbial mats are highly organized, metabolically interactive, self-sustaining communities that develop in extreme environments. Hot Lake is a hypersaline, epsomitic ($MgSO_4$), shallow meromictic lake near Oroville, Washington that contains a benthic microbial mat (Figure 1A). Because of the high alkalinity and divalent cation concentrations of saline alkaline lakes, carbonate minerals can precipitate in these environments, often in association with microbial mats. These lakes therefore represent natural models for investigating carbon cycling and the microbial processes that may accelerate carbonate mineralization. The measured pH of Hot Lake varies from 8.5-9.2 and the Mg and $SO_4^-$ concentrations can reach 1.7 and 2.2 molar, respectively, depending on season and depth. The top layer of the submerged mat (Figure 1B), immediately under the large crystals and on top of the green cyanobacterial layer contains magnesite ($MgCO_3$) crystallites that are encased in an amorphous layer of similar composition (i.e., Mg, C, and some Ca) (Figure 2). We hypothesize that organic polymers (e.g., extracellular polymeric substances) produced by the mat may be coating the particles, impeding crystallization at the surface.

Exploratory investigations of microbial diversity were conducted using pyrosequencing and revealed many taxonomic units consistent with phototrophy, chemoheterotrophy, sulfur cycling, and halotolerance. The most abundant phototrophs, based on nearest-neighbor phylogeny, included the filamentous cyanobacterium *Tychonema* and the phototrophic sulfur bacteria, *Halochromatium*. *Thiohalocapsula*, a purple sulfur bacterium, was also present as were purple non-sulfur bacteria affiliated with *Roseovarius* and *Rhodobaca*. Pyrosequencing also detectedmany phyla representative of halotolerant, sulfur-respiring bacteria including *Halothiobacillus* and sulfate-reducing bacteria including *Desulfonatronum* and *Desulfotignum*. Laboratory microcosms have been established from mat subsamples diluted into lake water to provide first-generation model systems where community structure and function responses to environmental variations can be measured. Future research plans include flow cytometry and sectioning in combination with metagenomics and metatranscriptomics to further characterize the functions and interactions among members of this mat community as a function of vertical position. In addition to systems biology investigations, we will also be investigating geomicrobial processes contributing to carbonate mineral formation.

Figure 1. Hot Lake benthic microbial mat (top, mat underside) and vertical section illustrating layering.

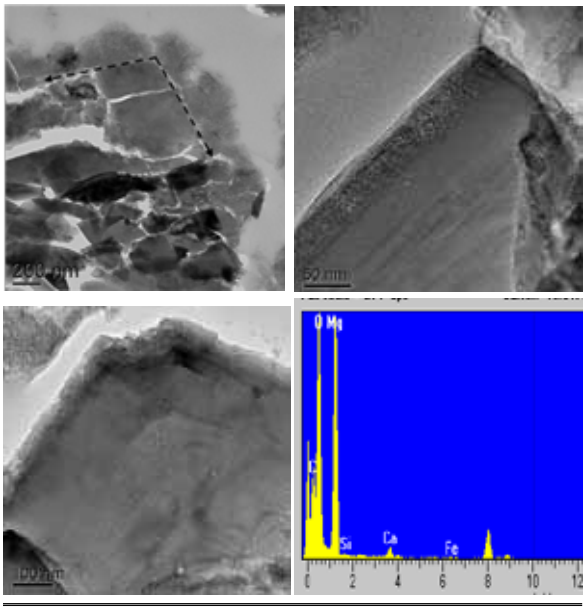



Figure 2. Crystallites from upper layer of the Hot Lake mat sampled just above the green cyanobacterial layer. The crystallites are magnesite ($MgCO_3$) as determined by selected area electron diffraction (SAED) with amorphous coatings of similar elemental composition as determined by energy dispersive X-ray spectroscopy (EDS).

This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP). This research is part of the Genomic Science Program FSFA at Pacific Northwest National Laboratory.

# 191

## Mobile Gene Pools and the Functions They Encode Give Insight into How Genome Evolution is Shaped by Interactions Between Microbes and Their Environment

Margrethe Serres[1]* (mserres@mbl.edu), Sebastian Jaramillo Riveri,[2] and **Margie Romine**[2]

[1]Marine Biological Laboratory, Woods Hole, Mass.; and [2]Pacific Northwest National Laboratory, Richland, Wash.

**Project goals: Differences in a microbe's ability to survive under various environmental conditions has been linked to the presence of genes conferring an advantage to its host. Some of these fitness genes have been obtained from other organisms residing in the same environment. Integron integrases represent one such mode of lateral gene transfer. The IntI integron integrase inserts foreign DNA through site-specific recombination between an AttI site in the integron and an AttC site present on a mobile DNA segment. The integron also encodes a promoter that facilitates the expression of the inserted gene(s). Detection of integron associated gene cassettes and the study of their encoded functions may give insight into how organisms enhance their functional efficiency and adapt to environmental factors. Analyzing the evolutionary history of the exchanged elements may also give insight into co-occurring organisms and their ability to affect genome evolution. The initial study is done on species belonging the genus *Shewanella* while future work will be directed at environmental samples of mats from Yellowstone National Park hot springs and meromictic saline lakes.**

**Current results**

We are studying integron associated gene cassettes in the genus *Shewanella*. IntI integrases have been reported to be present in strains belonging to this genus. Also, as a group the *Shewanellas* have adapted to many environments that vary in their nutrient sources, salinity, temperature, and pressure. The genome sequences and ortholog gene sets are available for 21 closely and distantly related *Shewanella* strains.

One or two copies of IntI integron integrases were detected in 12 of the 21 *Shewanella* genomes. The IntI genome neighborhoods were searched for AttI and AttC integration sites. Sequence patterns of the identified *Shewanella* Att sites were then used for Blastn analyses of the *Shewanella* genomes. We found recombination sites with adjacent gene cassettes in all 21 genomes. A total of 1137 genes were identified as belonging to the integron associated gene cassettes. The number of integron cassette genes per genome ranged from 96 (*S. woodyii*) to 29 (*S. halifaxensis*). Genomes that encode two IntI genes had an overall higher number of integron integrase cassette genes. Approximately half of the integron gene cassettes consisted of one or two genes, with the remaining containing 3 to 14 consecutive genes. The largest gene cassettes had 12-14 genes, and these were found in the genomes of *S. amazonensis*, *S. baltica* OS195, *S. baltica* OS185, and *S. frigidimarina*.

The protein sequences encoded by the 1137 genes were analyzed for their functions. We found that 65% or the proteins were annotated as (conserved) hypothetical proteins. Others had functional descriptions with the most abundant being GCN5-related acetyl transferase, glutathione-dependent formaldehyde-activating enzyme, cytoplasmic adenylate cyclase, and glyoxylase family protein. To improve the annotation of the integron cassette encoded proteins we searched for conserved regions, or protein domains, with similarity to HMM models of proteins in the Pfam and TIGRFAM databases. Families of proteins with common domain(s)

are known to encode similar functions. A blast analysis using trusted cutoffs detected similarity to proteins in the two databases for 879 of the integron cassette proteins. We detected 406 distinct Pfam and 355 distinct TIGRFAM domains, with the most prevalent domains encoding the most abundant protein functions listed above. The domain information is currently used to improve the annotation of the integron associated cassette proteins. Cellular roles associated with the TIGRFAMs indicate that the integron gene cassettes are enriched in proteins related to protein synthesis, cellular processes, energy metabolism, and regulation.

A blast analysis was done against the nr database to get insight into the evolutionary history of the laterally transferred gene pool. We did not include similarity to other *Shewanella* proteins as some of the integron cassette proteins had orthologs in the other *Shewanella* genomes.

The protein-pair with the lowest e-value was extracted for each of the integron cassette proteins. The most abundant sequence matches were to *Colwellia psyrerythraea* (40), *Pseudoalteromonas tunica* (38), *Alteromonas macleodii* (19), *Moritella* (18), and *Idiomarina loihiensis* (17). These microbes belong to the *Alteromonodales*, along with *Shewanella*. When counting sequence matches at the genus level, *Vibrio* (141) was significantly higher than the rest; *Pseudoalteromonas* (52), *Colwellia* (40), and *Pseudomonas* (35), likely reflecting the abundance of sequenced *Vibrio* genomes. This pattern of high similarity to *Vibrio* proteins did not differ when comparing different phylogenetic groups within *Shewanella* or when comparing proteins with or without *Shewanella* orthologs. Analyses of specific protein functions and of selected integron associated cassettes will be presented.

This research is part of the Genomic Science Program FSFA at Pacific Northwest National Laboratory.

# 192

## Metagenome Analysis of High-Temperature Chemotrophic Microbial Communities Provides a Foundation for Dissecting Microbial Community Structure and Function

W. Inskeep[1]* (binskeep@montana.edu), M. Kozubal,[1] J. Beam,[1] Z. Jay,[1] R. Jennings,[1] H. Bernstein,[2] R. Carlson,[2] D. Rusch,[3] S. Tringe,[4] M. Romine,[5] R. Brown,[5] M. Lipton,[5] J. Moran,[5] H. Kreuzer,[5] C. Ehrhardt,[5] and J. Fredrickson[5]

[1]Department of Land Resources and Environmental Sciences and Thermal Biology Institute, Montana State University, Bozeman; [2]Department of Chemical and Biological Engineering, Montana State University, Bozeman; [3]J. Craig Venter Institute, Rockville, Md.; [4]Department of Energy-Joint Genome Institute, Walnut Creek, Calif.; [5]Department of Energy-Pacific Northwest National Laboratory, Richland, Wash.

**Project Goals: Use systems biology approaches to dissect community-level structure and function in geothermal microbial communities of YNP.**

Microbial communities are a collection of interacting populations, each comprised of numerous individuals. However, a significant fraction of our knowledge base in microbiology originates from organisms grown and studied in pure culture, in the absence of other members of the community who may compete for resources or provide necessary cofactors and or substrates. Moreover, many of the organisms studied in pure culture have not represented the numerically dominant members of microbial communities found in situ. The advent of molecular tools (and -omics technologies) has provided opportunities for assessing the predominant and relevant indigenous organisms, as well as their likely function within a connected network of different populations (i.e., community). High-temperature microbial communities are often considerably less diverse than mesophilic environments and constrained by dominant geochemical attributes such as pH, dissolved oxygen, Fe, sulfide, and or trace elements including arsenic and mercury. Consequently, the goal of our work is to utilize high-temperature geothermal environments including acidic Fe-oxidizing communities as model systems for understanding microbial interactions among community members (Figure 1).

Recent metagenomic sequencing of high-temperature, acidic Fe-mats of Norris Geyser Basin, Yellowstone National Park (YNP) reveal communities dominated by novel archaea, members of the deeply-rooted bacterial Order Aquificales, and less-dominant Bacillales and Clostridiales. Phylogenetic and functional analysis of metagenome sequence is providing an excellent foundation for hypothesizing the role of individual populations in a network of interacting community members, and for testing specific hypotheses regarding the importance of biochemical pathways responsible for material and energy cycling. For example, we are using metagenome sequence in combination with information available from reference strains to identify protein-coding sequence of importance in the oxidation and or reduction of Fe, S, O, and As, as well as central C metabolism (including fixation of $CO_2$). Genes coding for proteins with hypothetical or putative roles in electron transfer, and C-cycling are being investigated using quantitative-reverse transcriptase-PCR (Q-RT-PCR) to evaluate mRNA levels in both pure-culture and mixed communities. Future transcriptomic and proteomic analyses will be coupled with detailed studies focused on the position of different organisms (spatial context) during Fe-mat development, as well as the role of $O_2$ flux across Fe-oxidizing boundary layers. Depositional studies have been conducted to correlate Fe-oxide deposition rates with $O_2$ flux rates measured using $O_2$-microelectrodes. Microelectrode measurements at the Fe-oxide-aqueous interface suggest significant $O_2$ consumption driven by the oxidation of Fe(II), but also show that sub-oxic conditions are common below the mat surface.

Metagenome sequence is being used to build consensus genome sequence of 5-6 dominant community members and will serve as a foundation for future sample analysis,

laboratory data integration, and modeling efforts. Metabolic models for individual populations are constructed using elementary flux mode analysis, and these sub-components are then combined in a community model to explore possible ramifications of substrate interaction patterns on microbial community structure and function. Pathway specific processes are also being elucidated using isotope measurements focused on $^{13}C$ and $^{34}S$ of different chemical fractions with the goal of coupling this information to population-specific energy and nutrient cycling. Transcriptomic and proteomic results will be used to assess and confirm the importance of specific pathways and processes, and in conjunction with complementary datsets on C-isotopes and metabolomics, will allow refinements to individual and or community network models. Application of genomic, proteomic, and metabolic information to dissect microbial community structure and function is tractable within high-temperature geothermal systems, in part due to the relative simplicity of the communities and the stability of several key geochemical variables (i.e. pH, Fe, $O_2$).

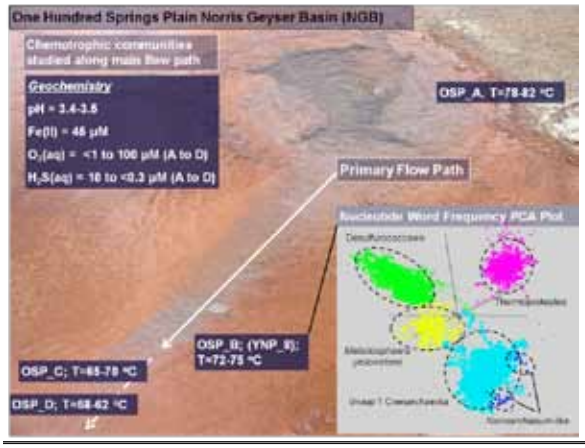



Figure 1. Site photograph of Fe-oxide depositing geothermal spring (One Hundred Spring Plain, OSP) located in Norris Geyser Basin (YNP). Reduced geothermal waters emerge with little to no detectable dissolved oxygen and significant levels of ferrous Fe (~50 µM). The oxidation of Fe(II) to produce amorphous Fe(III)-hydroxide is catalyzed by microbial populations such as *Metallosphaera yellowstonensis*. Inset at right indicates predominant microbial populations present at OSP_B (~72-75 C) identified using nucleotide word frequency plots of assembled metagenome sequence.

This research is part of the Genomic Science Program FSFA at Pacific Northwest National Laboratory.

# 193

## Energy Carbon and Nutrient Partitioning in Lab-Based Phototroph-Heterotroph Co-Cultures

Alex S. Beliaev[1]* (alex.beliaev@pnl.gov), Grigoriy E. Pinchuk,[1] Oleg V. Geydebrekht,[2] Jennifer L. Reed,[2] Donald A. Bryant,[3] Allan E. Konopka,[1] Thomas Metz,[1] Sergey Stolyar,[1] and **Jim K. Fredrickson**[1]

[1]Pacific Northwest National Laboratory, Richland, Wash.; [2]University of Wisconsin-Madison; and [3]Pennsylvania State University, University Park

**Project Goals: The overarching goal of the PNNL Foundational SFA is to understand the collective energy, carbon, and nutrient processing in phototrophic microbial communities that contributes to their stability and efficient utilization of resources.**

Experimental systems for hypothesis testing include a well-defined engineered co-culture (e.g., *Synechococcus* 7002 – *Shewanella putrefaciens* W3-18-1), individual organisms (e.g., *Synechococcus* spp., *Thermosynechococcus* spp., *Roseiflexus* spp.), and consortia derived from or representative of naturally-occurring microbial communities (e.g., *Thermosynechococcus-Roseiflexus*). This researchis organized around three primary objectives which, in addition to science-driven tasks, encompass technical milestones thus enabling the transition from constructed opportunistic co-cultures to consortia derived from natural communities. The methods development efforts are focused primarily on analytical approaches for measuring biomass composition and mRNA/protein abundances in mixed cultures as well establishing techniques for identification of interactions between microorganisms. To identify secreted organic compounds that may serve as the primary carbon and energy sources for heterotrophic microbes growing in association with autotrophic species, we have tested the applicability of MS- and NMR-based detection techniques in conjunction with Chenomx metabolite library. While GC-MS had limited utility due to high concentration of interfering cationic salts, cNMR approach correctly identified and quantified dissolved organic acids, alcohols, and sugars in complex growth media.

To begin understanding the mechanisms governing the growth of photoautotroph-heterotroph associations, we are focusing on developing an integrated constraint-based flux balance model of the co-culture metabolism under environmentally relevant conditions. In initial studies, we used *Synechococcus* 7002-*S. putrefaciens* W3-18-1 co-culture to better understand the pathways of carbon and energy partitioning under different types of growth limitations relevant to natural associations which include limitations by carbon and light. Specifically, *Synechococcus* 7002 and *S. putrefaciens* W3-18-1 were successfully grown together in both batch and chemostat modes with lactate serving as the only source of carbon for both cultures. No mass-transfer of gases was applied suggesting that $O_2$ required to for com-

plete lactate oxidation by *S. putrefaciens* W3-18-1 produced sufficient amounts of $CO_2$ to maintain growth of *Synechococcus* 7002. Our experiments also revealed that *Synechococcus* 7002 cannot grow in dark or in light using either lactate or acetate as the sole source of carbon while *S. putrefaciens* W3-18-1 cannot grow on lactate in the absence of $O_2$. Therefore, the co-culture utilizing light as the only source of energy and lactate as the sole source of carbon could only occur as a result of tight metabolic coupling between the phototrophic and heterotrophic organisms. *Synechococcus* 7002-*S. putrefaciens* W3-18-1 co-culture displayed balanced steady-state growth ($\mu$=0.05 $h^{-1}$) under 1600 $\mu$mol/$m^2$/sec irradiance when dissolved oxygen tensions (dOT) were kept below 1%. However, under high dOT (160% of air saturation) and light intensity (2040 $\mu$mol/$m^2$/sec) the co-culture formed aggregates that primary consisted of filament-like *S. putrefaciens* W3-18-1 cells. While factors underlying these morphological changes are yet to be determined, we hypothesize that filamentous growth may play an important role in maintaining stable phototrophic biofilms which provide protection against reactive oxygen species. Notably, we have achieved sustainable growth of binary culture on inorganic source of carbon under light conditions. *S. putrefaciens* W3-18-1 was maintained in a chemostat mode for more than 15 generations using excreted photosynthate being the only carbon and energy source. Although initial experiments and methods development focus on opportunistic *Synechococcus* 7002-*S. putrefaciens* W3-18-1 co-culture, the naturally-occurring phototrophic mat communities provide an excellent opportunity to explore inter- and intra-guild metabolic interactions among specific populations associated with these communities. In a parallel effort, we initiated a study of a binary co-culture of *Thermocynechococcus* sp. NAK55 and *Roseiflexus castenholzii* DSM 13941; both organisms were isolated from phototrophic mat of Nakabusa hot spring (Japan). *Thermosynechococcus* NAK55 (kindly provided by Dr. S. Haruta) was successfully grown in liquid BG11 at 52°C without agitation under 20µE light. *R. castenholzii* DSM 13941 also grew on BG11 supplemented with 0.4% yeast extract in oxic and anoxic atmosphere in the dark and anoxic in the light at 52°C without agitation. Our current work is directed at revealing the nature of interactions as well as understanding the pathways of carbon, energy, and nutrient partitioning in both opportunistic and naturally occurring co-cultures.

An *in silico* model of metabolic coupling is also being constructed using *Synechococcus* 7002 and *S. putrefaciens* W3-18-1 constraint-based models. For the first assessment of *Synechococcus* 7002 growth *in silico* we used previously developed metabolic model for *Cyanothece* sp. ATCC 51142. It was estimated that the maximal biomass yields of *Synechococcus* 7002 was 0.024gAFDW/mmol $CO_2$ assuming $CO_2$ is the limiting substrate and the ratio of $O_2$ produced per $CO_2$ consumed at the maximal biomass yield was 1.05 $O_2/CO_2$. These parameters were used to approximate the behavior of *Synechococcus* sp. PCC 7002 in the binary culture. The metabolic model for *S. oneidensis* MR-1 (*i*SO783) was used to predict the behavior of *S. putrefaciens* W3-18-1 in co-culture, since 699 out of 783 genes in the MR-1 model have orthologs in W3-18-1. Most of the 29 reactions associated with the 74 missing orthologs from *S. putrefaciens* W3-18-1 were excluded from the model, except for three reactions that were essential. We estimated the growth associated ATP requirements for W3-18-1 assuming that it has 15% higher biomass yields than MR-1 as it was previously estimated. The model-predicted total biomass concentration in binary culture using 5 mM lactate as the sole source of carbon was in the range 0.3579 - 0.3863 gAFDW. Experimental assessment of chemostat lactate-limited binary culture growing in chemostat revealed that sum of total biomass and extracellular organic carbon was 0. 379 g/l, therefore confirming model prediction.

This research is part of the Foundational SFA at Pacific Northwest National Laboratory.

# 194

## Decreased Protein Oxidative Damage Following Opportunistic Microbial Associations

Diana J. Bigelow,[1] Donald Bryant,[2] Baowei Chen,[1] **James K. Fredrickson,[1]** Na Fu,[1] Grigoriy E. Pinchuk,[1] Gaozhong Shen,[2] Thomas C. Squier[1]* (thomas.squier@pnl.gov), Sergey Stolyar,[1] and Yijia Xiong[1]

[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash.; and [2]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park

**Project Goals: To identify stress resistance and adaptation mechanisms in microbial communities, we have the following goals:**

- **Using detection assays against common oxidative modifications, use high-throughput screening approaches to explore how the efficient coupling of diverse metabolisms may prevent oxidative damage to biomolecules in model organisms *Shewanella* in culture with the cyanobacterium *Synechococcus*, and initiate measurements in natural systems involving lichins and co-cultures that reconstitute aspects of natural mat systems.**
- **Utilize advanced mass spectrometry capabilities available at PNNL to discover possible sites of oxidative and other post-translational modifications within microbial proteins from lysates isolated from axenic cultures of *Shewanella* and *Synechococcus*.**
- **Affinity isolate targeted oxidative modifications from natural populations as a function of the diel cycle, permitting the identification of likely sensors of oxidative stress in specific classes of bacteria.**

As part of PNNL's Scientific Focus Area "Biological Systems Interactions" this project is identifying regulatory proteins responsive to oxidative stress, which are hypothesized to play central roles in promoting stable associations between cyanobacteria and other microbes. Our strategy is to identify sequence-specific oxidative modifications in sensor proteins that could act to regulate

metabolic fluxes and thereby control the excess generation of reactive oxygen species (ROS), which would promote adaptive responses that enhance cell survival. To identify fundamental adaptive strategies, initial efforts are focusing on laboratory model systems involving the cyanobacterium *Synechococcus* sp. PCC 7002 in culture with *Shewanella oneidensis* W3-18-1. We find that for cultures of the photoautotroph *Synechococcus* sp. PCC 7002 grown in the presence of the heterotroph *Shewanella* that there are large decreases in the overall levels of protein oxidation (Figure 1). Reductions in oxidative stress are apparent despite the substantially higher oxygen present in the co-culture (160% dissolved air saturation) in comparison to axenic cultures (44% dissolved air saturation) (both cultures are grown in caged bioreactors using white light intensities of 240 µmol photons $m^{-2}$ $s^{-1}$) with shaking of 150 rpm. These results support the hypothesis that opportunistic interactions between heterotrophic (*Shewanella*) and photosynthetic (*Synechococcus*) microbes permit metabolic coupling to enhance energy efficiencies and community stability. Further, our results are consistent with prior indications that axenic isolates of the cyanobacterium *Synechococcus* sp. isolated from the microbial mat of Octopus Spring in YNP have a substantially enhanced sensitivity to light-induced oxidative stress in comparison to the natural mat community. This finding indicates an important role for metabolic coupling between community members in the mat that promote stress resistance.

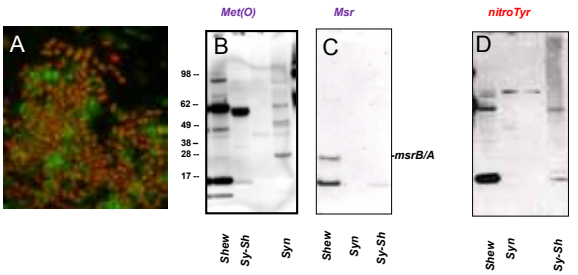



Figure 1. *Opportunistic Microbial Associations Diminish Protein Oxidation*. (Left) Image of co-culture showing *Synechococcus* (red) and *Shewanella* (green) (A). (Right) Immunoblots against methionine sulfoxide [Met(O)] (B), methionine sulfoxide reductase (Msr) (C), and nitrotyrosine (nitroTyr) (D) comparing extent of oxidative damage in co-culture (Sy-Sh) with that seen in axenic cultures of either *Synechococcus* (Syn) or *Shewanella* (Shew).

As hemes promote peroxidase and Fenton chemistries that can result in their oxidative modifications, we have examined the possibility that heme-containing proteins are oxidatively sensitive through a consideration of methionine sulfoxide formation. In the case of the major heme proteins, there is no significant methionine oxidation observed with the exception of one, fumarate reductase (Figure 2). The high level of methionine oxidation apparent for fumarate reductase, suggests possible linkages between oxidative stress conditions and alterations in metabolic flux that may arise due to oxidant-induced changes in fumarate metabolism. Given the positions of methionines within the fumarate reductase structure (SO0970; 1D4C.pdb), which occur within the active site and proximal to heme interfaces critical to efficient electron transfer, it is likely that methionine oxidation may functionally uncouple fumarate binding and its oxidation under aerobic conditions. As methionine oxidation is reversible through the actions of methionine sulfoxide reductases, these measurements suggest a means to maintain an ability to control the use of fumarate as an electron acceptor through its functional regulation in response to oxygen levels.

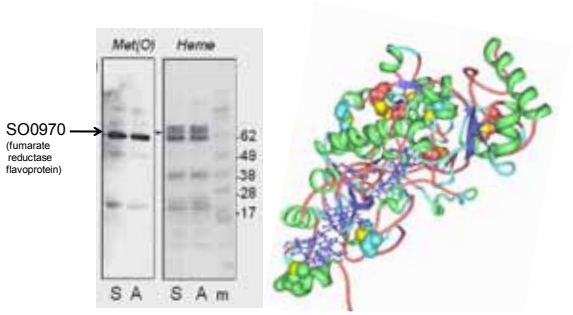



**Figure 2**: Targeted oxidation of methionines in fumarate reductase protein in Shewanella in the presence of suboxic (S) or aerobic (A) environmental conditions. (Left) Immunoblot against methionine sulfoxides [Met(O)] or heme stain (Heme) following SDS-PAGE protein separation. (Right) Structure of fumarate reductase (1D4C.pdb) highlighting positions of methionines (spacefilling) relative to hemes and FAD (purple stick representation). Colors of methionines correspond to backbone representation, where sulfurs are yellow.

Additional regulation through site-specific oxidative modifications involves the heme protein GlbN, which has been shown to protect *Synechococcus* sp. PCC 7002 against oxidative stress associated with growth on nitrate (Scott et al., 2010; Biochemistry 49, 7000). These growth conditions are suggested to mimic environmental conditions associated with denitrification, where cyanobacteria that express GlbN grow in association with soil organisms that produce nitric oxide as a result of anaerobic respiration using nitrate as an electron acceptor. We are examining the role of GlbN as an antioxidant protein against peroxynitrite, or possibly as a protein sensor that undergoes autocatalytic activation in response to oxidative stress, using mutant strains grown in the presence of high nitrate. Inspection of the GlbN structure (2KSC.pdb) indicates a close proximity between Tyr5 and Tyr22 and basic Lys side chains that we have previously demonstrated to result in the sensitive nitration of tyrosines in a range of proteins. Positions of these tyrosines within the tertiary structure suggest an ability to stabilize the fold of GlbN, which has the potential to create necessary binding pockets that enhance function against, for example, peroxynitrate. Isolation of the intact protein following growth on nitrate, in conjunction with *in vitro* measurements of function, offer a means to identify the role of GlbN and other sensor proteins that may function to control intracellular metabolisms that enhance community stability.

This research is part of the Genomic Science Program FSFA at Pacific Northwest National Laboratory.

# 195

## Genomic Reconstruction of Vitamin Metabolism in Microbial Communities

**Andrei Osterman**[1]* (osterman@burnham.org), Dmitry Rodionov,[1] Leonardo Sorci,[1] Margaret Romine,[2] Jim K. Fredrickson,[2] Samantha Reed,[2] and Nadia Raffaelli[3]

[1]Burnham Institute for Medical Research, La Jolla, Calif.; [2]Pacific Northwest National Laboratory, Richland, Wash.; and [3]Università Politecnica delle Marche, Ancona, Italy

**Project Goals: Genomics-based prediction and experimental validation of gene functions, pathways and networks in targeted groups of heterotrophic and phototrophic microbes for the fundamental understanding of microbial ecophysiology, evolution, adaptation and associations.**

Metabolic cross-feeding is believed to play an important role in microbial communities. Whereas typical metabolic byproducts may provide a major flux of carbon and energy, vitamins (precursors of key cofactors) are required in relatively small amounts. Therefore, vitamin exchange may be rather widespread phenomena contributing to "opportunistic" relationships between species. This notion is consistent with our recent genomic survey, which confirmed that most environmental bacteria harbor both, de novo and salvage pathways for biogenesis of major vitamins (such as niacin, pantothenate, biotin, thiamin, riboflavin). Such species have a potential to benefit from as well as contribute to the vitamin pool in the environmental niche. At the same time, the observed mosaic distribution of *de novo* and salvage pathways leads to the presence in communities of the strict auxotrophs and strict prototrophs with respect to one or another vitamin. An ability to accurately reconstruct vitamin metabolic pathways and predict respective phenotypes over a rapidly growing collection of sequenced microbial genomes would impact our understanding of the metabolic cross-talk in microbial communities. To address this long-range goal we combine comparative genomics and predictive bioinformatic techniques with the experimental assessment of vitamin biochemical, transport and regulatory pathways in individual bacterial species and in model co-cultures. Using a subsystems-based approach captured in The SEED genomic platform, we were able to elucidate major vitamin biosynthetic and salvage pathways over a collection of >1,000 diverse bacterial genomes (seed-viewer.theseed.org). In addition to accurate projection of knowledge from model species to many others, this approach allows prediction of previously unknown genes and pathway variants. This is illustrated by the example of NAD metabolism, which is indispensable in nearly all analyzed species and reveals remarkable variations of associated pathways and regulatory mechanisms. Our recent analysis is focused on two model environmental bacteria, *Synechococcus* and *Shewanella*. A co-culture formed by the representatives of these groups, *Synechococcus sp. PCC 7002* and *Shewanella sp. W3-18-1* (under study at PNNL), provides a tractable model of microbial interactions between phototrophs and heterotrophs. Notably, both species appear to harbor divergent and unusual salvage pathways. Thus, *Synechococcus* sp. *PCC 7002* genome encodes an unprecedented (for Cyanobacteria) combination of PncA/PncB-mediated salvage of nicotinamide (Nm) with a possible NadR-driven utilization of ribosyl-Nm. At the same time, *Shewanella* sp. W3-18-1 appears to be one of the few species in this group with the unusual version of NadV-mediated Nm salvage pathway. A previously unknown gene encoding NMN deamidase (NadH) postulated for the second step of this pathway was cloned, and the respective purified recombinant enzyme was characterized. NadH orthologs are conserved in many diverse bacteria where they are likely involved in NMN recycling. The physiological role and the contribution of salvage genes and pathways to the overall NAD biogenesis are tested by a combination of genetic and metabolic profiling techniques. Remarkably, some (but not all) of the species in both groups, despite their taxonomic and metabolic distinctions, share a novel transcriptional regulator of NAD metabolism, NrtR. Regulons controlled by NrtR were analyzed as a part of our systematic genomic reconstruction of regulatory networks in *Shewanella* and Cyanobacteria and captured in the RegPrecise database (regprecise.lbl.gov/RegPrecise). Genomics-driven searches for additional regulatory mechanisms and presently unknown transporters involved in vitamin salvage and/or excretion are presently in progress.

This research is part of the Genomic Science Program FSFA at Pacific Northwest National Laboratory.

# 196

## Global Metatranscriptomic Analyses of the Chlorophototrophic Microbial Mat Community of Mushroom Spring

Zhenfeng Liu,[1] Marcus Ludwig,[1] Christian G. Klatt,[3] David M. Ward,[3] and **Donald A. Bryant**[1,2]* (dab14@psu.edu)

**Lead Principal Investigator:** Jim K. Frederickson (Jim.Fredrickson@pnl.gov)

[1]Dept. of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park; [2]Dept. of Chemistry and Biochemistry, and [3]Dept. of Land Resources and Environmental Sciences, Montana State University, Bozeman

**Project Goals: The overarching goals of our studies are to understand the composition of the microbial mat communities associated with alkaline siliceous hot springs, to understand the physiological capabilities of the individual populations of organisms, and to understand how these populations interact metabolically within the mat ecosystem.**

The chlorophototrophic microbial mats of alkaline siliceous hot springs in Yellowstone National Park have served as models for studying the structure and function of microbial communities for decades. Culture-independent methods,

including metagenomics and metatranscriptomics, have been used to define the composition and physiological potential of the organisms comprising these mats. These studies have showed that the mats of Mushroom and Octopus Spring are composed of 8 major populations, belonging to six kingdoms: *Cyanobacteria* (*Synechococcus* spp.), *Chloroflexi* (*Roseiflexus*, *Chloroflexus*, and *Anaerolinea*-like spp.), *Chlorobi*, *Acidobacteria*, *Firmicutes*, and *Bacteroidetes*. Three new chlorophototrophs (*Candidatus* Chloracidobacterium thermophilum, *Candidatus* Thermochlorobacter aerophilum, and a novel phototrophic member of the *Chloroflexi*) were discovered in these mats. The overarching goals of our studies are to understand the composition of this community, to understand the physiological capabilities of the individual populations of organisms, and to understand how these populations interact metabolically within the mat ecosystem. In this study the global transcriptome of the chlorophototrophic mat community of Mushroom Spring at 60-62°C was characterized.

Total RNA was extracted from chlorophototrophic mat samples collected at Mushroom Spring. Small RNAs (<300 nt) were removed and the remaining RNAs were reverse-transcribed to produce cDNAs. Four cDNA samples collected during light transition periods were initially sequenced by pyrosequencing (GX20 FLX) and SOLiD-3.5 technologies. The sequences produced by pyrosequencing were mainly used for rRNA analyses, while the SOLiD datasets were analyzed to discern gene regulation patterns. Sequences generated by pyrosequencing were aligned to rRNA databases using BLASTN. The composition of rRNA sequences was assessed from MEGAN analyses of the sequence alignments. The SOLiD datasets were aligned to assembled metagenome scaffolds (primary reference) and selected complete genomes of isolates from similar environments (secondary references) using the bwa algorithm and allowing 5 mismatches for the 50-bp sequences. Using artificially generated datasets, simulations of alignments allowing 0 to 5 mismatches showed that allowing 5 mismatches ensured a maximal number of correctly aligned sequences, a very low error rate, and a reasonable computation time. Test alignments using metagenome and complete genomes as references showed that complete genomes of biologically relevant isolates could serve as satisfactory references only when the sequence differences from the metagenome consensus were negligible. Gene expression patterns could be inferred by plotting normalized mRNA sequence counts for each gene as a function of time during a diel cycle. To minimize differences in sequencing coverage, organism abundance and cellular energy levels among samples, mRNA sequence counts for each gene were normalized to the total mRNA sequence counts of the organismal population to which the gene belonged. Finally, the resulting normalized data could be organized by the program "Cluster" and visualized using "Java Treeview" to reveal groups of genes exhibiting transcription differences as a function of the diel cycle.

Members of four kingdoms (*Cyanobacteria, Chloroflexi, Acidobacteria*, and *Chlorobi*) accounted for ~84% of the rRNA sequences obtained by pyrosequencing. The remaining 16% of rRNA sequences showed similarity to many different rRNA sequence types. The most abundant sequences were similar to those of *Firmicutes* and *Bacteroidetes* (<2% each). Additional RNA samples were collected at 1-h intervals over a complete diel cycle. The amount of mRNA in cells varied significantly during the diel cycle and ranged from a low value of ~3% of the RNA sequences at night to ~12% in the late afternoon. For all chlorophototrophs in the mat, mRNA levels were highest during the midday when light intensity was highest. Changes in transcription occurred in all populations as a function of the diel cycle, and cluster analysis showed that transcripts could be assigned to two to four pattern classes for each organism during the course of the diel cycle. The transcription patterns (and the metabolism they represented) observed for oxygenic photoautotrophs were distinctly different from the patterns observed for aerobic anoxygenic photo(hetero)trophs (AAPs). Members of the *Cyanobacteria* were transcriptionally most active during the light period and maximally expressed their genes for components of the photosynthetic apparatus during the light period. As previously documented, expression of nitrogen fixation and fermentation-specific genes by cyanobacteria began in the late afternoon and continued until full sunlight again reached the mat in the morning. The global transcription patterns for members of the *Chloroflexi*, *Candidatus* Chloracidobacterium thermophilum and a *Chlorobiales* population were similar but differed from the patterns observed in the cyanobacteria. Transcript levels for genes encoding components of the photosynthetic apparatus peaked in *Roseiflexus* spp. populations were minimal during the daylight period but increased in the evening and were maximal in the early morning. Similarly, photosynthetic apparatus transcripts from *Candidatus C. thermophilum* and *T. aerophilum* peaked in the late afternoon and continued to be present at high levels throughout the night, but they were minimal during the daylight period. The transcription patterns suggested that factors other than light, most likely oxygen level, determine the periods of highest transcriptional activity for AAPs and specifically determine when the genes encoding components of the photosynthetic apparatus are expressed. The data provide an explanation for why organisms living in an environment characterized by very high light intensity have large antenna complexes such as chlorosomes: periods of high metabolic and transcriptional activity do not coincide with the periods of highest light intensity.

This microbial community is also being studied by microscopy, proteomics, and metabolomics to produce a complete model for the structure and metabolic interactions that occur in this complex but tractable model ecosystem.

This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP). This research is part of the Genomic Science Program FSFA at Pacific Northwest National Laboratory.

# 197

## Composition and Structure of Phototrophic Hot Spring Microbial Mat Communities: Natural Models for Systems Biology

**David M. Ward**[1]* (umbdw@montana.edu), Melanie C. Melendrez,[1] Eric D. Becraft,[1] Christian Klatt,[1] Jason Wood,[1] Donald A. Bryant,[2,3] and Frederick M. Cohan[4]

[1]Land Resources and Environmental Sciences Department and [2]Dept. of Chemistry and Biochemistry, Montana State University, Bozeman; [3]Dept. of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park; [4]Department of Biology, Wesleyan Univ., Middletown, Conn.

**Project Goals: Our overarching goal is to understand relationships between composition and function in a relatively simple natural phototrophic mat community inhabiting Yellowstone alkaline siliceous hot springs we have studied for many years in order to learn principles of how complex, multi-species biological systems work. Our specific goals are (i) identifying major populations using metagenomic assembly and recruitment with highly representative genomes, (ii) exploring species composition within these groups using rapidly evolving gene sequences or sets of gene sequences and theory-based evolutionary simulations to conduct single- and multi-locus population genetics analyses, (iii) examining the adaptations of different species through comparative genomics of representative isolates, (iv) observing patterns of community gene expression and protein synthesis through metatranscriptomic and metaproteomic analyses, (v) exploring ways to link important community functions, such as carbon sequestration.**

Understanding how a complex biological system, such as a microbial community, works requires knowledge of its component microorganisms. It is essential that community composition be understood at the species level, as species are the fundamental taxonomic units that are uniquely distributed into distinct niches, make unique functional contributions and uniquely respond to the complex and dynamic natural environment. We have conducted long-term studies of microbial mats inhabiting alkaline siliceous hot springs, which are relatively simple, stable, high-biomass and very accessible systems that are protected within Yellowstone National Park. Our goal is to understand relationships between composition and function in this relatively simple natural community in order to learn principles of how complex biological systems work.

Initial impressions of community composition came from analyzing the genetic variation in 16S rRNA sequences. These studies revealed that the dominant native cyanobacteria (*Synechococcus* spp.), which play a major role in constructing the mat by performing oxygenic photosynthesis, are very unlike, and more diverse than, readily cultivated *Synechococcus* spp. isolates. Similarly, the dominant filamentous anoxygenic bacteria inhabiting the mat, which are thought to use infrared light energy to assimilate organic compounds excreted by *Synechococcus* spp., were found to be *Roseiflexus* spp., not the readily cultivated *Chloroflexus* spp.

The differential distribution of 16S rRNA variants along the effluent flow path (temperature gradient) and vertically in the upper portion of the mat photic zone (light gradient) suggested the presence of multiple, ecologically distinct *Synechococcus* populations. These populations might, like the Galapagos finches, have resulted from adaptive evolutionary radiation to fill distinct niches. However, analyses with more rapidly evolving genes, which offer higher molecular resolution, in combination with computer simulations of the evolution of ecological species, revealed that 16S rRNA sequence variation is unable to resolve all *Synechococcus* spp. in the mat. These approaches predict the existence of several tens of ecologically distinct *Synechococcus* spp., many of which exhibit unique spatial distributions, unique patterns of gene expression during the diel cycle and respond uniquely after imposing environmental change, as expected of ecological species. Cultivation-independent multi-locus sequence analyses, conducted by using bacterial artificial chromosome cloning to retrieve large genome segments of community members (*i.e.*, about 100 genes each), have demonstrated that genetic recombination among *Synechococcus* spp. has occurred, but has not been frequent enough to prevent the existence and detection of ecological species.

The relatively low genetic diversity of these mat communities has enabled the use of metagenomic technologies to describe its major types of inhabitants comprehensively. In particular, the low diversity permits assembly of paired-end sequences from individual metagenomic clones (2,000-12,000 nucleotides in length) into much longer contiguous genomic assemblies up to 1,600,000 nucleotides long), based on the similarity of overlapping regions of closely related sequences from the individual clones. These longer sequences are referred to as "metagenomic scaffolds", where "meta-" indicates that the scaffolds are comprised of sequences from the genomes of phylogenetically similar organisms, not a single genome. Scaffolds that have similar phylogenetic characteristics can be grouped together using k-means cluster analysis, which also separates scaffolds comprised of phylogenetically distinct sequences. Once annotated, these clustered scaffolds provide a means of linking genes that indicate phylogenetic affiliation with genes that give insight into the functional potential of the organisms represented by each cluster. Cluster analyses have revealed that 8 major populations predominate in the upper photic region of these mats. Clusters representing *Synechococcus* spp., *Roseiflexus* spp. and *Chloroflexus* spp. and a novel phototrophic acidobacterium (*Candidatus* Chloracidobacterium thermophilum), which we had discovered through analyses of functional gene sequences (photosystem reaction center genes) and later cultivated, were expected. A fifth cluster, confirmed the presence of a novel *Chlorobiales* population, which had been suspected based on the independent detection of 16S rRNA and reaction center sequences typical of members of the *Chlorobiales*. Three clusters represented predominant populations which had heretofore evaded detection, a remarkable finding for a system that has been

so thoroughly investigated by numerous microbial ecologists for so long. One of these populations contains genes that indicate that it is a novel phototrophic member of Kingdom *Chloroflexi*. Its distant relationship to *Chloroflexus* spp. and *Roseiflexus* spp, suggests that phototrophy may have been an ancient phenotype in this kingdom. The 6 phototrophic clusters and their different gene contents suggest that diverse phototrophic guilds specialized to use different light wavelengths and autotrophic pathways, participate to coordinate efficient light capture and carbon sequestration in these mats. Two clusters, representing as yet unknown members of *Firmicutes*, and *Bacteroidetes*, do not contain genes indicating a potential to conduct phototrophy, and are likely contributed by heterotrophic members of the mat community, which are unrelated to heterotrophic bacteria previously cultivated from these mats.

Metagenomic assembly clusters appear to have coarse taxonomic resolution, lumping many individual phylogenetically related species. For instance, a single cluster is comprised of the many *Synechococcus* species described above. There is evidence that other clusters also contain multiple species, which, like *Synechococcus* spp., are likely adapted to distinct niches. We are using high-throughput sequencing approaches (specifically, Ti454 bar code analyses) to investigate species diversity of all 6 mat phototrophic populations simultaneously, based on variation in their photosystem reaction center genes. These analyses are being conducted on a large number and variety of samples, thus enabling simultaneous deep sampling of genetic diversity (thousands of sequences per sample), prediction of ecological species therefrom and validation of their unique ecological character. We are also in the process of obtaining deeper-coverage metagenomes for the entire mat community that will enable us to better understand the predominant populations involved in mat decomposition and carbon recycling.

This thorough understanding of community composition provides a solid basis for metatranscriptomic and metaproteomic analyses of gene expression and protein synthesis in the mats. The data from these systems-based approaches will also help us begin to understand and model how these species interact in space and time to coordinate community functions.

This research is part of the Genomic Science Program FSFA at Pacific Northwest National Laboratory.

# 198

## Systems Biology of Lichen Systems: Pure Cultures of *Cladonia grayi* and Lichen-Dominated Biological Soil Crusts

PIs: **Daniele Armaleo,**[1] **Francois Lutzoni**,[1] **Frank Collart**,[2] **Scott Baker**[3]* (scott.baker@pnl.gov), and **Jon Magnuson**[3]* (jon.magnuson@pnl.gov)
**Co-authors:** Olafur Andresson,[4] Deanna Auberry,[3] Dave Culley,[3] Fred Dietrich,[1] Igor Grigoriev,[5] Brendan Hodkinson,[1] Steven Karpowicz,[6] Alan Kuo,[5] Peter Larsen,[2] Francis Martin,[7] Tami McDonald,[1] Sabeeha Merchant,[6] Emmanuelle Morin,[7] Olaf Mueller,[1] Ellen Panisko,[3] Maria Virginia Sanchez,[8] Ian Small,[9] and Basil Britto Xavier[4]

[1]Duke University, Durham, N.C.; [2]Argonne National Laboratory, Argonne, Ill.; [3]Pacific Northwest National Laboratory, Richland Wash.; [4]University of Iceland; Reykjavik, Iceland; [5]Joint Genome Institute, Walnut Creek, Calif.; [6]University of California, Los Angeles; [7]Institut National de la Recherche Agronomique, Nancy, France; and [8]Universidad Nacional de Cuyo, Mendoza, Argentina; and [9]The University of Western Australia, Crawley

Lichens arguably represent the most successful symbiotic relationship in nature. A lichen is composed of a heterotrophic fungus with a phototrophic partner; a green alga or a cyanobacterium or sometimes both. This project is a collaboration of Duke University, Argonne National Laboratory (ANL) and Pacific Northwest National Laboratory (PNNL) to explore the systems biology of lichens at two extremes of natural complexity; the partners of one species of lichen, *Cladonia grayi*, and complex biological soil crust (BSC) communities dominated by lichens. Lichens are found in all environments, and wherever they are found they have an important role in the ecosystem. They are sensitive to environmental changes. Therefore, they may be valuable sentinels for assessing the effects of climate change. Though slow growing, they are important in carbon flux in many environments, e.g., in BSCs and tundra. From a fundamental scientific standpoint lichens represent the best and most diverse examples of co-evolved phototrophic-heterotrophic systems.

The principal goal of this project with respect to *Cladonia grayi* is to use systems biology and molecular biology tools to advance fundamental understanding of the lichen symbiosis. Both symbiotic partners of *C. grayi* have been successfully cultured in isolation by the Duke University team members. This has enabled them to sequence the genome of each partner, the fungus *C. grayi*, and the green alga *Asterochloris* sp. Together with JGI and many others the assembly and annotation of the genomes has begun. Furthermore, the ability to culture the partner organisms separately and reestablish the early stages of the symbiotic relationship provides an unprecedented opportunity to study the biology of lichens using modern systems biology tools.

The genomes of the symbiotic partners from the lichen *Cladonia grayi* were assembled from 454 and Solexa data and partially annotated. The data, at JGI (http://genome.jgi-psf.org/Clagr2/Clagr2.home.html), have not yet been released publicly. The 40 Mb genome of the fungal partner (*C. grayi*) includes ~12,000 gene models and the 56 Mb genome of the green algal partner *Asterochloris sp*. has ~9,000 gene models. Signatures of symbiosis are being sought by surveying expansion or contraction of gene families, genes of ancient origin under stabilizing selection, organelle genomes, and gene expression during early development. A peculiar class of ammonium transporters unique to lichens and land plants was identified, as well as a polyketide synthase gene cluster responsible for lichen-specific compounds. The prokaryotic communities associated with *C. grayi* have also been surveyed. Transcriptomic investigations of the various cultured states are underway at ANL to enable the generation of predictive models relevant to complex environmental systems under conditions of climate change. A task that preceded the transcriptomic work was the generation of a map of the JGI annotated enzyme activities for *Cladonia* and *Asterochloris* to the KEGG map01100 (complete metabolism network). ANL has generated a series of interactive maps for the individual organisms as well as combined map that illustrates annotated enzyme classification (EC) activities that are unique to the individual organisms. Proteomics studies of liquid and plate cultures of the individual symbiotic partners from *Cladonia grayi* have begun at PNNL. The initial emphasis has been on the development of methods for cell lysis and proteomic analysis of the two organisms and determining the minimum amount of biomass required for a proteomic sample. The latter point is important in examining the early stages of the symbiosis establishment in the Duke culturing system, where biomass quantities are limited. Analysis of the initial LC-MS proteomics data for the fungal partner is underway and will be presented.

The goals of this project with respect to the BSCs are: to examine the phylogeny of selected lichen-dominated BSC communities obtained from arid land ecosystems; and apply transcriptomics and/or proteomics techniques to investigate these BSCs under different physiological states, e.g., naturally desiccated and rehydrated. BSCs are a remarkable and extremely sensitive type of microbial community important for retention and conditioning of soil and providing an environment where arid land plants can germinate and prosper. They may be an important indicator of climate change and undoubtedly impact carbon flux in arid environments throughout the world. A site for collection of BSCs has been established on Bureau of Land Management land in south-central Washington state. The study site has BSCs of various stages of maturity in a shrub-steppe ecosystem on land with steep topography that has not been used for agriculture or burned over in recent time. Initial studies of the phylogeny of these communities have revealed the genus of two of the major lichens, *Lecanora* and *Caloplaca* spp., in the selected BSCs. Broader studies of the eukaryotic and prokaryotic community will be conducted in the near future.

This research is part of the Genome Sciences Program FSFA at Pacific Northwest National Laboratory.

# 199

## Isotopic Studies of Biological Systems

**James Fredrickson** and **Helen Kreuzer*** (Helen.kreuzer@pnl.gov)

Pacific Northwest National Laboratory

**Project Goals: This project has two parts. The goal of the first part is to use hydrogen stable isotope measurements to elucidate biological hydrogen production pathways. The goal of the second part is to use carbon stable isotope measurements to trace carbon flow through lithoautotrophic acid hot springs microbial communities in Yellowstone National Park.**

### 1. Using H Isotopes to Elucidate Biological Hydrogen Production Pathways

Eric L. Hegg,[1]* H. Yang,[1] N. Ostrom,[1] H. Ghandi,[1] H. Kreuzer,[2] E. Hill,[2] and J. Moran[2]

[1]Michigan State University; and [2]Pacific Northwest National Laboratory

Biological $H_2$ production by hydrogenase ($H_2$ase) enzymes has enormous potential as an environmentally sustainable energy source. $H_2$ases, found throughout nature in many diverse organisms, are among the most efficient $H_2$-producing catalysts known. Although considerable progress has been made in elucidating the metabolic pathways involved in $H_2$ metabolism, one major impediment to improving our understanding of $H_2$ metabolism is our inability to adequately define the regulation of and the flux through key pathways involved in $H_2$ production. **The goal of this project is to develop stable isotopic approaches for improving understanding of biological $H_2$ production.**

We predicted that the isotope ratio of $H_2$ produced by various $H_2$ases would differ because of slight differences in their active sites and proton transfer pathways. We further predicted that we could measure this difference via isotope-ratio mass spectrometry, and that the H/D isotope ratios would allow us to address fundamental questions concerning biological $H_2$ production. To test this predictions, we purified five different $H_2$ases [three [FeFe]-$H_2$ases (*Clostridium pasteurianum, Shewanella oneidensis,* and *Chlamydomonas reinhardtii*) and two [NiFe]-$H_2$ases (*S. oneidensis* and *Desulfovibrio fructosovorans*)] and established conditions that allowed us to quantify the specific activity of the purified $H_2$ases, the amount of $H_2$, and its isotopic content. Using the enzymes and optimized protocols established above, we determined the isotope ratio of the $H_2$ produced by the purified Significantly, the data indicate that all five $H_2$ases produce $H_2$ with a unique isotopic signature, demonstrating that different $H_2$-producing enzymes have different fractionation factors reflected in the isotope ratio of the $H_2$.

*Shewanella oneidensis* MR-1 is a facultative anaerobe capable of transferring electrons to a variety of terminal acceptors including iron, manganese, and other metals. *S. oneidensis* encodes two $H_2$ases, the [FeFe]-$H_2$ase HydA and the [NiFe]-$H_2$ase HyaB. We measured the isotopic content

of $H_2$ produced in the headspace above cultures that had either the [FeFe]-$H_2$ase or the [NiFe]-$H_2$ase deleted, and we found that the isotope ratios closely matched those predicted by the in vitro studies with the purified enzymes. We next monitored both concentration and the isotope ratio of the $H_2$ gas in the headspace of *S. oneidensis* cultures. We grew MR-1 in 69 mL sealed serum bottles containing 20 mL of the defined growth medium M1. We used 60 mM lactate as the electron donor, and oxygen (from air-saturated water) as the electron acceptor. $H_2$ gas production began at the point of electron acceptor limitation in the serum bottles and went through two distinct phases, initial and late production. Although mRNA from both $H_2$ases was present during each phase of $H_2$ production, isotope ratio data indicated that initial $H_2$ production, starting at about 20 hours, was by the HyaB (Ni-Fe) $H_2$ase. This $H_2$ase then consumed $H_2$ from about 50 to 100 hours after inoculation. Late phase $H_2$ production after 100 hours was driven by the HydA $H_2$ase.

**2. Carbon Flow in Lithoautotrophic Acid Hot Springs Microbial Communities, Yellowstone National Park**

Helen Kreuzer,[1] James Moran,[1] Christopher Ehrhardt,[1] and Bill Inskeep[2]

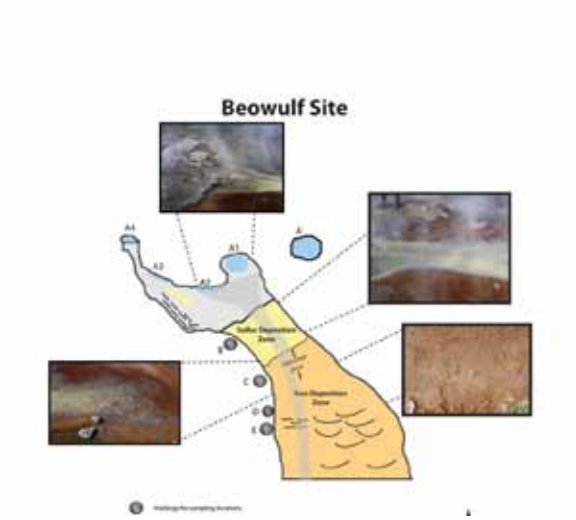
[1]Pacific Northwest National Laboratory; and [2]Montana State University

Beowulf Springs and Obsidian Pool in Yellowstone National Park are acidic hot springs. Beowulf is a sulfidic spring, while Obsidian Pool is not. The microbial communities in the hottest portions of these springs are thought to be litho-autotrophic, but carbon flow in these communities has not been characterized. Geochemical analysis shows that there are several potential sources of C present, including $CO_2$, methane, dissolved inorganic carbon, and dissolved organic carbon. Although $CO_2$ has been assumed to be the ultimate C source, any of these other C-containing materials could potentially be a C source, and the topography of Obsidian Pool suggests it might receive heterotrophic C input from the surrounding landscape. **The goal of this project is to trace C flow through these communities.**

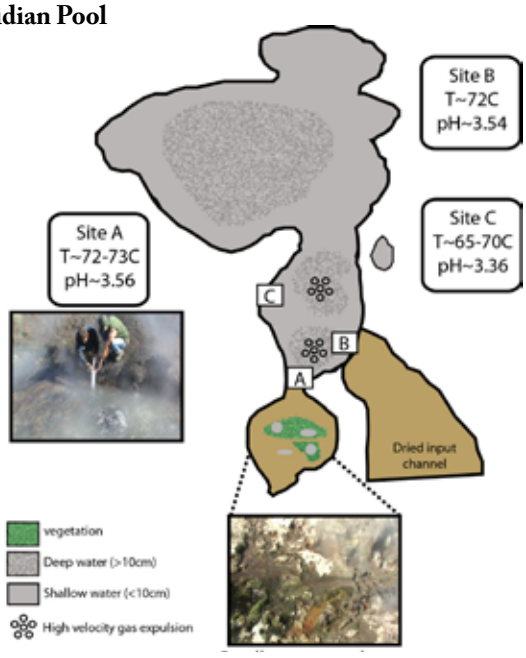
We hypothesize that the C sources will have different C isotope content, enabling an initial tracing of C into the community through isotope profiling. We collected samples from each of the sites (see figure), and are in the process of analyzing the C stable isotopic content of dissolved methane, outgassed $CO_2$, dissolved organic and inorganic carbon, fatty acid methyl esters and archaeal lipids from the in situ microbial communities, as well as potential organic substrates from surrounding environmental inputs. We will use lipid isotope data to associate potential C sources with phylogenetic groups of organisms and base the design of future stable isotope probing experiments on these results. In a collaborative effort (T. Woyke), samples from the same Obsidian Pool sites and one of the Beowulf sites are undergoing cell sorting at the JGI. If successful, we may be generating single-cell sequence data that can be coupled with the isotope data to help better link and resolve carbon flow in these communities.

This research is part of the Genomic Science Program FSFA at Pacific Northwest National Laboratory.

**Beowulf Springs**




**Obsidian Pool**



# 200

## The Development of RNA-Sequencing Technologies for the Analysis of Complex Microbial Communities

Bryan E. Linggi,[1] Meng L. Markillie,[2] Chris S. Oehmen,[3] Ronald C. Taylor,[3] and **H. Steven Wiley**[1]* (steven.wiley@pnl.gov)

[1]Environmental Molecular Sciences Laboratory, [2]Biological Sciences Division, [3]Computational Biosciences, Pacific Northwest National Laboratory, Richland, Wash.

**Project Goals: Develop a suite of technologies to enable the accurate measurement of molecular species (DNA, RNA, protein, metabolites) in microbial communities as well as building the computational infrastructure needed to analyze the resulting data.**

Microbial communities are complex systems containing 10s to 100s of different organisms performing a variety of interrelated functions. To understand and develop ways to utilize these communities for human benefit, technologies are needed to probe their function and accurately measure their composition as well as the biochemical and molecular species that are present. To address this challenge, we are developing a suite of technologies to enable the accurate measurement of molecular species (DNA, RNA, protein, metabolites) in microbial communities as well as building the computational infrastructure needed to analyze the resulting data. As test cases, we are examining both 'synthetic' as well as natural microbial communities.

**Metatranscriptomics**

At the broadest scale, 'omics' experiments, such as genomics, transcriptomics, proteomics and metabolomics, can provide estimates of both the species composition of communities as well as the expression levels of key metabolic pathways. Ideally, the integration of multiple types of 'omics data could provide a comprehensive profile of community structure and function, but there are technical difficulties and tradeoffs with different technologies that complicated direct comparison of analytical results. For example, transcriptional profiling through sequencing-based technologies (RNA-Seq) requires removal of ribosomes to enable adequate sampling of complex microbial communities. In addition, RNA extraction efficiency can vary by organism, which could bias the estimated abundance of certain transcripts. To develop approaches to address these types of complexities, we have created an artificial community comprised of a mixture of 12 different species (12 bacteria, 2 archaea) in which the origin of the different transcripts could be identified though a ligated "barcode" sequence. Isolated RNA from each species was mixed in defined proportions and then sequenced together using the ABI SOLiD4 sequencing platform. We then tested different analytical and computational approaches to determine which combination could most accurately identify the source and abundance of each RNA fragment.

We found that even direct matching of each read to its source genome generates a fraction of unmapped reads. To discriminate between sequences that don't match known genomes because of sequencing errors or inefficient mapping algorithm versus those that arise from an unknown organism, we generated a "gold standard" match for each sequence using ScalaBlast running on the EMSL Chinook supercomputer. While computationally demanding, ScalaBlast provides a much more reliable match relative to RNA-Seq alignment algorithms such as Bowtie, which sacrifice accuracy for speed. A number of other read matching algorithms were also tested to determine their relative accuracy. This allowed us to determine the optimal tools to use for metatranscriptome matching.

Another issue that must be resolved when using RNA-Seq technology for community profiling is the extent of ribosome removal that is necessary. Because of the complexity of microbial communities, extensive sampling is required to ensure unbiased coverage. Unfortunately, most of the RNA in microbes is ribosomal (rRNA), which is usually non-informative with respect to community function. If the ribosomes are not removed, estimates of transcript abundance between different organisms could be erroneous. Removal of the rRNA, however, could also potentially bias mRNA representation. Thus, we benchmarked several different approaches for rRNA removal to determine the extent to which different methods actually alter mRNA composition. Data from both the rRNA removal experiments and sequencing matching trials were used to design optimal strategies for transcriptional profiling of microbial communities.

**Computational Infrastructure**

The acquisition of large amounts of both proteomics and transcriptomics data from microbial communities presents difficulties for data query and analysis. However, effective integration of these types of data could provide a wealth of information on control of protein levels and post-transcriptional regulatory mechanisms. To address this, we recently completed a pilot project for development of a large-scale data warehouse / workspace for the analysis of extremely large data sets. The workspace was constructed as a terabyte-size parallel processing platform based on the Hadoop / MapReduce / HBase framework, implemented on one cluster with a distributed file system. Sample data were generated by performing comprehensive RNA-Seq and proteomics analysis of *Shewanella* grown under aerobic and oxygen-limited conditions in chemostats. The data consisted of over 29,000 different peptides observed in 4 different runs of the two samples. RNA-Seq runs provided transcript data in duplicate from each sample. Following data normalization, we matched RNA data that corresponded to each observed peptide in each sample and calculated a peptide/RNA ratio for each observed peptide. Confidence values and statistics were assigned to each peptide/RNA ratio across duplicates and across conditions. Preliminary analysis of the data suggests a strong correlation between the level of mRNA and most proteins, but that there are several membrane proteins that behave anomalously. We found that several technical issues must be addressed before

peptide/RNA ratios can become a reliable measure of post-transcriptional regulation, such as correcting for the different dynamic ranges of RNA-Seq and proteomics measurements and developing appropriate error models for combined 'omic measurements. Nevertheless, we were successful in creating a robust computational infrastructure for rapidly integrating large amounts of genome-centric data and this should greatly facilitate its exploration and use in defining microbial regulatory processes.

This research is part of the Genomic Science Program FSFA at Pacific Northwest National Laboratory.

# 201

## Information and Informatics Resource for Collaborative Research on Biological Systems Interactions

Gordon Anderson,[1]* William Cannon,[1]* Lee Ann McCue,[1]* Zoë Guillen,[1] Linda Angel,[1] Sebastian Jaramillo Riveri,[1] Margrethe Serres,[2] Margie Romine,[1] and **Jim Fredrickson**[1]

[1]Pacific Northwest National Laboratory, Richland, Wash.; and [2]Marine Biological Laboratory, Woods Hole, Mass.

**Project Goals: As part of the Foundational Scientific Focus Area research program on Biological Systems Interactions at the Pacific Northwest National Laboratory, we have developed an information and informatics resource for collaborative research. The resource provides a computational infrastructure for information dissemination, analysis, discussion and data sharing. The resource currently combines three core technologies to enable collaborative research: semantic wiki technologies, an open source content management-based global file system, and pathway-genome databases and tools. The url for the Biological Systems Interactions research program is microbes.pnl.gov.**

### Wiki-enabled Collaborations

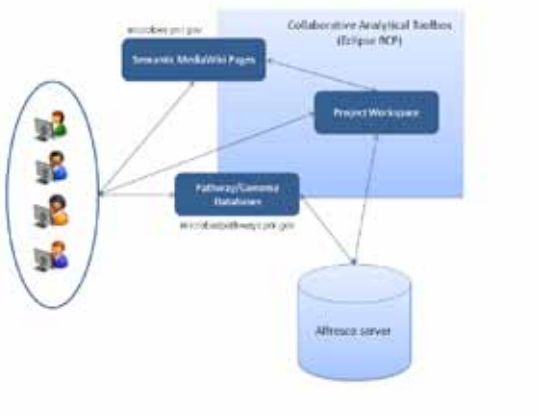
Research projects need the ability to collect and exchange both structured and ad hoc information collectively. We are using semantic wiki technology to allow researchers to create and edit interlinked wiki pages with ad hoc descriptive information about their experiments. A researcher can create a wiki page to provide the background information, research goals, hypotheses, and experimental outcomes of their project. The wiki also makes it easy to show the relationships between the projects in our research program by linking the project description wiki pages for related projects. We are also working to integrate the wiki with our content management-based global file system (CAT), so that there is a direct link between the ad hoc description of an experiment and its associated structured data stored in CAT.

### Collaborative Analytical Toolbox (CAT)

CAT, developed by Pacific Northwest National Laboratory, is a client-server analytic framework for building and organizing a knowledge repository. The CAT environment provides a familiar, customizable interface that enables users to:

- build, organize and share their information;
- view/browse their information in any number of ways and via any number of arbitrary hierarchies;
- browse or search for information from a variety of sources using many different search tools;
- pull data back into their project space;
- integrate with existing tools to analyze their data;
- collaborate with other users by sharing data, templates, annotations, etc.

CAT was designed with flexible integration as a key requirement, so that it is easy for new or existing tools to be integrated with minimal development cost. Specifically, it is based on well-documented, well-supported, robust Eclipse Rich Client Platform (RCP), using the Alfresco content management system on the server side.




The RCP framework is a cross-platform architecture for building and deploying rich client applications. This framework is extendable through plug in applications to extend its functionally and customize it for the application domain. Our developments will extend CAT through the addition of new functionally and interfaces such as:

- Genome and ortholog interactive editor / curation tools
- Pathway and genome databases integration
- Enhanced features – links to KEGG/SEED/UniProt/IMG

Additionally, we are investigating integration of Alfresco content management with MediaWiki to support seamless integration of the information in CAT and the Wiki. This poster will present our current architecture as well as illustrate our vision for information assimilation designed to enable collaborative research across institutional boundaries.

### Pathway and Genome Databases

Integration of the Collaborative Analytical Toolbox with pathway databases will greatly facilitate model building. Refinement of gene annotations is a continual process and takes place both prior to and after automated pathway prediction. However, the refinement of gene annotations

and pathway annotations need to be coordinated, because refinement of gene annotations may result in the need to reevaluate the putative pathway annotations. The coordination between these activities is achieved within CAT. In addition, the initial pathway models must be evaluated with respect to predictions from other sources. Continuous evaluation of the pathway annotations is aided by the construction of a workflow that draws data from multiple sources such as MetaCyc, KEGG, IMG, and SEED, and allows for side-by-side comparisons of pathway models and membership. The revisions to the models are captured in BioCyc pathway-genome databases.

Currently available pathway databases include

- *Chloroflexus aggregans* DSM 9485
- *Chloroflexus* sp. Y-400-fl
- *Roseiflexus castenholzii* DSM 13941
- *Roseiflexus* sp. RS-1
- *Shewanella oneidensis* MR-1
- *Shewanella* sp. W3-18-1
- *Synechococcus* sp. JA-2-3B'a(2-13)
- *Synechococcus* sp JA-3-3Ab
- *Synechococcus* sp. PCC 7002
- *Synechocystis* sp. PCC 6803
- *Thermosynechococcus elongatus* BP-1

Support: This research is part of the GSP FSFA at Pacific Northwest National Laboratory

## Plant-Microbe Interfaces

# 202
## Plant-Microbe Interfaces

**Mitchel J. Doktycz**[1*] (doktyczmj@ornl.gov), Gerald A. Tuskan,[1] Christopher W. Schadt,[1] Gregory B. Hurst,[2] Edward Uberbacher,[1] Dale A. Pelletier,[1] Timothy J. Tschaplinski,[1] Francis Martin,[4] Rytas Vilgalys,[5] Amy Schaefer,[6] Caroline Harwood,[6] Jennifer Morrell-Falvey,[1] David J. Weston,[1] Scott T. Retterer,[1] Andrey Gorin,[3] Yunfeng Yang,[1] Robert Hettich,[2] Udaya C. Kalluri,[1] Xiaohan Yang,[1] Abhijit Karve,[1] Mircea Podar,[1] Steven D. Brown,[1] Robert Cottingham,[1] Tatiana Karpinets,[1] Chongle Pan,[3] Guru Kora,[3] and Susan Holladay[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]Computer Science and Mathematics, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [4]INRA, Nancy, Champenoux, France; [5]Biology Department, Duke University, Durham N.C.; and [6]Department of Microbiology, University of Washington, Seattle

http://PMI.ornl.gov

**Project Goals: The goal of the Plant-Microbe Interfaces science focus area 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 an initial test 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 will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.**

Rapid progress in biological and environmental sciences has been enabled by the availability of genome sequences and the tools and technologies involved in interpreting genome function. As our understanding of biological systems grows, it becomes increasingly clear that the functional expression of individual genomes is affected by an organism's environment and the community of organisms with which it associates. 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. These plant–microbe interactions can benefit plant health and biomass production by affecting nutrient uptake, influencing hormone signaling, effecting water and element cycling in the rhizosphere, or conferring resistance to pathogens. Studying the integral plant–microbe system in native, perennial plant environments, such as *Populus* and its associated microbial community, provides the greatest opportunity for discovering plant–microbial system functions relevant to DOE missions related to bioenergy and carbon-cycle research and understanding of ecosystem processes.

The functional attributes of *Populus* depend on the microbial communities with which it associates. Bacteria and fungi can be found within *Populus* tissues and closely associated with the roots in the rhizosphere. Understanding these communities, and the interfaces between organisms, is critical to realizing fundamental scientific knowledge that may enable increased plant productivity, ecosystem sustainability, disease resistance, drought tolerance, and ecosystem carbon budgets. This interface can also influence the processes, or mechanisms, by which adaptive traits arise from genetic variation and community function. Microbial rhizosphere structure, plant root bacterial and fungal colonization patterns, and the microbe–plant signaling pathways inherent in each type of association are all found within *Populus* and can be functionally translated hierarchically across scales into ecosystem patterns and processes.

Understanding the mechanisms by which plants and microbes interact represents a grand challenge facing biological and environmental science. How microbial selection and colonization occurs, what reciprocal benefits are bestowed upon the plant and microbe, and how these interactions ultimately affect, and are affected by, the environment are just some of the intrinsic scientific questions. The multiple spatial and temporal scales involved in these interfaces, the complexity of the component systems, and the need for better tools that use and build upon growing genomics resources to probe and interpret these combined systems represent some of the essential technical challenges. An overview of the research being carried out in the ORNL Plant Microbe Interfaces science focus area will be presented.

# 203

## Plant-Microbe Interfaces: Proteomics Studies of Plant-Microbe Interfaces

Karuna Chourey[1] * (choureyk@ornl.gov), Xiaohan Yang,[2] Ting Li,[2] Chongle Pan,[1,3]Zhou Li,[4] Rachel M. Adams,[4] Robert L. Hettich,[1] Patricia K. Lankford,[2] Keiji G. Asano,[1] Andrey A. Gorin,[3] Udaya C. Kalluri,[2] Poornima Sukumar,[2] David J. Weston,[2] Sara M. Allen,[2] Dale A. Pelletier,[2] Timothy J. Tschaplinski,[2] **Gerald A. Tuskan**,[2] **Mitchel J. Doktycz**,[2] and **Gregory B. Hurst**[1]

[1]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]BioSciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]Computer Science and Mathematics, Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [4]UT-ORNL Graduate School of Genome Science and Technology, Knoxville, Tenn.

http://PMI.ornl.gov

**Project Goals: The goal of the Plant-Microbe Interfaces science focus area 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 an initial test 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 will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.**

Characterization of proteomes of organisms and communities has the potential to contribute substantially to our understanding of plant-microbe interfaces, in particular the interactions among *Populus* and its microbiome. To realize this potential, a number of challenges posed by these complex systems must be overcome, requiring the development of tools and protocols that will extend the state of the art in proteomics. These challenges include (1) the complexity of microbial communities, and the lack of readily available metagenomes (2) partial protein extraction from plant and fungal tissues, often impeded by interfering small compounds, and protein degradation by endogenous enzymes, (3) collection of extracellular proteins that may play roles in signaling, (4) informatics associated with proteomics of eukaryotic organisms (plant, fungal) for which genomes reflect gene duplication events, alternate splicing, etc. A major future challenge for which we are currently designing approaches is the metaproteomics of *Populus* root microbial communities, including rhizosphere and endophyte components. To address these challenges, we have initiated proteomics studies in several critical areas for the Plant-Microbe Interfaces (PMI)-SFA:

-Analysis of small proteins encoded by small genes in bacterial and plant species To complement work pioneered by Xiaohan Yang to identify small-protein-encoding genes (see poster by Li et al.), we are implementing size fractionation to enrich small proteins from plant tissues in order to increase the sensitivity of LC-MS-MS analysis toward these analytes. Proteins identified by LC-MS-MS from several fractionation methods exhibited molecular mass distributions with medians significantly lower than those of unfractionated lysates. A number of the identified proteins were unique to small protein enrichment fractions. Among these proteins are several with annotations indicating that their functions are not yet characterized. Evidence for expression of these proteins supports improved annotation of the corresponding small genes, and provides candidates for further studies of biological function.

-Proteomics studies of roots from field-sampled mature *Populus* trees The ability to study the root proteome of plants is the first step towards unraveling the plant-microbe interactions in the rhizosphere. We evaluated different proteome extraction methods on root tips harvested from subsurface roots of naturally occurring *Populus* trees (Clinch River, East TN). Extraction of proteins via acetone precipitation worked best, and was employed towards proteomic profiling of a subset of *Populus* root samples obtained during PMI sampling trips to the Yadkin River in N. Carolina. Raw tandem mass spectra were searched against predicted proteins from genome sequences of *Populus trichocarpa* and *Laccaria bicolor*. Preliminary results show reproducible identification of >1000 proteins from each root sample, despite heterogeneity in morphology, differences in location, soil type, etc. Distributions of plant proteins across KOG (euKaryotic Orthologous Groups) functional categories were comparable across samples, suggesting similar protein metabolism in roots of sampled mature trees. Identified proteins occupied a variety of cellular locations (predicted by WolfPsort), proving the method to be efficient at extracting proteins in an unbiased manner. To date, >500 proteins were detected across all samples, forming an initial 'core' root proteome for natural *Populus* roots. Prominent in the core

proteome were pectin methylesterases, major latex protein, peroxidases, enolase, glutathione S-transferase, alcohol dehydrogenase, ubiquitin, actin and histones. Relatively few *Laccaria* proteins were identified, with strong representation by histones and ubiquitin. Future studies on additional archived roots will provide information on whether observed proteome differences can be correlated with tree location, age, soil conditions, or other data acquired during sampling.

-Characterization of proteome changes resulting from RNAi knockdown of an auxin-related gene in *Populus* *We* have interrogated the root and shoot proteomes of wild-type (WT) *Populus*, and a strain with altered auxin signaling, IAA7.1 1-3 (see poster by Sukumar *et al.*). Initial measurements indicate that of >3900 proteins detected in shoots, 20 (28) were more (less) abundant in the IAA7.1 1-3 strain than in the WT. In roots, from >5500 proteins identified, 24 (1) proteins were more (less) abundant in the IAA7.1 1-3 strain. These measurements will potentially provide additional information on metabolic pathways affected by auxin signaling.

-Stable isotope labeling for comparisons via quantitative proteomics among multiple treatments, experimental conditions or time points Various quantitative proteomics methods have been developed, each with unique advantages. Label free quantification allows simultaneous protein identification and quantification without laboriously incorporating costly isotopes into samples. Metabolic labeling minimizes variability in sample preparation and measurement. Chemical labeling via iTRAQ/TMT permits multiplexed quantification. A comparison among these three techniques, using the latest-generation high performance mass spectrometer (LTQ-Orbitrap-Velos) provided guidance on selecting the most appropriate method for a proteomics study. The results indicate that iTRAQ/TMT chemical labeling has the highest quantification precision, label-free quantification provides the largest number of protein identifications, and metabolic labeling is intermediate in both measures. We have initiated experiments using iTRAQ/TMT quantification on experimental systems relevant to the PMI-SFA, including analysis of laboratory-grown *Populus* plants with and without inoculation of the roots with *L. bicolor*.

# 204

## Plant-Microbe Interfaces: Isolation and Functional Characterization of Cultivatable Bacteria from the *Populus* Rhizo-Endosphere

**Dale A. Pelletier*** (pelletierda@ornl.gov), Tse-Yuan S Lu, Christopher W. Schadt, Neil Gottel, Matthew Foster, Marilyn Kerley, Patricia K. Lankford, and Mitchel J. Doktycz

Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

**Project Goals: 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 an initial test 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 will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.**

*Populus* trees are host to a variety of mutualistic microorganisms within their endosphere and rhizosphere that can have positive effects on the host. How these interactions manifest themselves within the spatially, structurally, and temporally complex scales of natural ecosystems is an open question. Our goal is to understand the diversity of the *Populus* microbiome and to elucidate the metabolic and molecular mechanisms responsible for shaping the *Populus*-microbial interface. In order to better understand the microbial communities associated with native *Populus deltoides* (Eastern cottonwood) we have under taken both cultivation independent and cultivation dependent assessment of *P. deltoides* rhizosphere and endosphere microbial communities. We have sampled *P. deltoides* at sites along the Caney Fork River in central Tennessee and Yadkin River in North Carolina. These sites represent ecotypes and soil conditions that are common to this region. We have sampled these sites in spring and fall to investigate seasonal changes in communities. This poster will focus on the microbial communities that were isolated from both rhizo- and endosphere sample utilizing direct plating methods. Isolated strains where identified by 16S rDNA sequence analysis and traits of interest in plant microbe interactions were investigated including plant colonization, both microbial and plant phenotyping and physiological properties.

A diverse array of bacterial strains (>1000) has been isolated from the rhizo- and endosphere of *Populus* roots. The isolates comprise 7 classes and 85 genera of bacteria including, Actinobacteria (14%), Bacilli (17%), Flavobacteria (6%), Sphingobacteria (3%), and a- (22%) b- (16%) and g- (22%) proteobacteria. Some general conclusions from isolation experiments are that native *Populus deltoides* roots are colonized by a diverse community of cultivable bacteria. The rhizosphere isolates are dominated by Actinobacteria and Bacilli strains, mainly *Streptomyces* and *Bacillus* species, the endophytes are dominated by a-proteobacteria while g-proteobacteria are prevalent in both environments. While we have cultivated a diverse array of organisms and our results are in general agreement with 454 16S pyrotyping results, the data suggests that specific groups of microbes may be especially underrepresented within the culture-based collection (e.g. *Flexibacter/Cytophaga/Bacteroides*, *Planctomycetales*, and *Acidobacteria*). A number of our isolates are

*Pseudomonas* species. This group of bacteria is known to have considerable genetic and phenotypic variability and is a common biocontrol and plant growth promoting bacteria, therefore more extensive characterization of 84 of these strains was performed. Full length 16S rDNA, rpoD and gyrB sequences indicate genotypic diversity of these isolates and phenotypic variability was found in a number of rhizosphere related traits including siderophore (95%), protease (51%), indole-3-acetic acid (50%) and phosphate solubilizing (41%) activities. Additionally root colonization of *Arabidopsis* and *Populus* identified plant growth promotion and modified root architecture phenotypes. To identify genes potentially responsible for plant phenotypes and rhizosphere colonization traits and comparative genome analysis we are sequencing the genomes of a number of isolates. A subset of these isolates are undergoing more extensive genomic and physiological analysis as well as plant-microbe co-culture experiments where plant physiology, transcriptome and metabolome are assayed which will lead to better understanding of the molecular mechanisms responsible for these interactions.

# 205

## Plant-Microbe Interfaces: Bacterial and Fungal Communities Within the Roots and Rhizosphere of *Populus deltoides* in Upland and Lowland Soils

Neil Gottel,[1] Hector F. Castro Gonzalez,[1] M. Kerley, Zamin K. Yang,[1] Wellington Muchero,[1] Jessy L. Labbé,[1] Gregory Bonito,[2] Migun Shakya,[1,3] Dale A. Pelletier,[1,3] Mircea Podar,[1,3] Tatiana Karpinets,[1] Edward C. Uberbacher,[1,3] Gerald A. Tuskan,[1] Rytas Vilgalys,[2] Mitchel J. Doktycz,[1,3] and **Christopher W. Schadt**[1,3]* (schadtcw@ornl.gov)

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Biology Department, Duke University, Durham, N.C.; and [3]UT-ORNL Graduate School of Genome Science and Technology, Knoxville, Tenn.

http://PMI.ornl.gov

**Project Goals: The goal of the Plant-Microbe Interfaces science focus area 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 an initial test 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 will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.**

*Populus* trees are a genetically diverse and ecologically widespread riparian species, a potential cellulosic feedstock for biofuels, the first woody plant species to have a genome sequence, and are host to a wide variety of symbiotic microbial associations within their roots and rhizosphere. Thus they serve as an ideal model to study interactions between plants and microorganisms. However, most of our knowledge of microbial associations to date comes from greenhouse and young plantation-based trees; there have been no published efforts to comprehensively describe microbial communities of mature natural communities of *Populus*.

We compared root endophyte and rhizosphere samples collected at upland and lowland sites in Tennessee, to begin to understand the variation that might exist within and between soil types of these communities in both bacterial and fungal populations. 454 pyrosequencing was used to survey the microbial community of *P. deltoides*, using primers targeting the V4 region of the bacterial 16S rRNA gene and the D1 region of the fungal 28S rRNA gene. Further, genetic relatedness among the *Populus* trees was evaluated using 20 SSR markers chosen for distribution across all 19 linkage groups of the *Populus* genetic map. Jaccard's similarity coefficients were calculated based on segregation data generated from SSR marker-based assay. Soil physical, chemical and nutrient status, as well as tree growth and age characteristics were also evaluated.

121,540 bacterial and 322,100 fungal sequences were obtained, representing profiles of 20 endophyte and 20 corresponding rhizosphere samples. Bacterial rhizosphere communities were dominated by Acidobacteria (31%) and α-Proteobacteria (30%). Endophytic samples retained a lower proportion of α-Proteobacteria (23%), and were dominated by γ-Proteobacteria (54%). The fungal rhizosphere and endophyte samples all had equal amounts of the Pezizomycotina (40%), but differences were seen in the Agaricomycotina, which were more dominant in the rhizosphere (34%) than the roots (17%) and more specifically some endophytic root samples had, a large populations of a specific member of the Pucciniomycotina similar to known non-pathogenic basidiomycetous yeasts, however colonization of these OTUs was highly variable. Endophytic bacterial richness was also more variable and on average tenfold lower than the rhizosphere samples, suggesting root tissues provide a distinct environment supporting relatively few microbial types and that colonization events may be sporadic. Both fungal and bacterial rhizosphere samples showed distinct phylogenetic composition patterns compared to endophyte samples using UNIFRAC-PCoA analysis. PCoA analysis did not reveal changes in microbial communities between upland and lowland soil types in either rhizosphere or endophyte samples. Similarly, there was no strict adherence to soil type with regards to genetic similarity of sampled trees. In general, diversity among sampled *Populus* trees was lower among lowland genotypes with Jaccard's coefficients

ranging from 0.02 - 0.08 compared to 0.08 - 0.13 for upland genotypes.

These findings indicate that the plant characteristics that influence the *Populus* root environment may represent a relatively stronger selective force than the soil environment in shaping the endophyte and rhizosphere microbial communities. However additional studies that carefully examine the variation in host genotype/phenotype vs. environmental effects will be required to fully describe these influences.

# 206

## Plant-Microbe Interfaces: Acyl-homoserine Lactone Quorum Sensing in *Populus* Bacterial Communities

**Amy L. Schaefer**[1]* (amyschae@uw.edu), Joshua F. Emory,[3] **Dale A. Pelletier**,[2] Ryan Morlen,[1] Colin R. Lappala,[1] Bruce A. Tompkins,[3] Gary J. van Berkel,[3] E. Peter Greenberg,[1] and **Caroline S. Harwood**[1]

[1]Department of Microbiology, University of Washington, Seattle; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [3]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

**Project Goals: As part of our goal to understand the molecular and cellular events involved in establishing and maintaining beneficial interactions between *Populus* and its associated microbial communities, we are examining the role acyl-homoserine lactone signaling plays in this interaction. The long term project goals are to elucidate the AHL signal inventory of the *Populus*-associated bacterial community, define what these signals control in the bacteria that synthesize them as well as any effect AHLs may have on the plant host, as well as investigate how disruption of AHL signaling (either by genetic mutation or by addition of AHL-degrading enzymes) influences bacterial-plant interactions.**

As part of the newly initiated ORNL Plant-Microbe Interfaces Science Focus area, we are characterizing the natural diversity of microbial associates of *Populus* and elucidating the molecular mechanisms by which these organisms interact. We sampled a population of *P. deltoides* as it occurs along the Caney Fork River in Tennessee in 2009. Analysis of 16s rDNA sequences indicates the *Populus* bacterial communities are dominated by Acidobacteria, Alphaproteobacteria, and Gammaproteobacteria (see poster by Gottel el al.) and the Proteobacteria are the predominant group isolated from *Populus* endophyte (86% of isolates, n=105) and rhizosphere (49% of isolates, n=157) samples (see poster by Gottel et al.). Many Proteobacteria use acyl-homoserine lactone (AHL) signals for cell density-dependent gene regulation, in a process known as quorum sensing and response. LuxI-type proteins synthesize small, diffusible AHL signals that function with LuxR-type signal receptors to control gene expression. Most known AHLs possess a fatty acyl side chain, derived from fatty acid biosynthesis, of varying side chain length and substitution. Recently we discovered that *Rhodopseudomonas palustris* makes a novel aryl signal, *p*-coumaroyl-HSL, which derives its side chain from an exogenously provided plant metabolite. This suggests that there may be additional novel HSL-type signals made by bacteria.

We screened 130 Proteobacteria isolated from *P. deltoides* for AHL production and found >80% Alphaproteobacteria and >20% of the Gammaproteobacteria isolates to be positive. This demonstrates that AHL signaling is prevalent in *Populus* microbial communities. We have also screened isolates for AHL production that is dependent upon exogenous addition of plant-derived aromatic compounds and found that *Enterobacter sp.* GM1 synthesizes cinnamoyl-HSL. This is the second aryl-HSL compound described to date and the first Gammaproteobacteria known to produce an aryl-HSL. We are sequencing the genome of GM1 to identify the cinnamoyl-HSL synthase and receptor genes. To enable a high-throughput means of detecting and identifying AHLs from culture supernatants, we developed a protocol that uses ultra high performance liquid chromatography (UPLC) coupled with multiple reaction monitoring (MRM) mass spectrometric detection. Using UPLC-MRM we can separate 21 synthetic AHL compounds, including the newly described aryl-HSLs, within four minutes and detect them at low concentrations (<0.2 pmol). AHL quorum sensing often controls the production of "public goods" such as antimicrobials and exoenzymes, as well as aggregation factors and conjugal transfer processes. In order to define the AHL regulon of a particular bacterium, mutants in either the *luxI*- or *luxR*-type genes are often constructed and analyzed relative to wild-type. However, not all AHL-producing bacteria are genetically tractable. To examine AHL-regulons in bacteria without constructing AHL-mutants we have demonstrated that purified AiiA lactonase, an enzyme that hydrolyzes the HSL ring of AHL signals, can be added to bacterial cultures to inhibit AHL-regulated phenotypes and gene expression. This protocol should enable future studies to define the AHL-regulons of *Populus*-associated bacteria.

# 207

## Plant-Microbe Interfaces: The Effect of Host Species, Genotype, and Edaphic Factors on Rhizosphere Fungi Associated with *Populus deltoides*

Gregory Bonito[1]* (gmb2@duke.edu), Hannah Reynolds,[1] **Christopher W. Schadt**,[2] Jessy Labbé,[2] **Mitchel J. Doktycz**,[2] **Gerald A. Tuskan**,[2] and **Rytas Vilgalys**[1]

[1]Biology Department, Duke University, Durham, N.C.; and [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

**Project Goals: The goal of the Plant-Microbe Interfaces science focus area 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 an initial test 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 will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.**

*Populus deltoides* is a common riparian tree species in areas of southeastern North America where regular flooding occurs. This species is reported to form both arbuscular and ecto- mycorrhizas in addition to harboring root endophytes. However, there is little understanding on the influence of edaphic or genotypic factors on the structuring of these rhizosphere fungal assemblages. To address the influences of genetic and soil factors on microbial root communities associated with *Populus* we designed a series of bioassay experiments using rooted cuttings to assess the diversity of fungi from different soils in an experimental greenhouse/growth chamber environment.

Specific objectives of this research were to:

1. Determine the influence of host species (*P. deltoides*, *Q. alba*, *P. taeda*) on the structuring of mycorrhizal communities
2. Determine the influence of *P. deltoides* genotype on the structuring of mycorrhizal communities
3. Compare the inoculum potential and influence of different soils on bacterial and fungal rhizosphere communities associated with a single *P. deltoides* genotype

Field soils from our ORNL *P. deltoides* research sites were used as the source of microbial inoculum in these bioassay studies. Cuttings from different *P. deltoides* genotypes were planted into a mixture of sterile sand and potting media amended with different field soils. Upon rooting, the cuttings form associations with fungi from different field soils, including mycorrhizal fungi. Experimental treatments included multiple field soils (inoculum) inoculated onto multiple *P. deltoides* clones. We also included a *P.deltoides x P.trichocarpa* hybrid, oak (*Quercus alba*) and pine (*Pinus taeda*) as positive controls and alternative hosts, as well as negative controls (no soil additions) in our experimental design. The plants were harvested after a five-month growing period. Soils were washed off the root systems, roots were visually assessed for ectomycorrhizas, and samples of bulk roots (representative of the whole root system) were taken from each plant for DNA extraction. Data on plant survivorship, number of shoots and shoot height was also recorded. Bulk root samples were then freeze-dried, pulverized, and DNA was extracted from them using a modified CTAB-chloroform extraction protocol. A number of different primer sets were tested and a subset of these was selected to amplify targeted microbial groups for pyrosequencing. The fungal community from each root will be sampled for both Fungi were ITS and LSU rDNA regions using the fungal specific primers ITS1f & ITS4 and LROR & LR3. Arbuscular mycorrhizae were amplified selectively using the primer set AML1 & AML2. Bacterial 16S rDNA primers that discriminate against plastid DNA were used to compare rhizosphere bacterial communities in selected samples. We made clone libraries from amplicon pools using these primer sets. From these sequences we have assessed these primer sets are efficient at amplifying the targeted groups. From our visual assessments, pine and oak seedlings had high ectomycorrhizal colonization (>80%) while most of the *P. deltoides* genotypes appeared to have generally low ectomycorrhizal colonization (<10%). One exception was the D124 genotype, which had the highest rates of ectomycorrhizal colonization of any of the *Populus* genotypes tested. Ectomycorrhizal species in the genus *Tuber*, *Hebeloma*, and *Laccaria*, were recovered from both oak and *Populus* roots. Further, one of the species of *Tuber* recovered from *Populus* roots in our bioassay was identical to the sequence for the only ectomycorrhizal fruitbodies (of truffles) that have found fruiting in our *Populus* field plots so far. We have collected this undescribed truffle species (*Tuber*) under *P. deltoides* at our ORNL sites in Tennessee and North Carolina. From our arbuscular mycorrhizae clone libraries we recovered two species of *Paraglomus*, one species of *Gomus*, and a novel species that falls in-between the currently known families within the Glomeromycota. We are now in the process of single direction pyrosequencing using 454 Titanium Lib L chemistry. Results from these experiments will be presented.

# 208

## Plant-Microbe Interfaces: Identification of Quantitative Trait Loci and Targeting of Genes Affecting Ectomycorrhizal Symbiosis in *Populus*

Jessy Labbé[1]* (labbejj@ornl.gov), Wellington Muchero,[1] Lee E. Gunter,[1] Annegret Kohler,[2] Véronique Jorge,[3] Catherine Bastien,[3] Francis Martin,[2] François Le Tacon,[2] **Gerald A. Tuskan**,[1] and **Mitchel J. Doktycz**[1]

[1]BioSciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]INRA, Nancy, Champenoux, France; and [3]INRA, Orléans, France

http://PMI.ornl.gov

**Project Goals: The goal of the Plant-Microbe Interfaces science focus area 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 an initial test 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 will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.**

The tree rhizosphere hosts a large community of microbes that compete and interact with each other and with plant roots. Within this community of microorganisms, ectomycorrhizal fungi are almost ubiquitous. Mycelium of symbiotic fungi and root tips form a novel composite organ, so-called ectomycorrhiza, which is the site of nutrient transfer between the symbionts. Because genome sequence is available for both *Populus trichocarpa* and the basidiomycete *Laccaria bicolor*, the *Populus-Laccaria* symbionts are an excellent system to study ectomycorrhizal interactions.

We have analyzed a *Populus deltoides* × *P. trichocarpa* $F_1$ pedigree (Family 54B, INRA-Orléans, France) for quantitative trait loci (QTLs) affecting ectomycorrhizal development and for microarray characterization of gene networks involved in this symbiosis. A 300 genotype progeny set was evaluated for its ability to form ectomycorrhiza with *L. bicolor*. The percentage of mycorrhizal root tips was determined on the root systems of all 300 progeny and their two parents. QTL analysis identified four significant QTLs, one on the *P. deltoides* and three on the *P. trichocarpa* unsaturated genetic maps (Jorge et al. 2005). These QTLs were aligned to the *P. trichocarpa* genome and each contained several megabases and encompass numerous genes. Using cDNA from RNA extracts of ectomycorrhizal root tips from the parental genotypes *P. trichocarpa* and *P. deltoides*, expression analysis from a NimbleGen whole-genome microarray, was used to narrow the candidate gene list. About 3.4% of the *Populus* gene models were differentially expressed (1,543 genes; p-value≤0.05; ≥5.0-fold change in transcript level) in mycorrhiza of the two parents including genes coding for the lignin metabolism and the NBS-LRR class of disease resistance proteins. Forty-one transcripts were located in the QTL intervals. Among these 41 transcripts, 25 were overrepresented in *P. deltoides* relative to *P. trichocarpa*; 16 were overrepresented in *P. trichocarpa*. The transcript showing the highest overrepresentation in *P. trichocarpa* mycorrhiza libraries compared to *P. deltoides* mycorrhiza codes for an ethylene-sensitive EREBP-4 protein that may repress defense mechanisms in *P. trichocarpa* while the highest overrepresented transcripts in *P. deltoides* code for proteins/genes typically associated with pathogen resistance. Recently we genotyped 300 mapping progeny on a 6K Illumina *Populus* SNP array to improve the genetic maps and increase accuracy of gene targeting. Finally, these results suggest that there is a shared molecular communication network between these two organisms and that modification of metabolic pathways may be occurring before, during and after colonization.

# 209

## Plant-Microbe Interfaces: The Discovery of Novel Secretory Motifs Modulating Plant-Microbe Interactions

**Ting Li**[1]* (lit1@ornl.gov), Jennifer L. Morrell-Falvey,[1] Gregory B. Hurst,[2] Timothy J. Tschaplinski,[1] **Gerald A. Tuskan**,[1] and **Xiaohan Yang**[1]

[1]Biosciences Division and [2]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

**Project Goals: The goal of the Plant-Microbe Interfaces science focus area 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 an initial test 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 will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.**

Recent data has shown that proteins less than 200 amino acids in length that are encoded in short open reading

frames (i.e., small proteins) can modulate diverse biological processes, including signal transduction between plants and their bacterial or fungal associates. The goal of this effort is to systematically investigate the functional genomics of the small signaling proteins mediating plant-microbe interactions.

It is widely accepted that small proteins play important roles in plant growth and development, such as transcriptional regulation, signal transduction, stress response and defense response. Transcriptomic analyses in *Populus* revealed thousands of short open reading frames expressed under normal and drought conditions and putative small signaling proteins were identified by additional comparative genomics analysis. Despite these efforts, the prediction and annotation of small proteins remain challenging. We report here a computational approach to predict small signaling proteins mediating plant-microbe interactions using protein signatures.

We hypothesize that novel conserved protein domains/motifs are signatures representing the functions of small proteins. A large-scale analysis of the conserved domains in small proteins from five plant species, including *Populus trichocarpa*, *Vitis vinifera, Arabidopsis thaliana, Cucumis sativus* and *Glycine max*, was performed. We first identified conserved protein domains using sequence-based probabilistic models, then known protein motifs were removed by querying 14 current protein domain databases. Our analysis identified 732 motifs that are not documented in the public protein domain databases. Most of these novel motifs are over-represented in small proteins and a larger portion of these motifs are located in the N- or C-terminus of the small protein sequences compared to known motifs. In addition, we found a distinctive expression pattern for the small proteins containing novel motifs as compared with those containing known domains. A significantly higher percentage of the small proteins containing the novel motifs, relative to those of the known domains, were predicted to locate in the extracellular space, suggesting that some of these novel protein motifs may be signatures for protein secretion or intercellular signaling. Computational and experimental characterizations are underway to determine the potential functions of these novel protein motifs in plant-microbe interactions. The novel motifs uncovered in this research will facilitate the genome-wide discovery of small proteins functioning in intercellular signaling in plant species.

# 210

## Plant-Microbe Interfaces: Transcript and Metabolic Networks Underlying Induced Systemic Resistance in *Arabidopsis* Co-cultured with *Pseudomonas* Strains

Abhijit A. Karve[1]* (karveaa@ornl.gov), Sara M. Allen,[1] Sara S. Jawdy,[1] Lee E. Gunter,[1] Nancy L. Engle,[1] Timothy J. Tschaplinski,[1] Tse-Yuan S. Lu,[1] Dale A. Pelletier,[1] Jennifer L. Morrell-Falvey,[1] Jay Chen,[1] Andrey A. Gorin,[2] Nikita D. Arnold,[2] Tamah Fridman,[2] Gerald A. Tuskan,[1] Mitchel J. Doktycz,[1] and **David J. Weston**[1]

[1]Biosciences Division and [2]Computer Science and Mathematics, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

**Project Goals: The goal of the Plant-Microbe Interfaces science focus area 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 an initial test 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 will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.**

In addition to providing support, nutrients and water, plant roots also act as communication conduits with soil microflora. Such plant-microbe interactions can elicit an array of beneficial or unfavorable phenotypes via systemic signaling. In the case of rhizobacteria including some *Pseudomonas* strains, plant growth is promoted by the suppression of diseases and insect herbivory. This phenotype termed as induced systemic resistance (ISR) is mediated through complex metabolic and hormonal networks. This is phenotypically similar to the pathogen-induced systemic acquired resistance (SAR) except that SAR is dependent on the accumulation of salicylic acid (SA) and pathogenesis related (PR) proteins. Alternatively, ISR relies on jasmonic acid (JA) and ethylene signaling in *Arabidopsis*. Our recent discovery of a putative *Pseudomonas* strain (GM-30) from the rhizosphere of *Populus deltoides* brings to question whether this strain elicits a novel beneficial, neutral or antagonistic plant phenotype. Here, we report the initial characterization of the plant systemic response induced from co-cultures with *Pseudomonas* strains GM-30 and Pf-5. Root colonization of *Arabidopsis* by both Pf-5 and GM30 promoted plant growth and modified plant root architecture. Transcript

profiles were collected from root and shoot tissue on days 3 and 10 after inoculation. Ontology analysis of microarray data suggests that jasmonic acid and ethylene biosynthesis were induced in shoots of plants co-cultured with Pf-5 relative to those co-cultured with GM30. Furthermore, plants co-cultured with GM30 showed significant enrichment (Wilcoxon rank sum, p = 0.007) for genes encoding PR proteins. These results suggest that both *Pseudomonas* strains trigger plant systemic defense but through alternate pathways. To confirm that both *Pseudomonas* strains colonize the plant roots, fluorescence in situ hybridization (FISH) was performed using a probe specific for γ-Proteobacteria (GAM42a-Alexa488) and the universal bacterial probe EUB338 (EUB388-Alexa594). These data indicate that the bacteria were physically associated with and formed colonies on the roots. To validate our array results, a qPCR pathway index populated with 42 published marker genes associate with both beneficial and antagonist plant-microbe interactions was created using a high throughput qPCR platform. Results from this analysis further suggest that Pf-5 elicits ISR through JA and ethylene signaling, while GM-30 altered the expression of genes associated with SA biosynthesis and PR genes. Metabolite profiles of plant shoots at day 3 after inoculation found that Pf-5 co-cultures induced methionine, a precursor for ethylene signaling relative to GM-30 co-cultures. Shoots from plants co-cultured with GM-30 had higher levels of phenylalanine and tryptophan, which are precursors for SA and auxin biosynthesis, respectively. Taken together, these results suggest that Pf-5 and GM-30 play a role in ISR and SAR. This hypothesis is currently being tested with a *Pseudomonas syringe* challenge on *Arabidopsis* seedlings co-cultured with either Pf-5 or GM-30. Current research is being conducted to compare the *Arabidopsis* and *Populus* systemic responses when co-cultured with Pf-5 and GM-30 by using comparative network analyses for genes, metabolites and proteins.

Protein network comparisons require comprehensive characterization of the proteomes in the co-cultivated and control samples, as incomplete or false protein identifications are recognized as the source of incorrect conclusions about induced biological changes. We are developing novel mathematical and computational algorithms for MS/MS-based proteomics as a promising way to increase precision and improve reproducibility of proteome profiles. So far new approaches were developed and tested on two key stages of the pipeline that transforms raw MS/MS spectra into protein ids: (1) assignment of MS/MS spectra to peptides; and (2) conversion of reliably assigned peptides to proteins.

At the peptide identification stage rigorous analytical formulas were used to estimate confidence of each assignment rather than usual empirical or semi-empirical generic cutoff approaches. The established confidence values were integrated for calculations of protein reliability by Bayesian statistics. In the preliminary trials we observed an increase of 2-3 times in the number of useable peptides and 1.25-1.5 boosts in the number of unique protein ids. The ongoing experiments will evaluate effects of these advances on the reproducibility and dynamic range of proteome measurements.

# 211

## Plant-Microbe Interfaces: Differential Involvement of Auxin Signaling Components in Modulating Root Responses to Various Rhizosphere Microbes

Poornima Sukumar[1]* (sukumarp@ornl.gov), Alyssa Delong,[2] Whitney McNutt,[1] Jennifer Morrell-Falvey,[1] David J. Weston,[1] Dale A. Pelletier,[1] Karuna Chourey,[1] Gregory B. Hurst,[1] Valérie Legué,[3] Annegret Kohler,[3] Francis Martin,[3] Mitchel J. Doktycz,[1] Gerald A. Tuskan,[1] and **Udaya C. Kalluri**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Purdue University, West Lafayette, Ind.; and [3]INRA, Nancy, Champenoux, France

http://PMI.ornl.gov

**Project Goals: The goal of the Plant-Microbe Interfaces science focus area 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 an initial test 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 will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.**

Understanding the mechanism of plant-microbe interactions is pertinent as microbes can influence plant growth, positively or negatively. Many mechanisms of interaction have been postulated including the involvement of phytohormones such as auxin. Several microbes in the rhizosphere have been shown to produce auxin and have been shown to affect root architecture including changes in primary and lateral root length, number of laterals, production of tertiary roots and nodulation. This study looks at the influence of three different microbes; *Piriformospora indica*, *Laccaria bicolor* strain S238N and a *Pseudomonas* strain GM30 on root architecture of *Arabidopsis* and *Populus*. 5-day-old *Arabidopsis* seedlings were co-cultured with the listed microbes and the root architectural modifications were examined 7 days later. We find that the 3 organisms used in this study have different affects on plant roots. While *P. indica* enhanced the number of secondary roots that emerge, *L. bicolor* co-culture produced longer laterals and GM30 enhanced the density of lateral root formation. In *Populus* tissue culture plants, there was an increase in the development of secondary adventitious roots with *L bicolor* and *P. indica*. Exogenous auxin application mimicked the pheno-

type observed with some of the strains. Additionally, there was enhanced expression of AtGH3-GUS, an auxin induced reporter line with GM30 co-culture, in the roots, indicating that this microbe could alter auxin accumulation in roots. The inhibition of lateral root formation through local application of auxin transport inhibitor can be compensated by *P. indica* and GM30, and to a smaller extent by *L. bicolor* co-culture. This supports previous reports that auxin is an important modulator of microbe-altered root architecture. To genetically dissect the components of auxin signaling, and transport involved in modulating various architectural phenotypes, we examined the root phenotypes of mutants available in *Arabidopsis* and a few AUX/IAA RNAi mutants in *Populus*, with co-culture. Among the signaling mutants, we find that *iaa34, iaa7* and *gh3-17* display altered response *to P. indica,* while *arf7* and *arf19* display altered response to GM30. Interestingly, *iaa3* displays altered response to all the 3 microbes tested in this study. The *Populus* mutant *iaa7* also seemed to have reduced responses to *L. bicolor* and *P. indica*. Among the mutants defective in auxin transport, we find that *aux1-7, lax2, lax3, pin3, pin7, pin3-pin7* and *abcb4-1* display altered response to *P. indica* and *L. bicolor*. These results suggest that the different microbes may use alternate components of the signaling and transport pathway of auxin to produce respective architectural alterations. To see if the identified auxin signaling, and transport proteins change with co-culture, RT-PCR experiments are underway to understand changes of respective signaling and transport genes over the duration of the experiment. Additionally, transgenic plants that have GFP/YFP fused with auxin transport proteins are being used to determine if the localization and amount of these proteins change over time. We find that PIN2-GFP and AUX1-YFP expression reduce with *P. indica* and GM30 at 5-6 days after infection. The accumulation of flavonoids, a class of secondary metabolites that have been shown to inhibit auxin movement, was also found to be altered in the presence of the tested microbes. Together, these results indicate that the alteration of root architecture during plant-microbe interaction involves complex regulation of auxin transport and signaling, possibly unique to individual microbes.

Student Oral Presentation–Tuesday

# 212

## Plant-Microbe Interfaces: Experimental and Computational Approaches for Microbial Diversity Characterization Using Artificial Communities

Migun Shakya[1,2*] (shakyam@ornl.gov), Zamin K. Yang,[2] **Christopher W. Schadt**,[2] and **Mircea Podar**[2]

[1]UT-ORNL Graduate School of Genome Science and Technology, Knoxville Tenn.; and [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

**Project Goals: The goal of the Plant-Microbe Interfaces science focus area 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 an initial test 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 will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.**

The use of 16S ribosomal RNA gene sequencing (rDNA) to characterize the taxonomic diversity and abundance of organisms plays an important role in microbial ecology studies. Microbial diversity characterization by pyrosequencing dramatically increases the scale at which environmental samples can be analyzed both in number and sequence depth. Nevertheless, the approach often has limitations for which it is difficult to control or account. Such limitations include unequal efficiency of DNA isolation, unknown ribosomal gene copy number, biases in DNA amplification, and sequencing errors. In addition, the myriad of computational tools available to analyze the data makes the data analysis even more confusing. Previous studies aimed at improving the techniques and overcoming some of those limitations were limited by sequencing depth, as well as the diversity of completed reference genomes. To determine the effects of experimental and computational steps involved in characterization of microbial diversity by 454 sequencing, we constructed artificial genomic DNA communities using cultivated organisms that have known genomic sequence and rDNA copy number. More than 60 species representing 14 bacterial and 3 archival phyla were included in the analysis. Genomic DNAs were mixed at known concentrations and hypervariable regions of the 16S rDNA gene (V1-2, V1-3, V4, V3-5, V3-9, and V6-9 for bacteria and V4, V3-9 for archaea) were amplified and sequenced with the 454 FLX and 454 Titanium system. In order to test the primer bias we also carried out a metagenomic study of the artificial community. The data was analyzed in terms of PCR/sequencing errors, chimeras, number of OTUs, abundance of individual alleles and species. For many taxa, the inferred abundance matched relatively well to the composition of the assembled communities. However, significant primer-dependent biases were observed for particular species or even phyla. A single set of 16S rDNA primers may, therefore, incompletely represent the diversity present in natural microbial communities. Similarly, the choice of data analysis pipelines also produced different results terms in number of OTUs. The computational tools should, thus, be carefully chosen based on the posed question. In addition, mock communities with known members should be constantly tested to authenticate the sequencing protocol and the computational tools.

# 213

## Plant-Microbe Interfaces: Emerging Technologies for the Functional Characterization of Isolates from *Populus*

**Jennifer Morrell-Falvey**[1]* (morelljl1@ornl.gov), Suresh R. Neethirajan, A. Nicole Edwards,[1,2] Bernadeta Srijanto,[3] Roy Dar,[3] Sarah Melton,[1] Dale Pelletier,[1] **Scott T. Retterer**,[1,3] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]UT-ORNL Graduate School of Genome Science and Technology, Knoxville, Tenn.; and [3]Center for Nanophase Materials Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

**Project Goals: The goal of the Plant-Microbe Interfaces science focus area 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 an initial test 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 will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.**

The composition of microbial communities within and around plant species is dependent on dynamic physical and chemical signaling events that occur within the local environment and at the root surface. Visualization and quantification of these events in natural systems is challenging. However, emerging technologies that combine advances in nanostructure fabrication, microfluidics and imaging provide a means of recreating these events within model systems. These systems mimic aspects of their natural counterparts while providing tractable experimental platforms in which both individual cellular responses and population dynamics can be recorded and analyzed.

Model systems, amenable to imaging, that allow dynamic modulation of local physicochemical cues in a controllable manner have been developed to recreate the interactions between microbes and their hosts. Building from work aimed at sampling and cultivating isolates from the *Populus* rhizo- and endospheres these tools will provide a means of screening the chemotactic response, surface adherence, and colonization dynamics of individual *Populus* isolates.

A nanostructured microfluidic platform has been created in order to examine the chemotactic responses of isolates to specific plant-associated signals. This platform is created using a combination of electron beam lithography and anisotropic silicon etching techniques. It can be easily replicated via silicone molding and facilitates the physical tracking of hundreds of microbes within a quasi-two-dimensional space that confines microbes within the focal volume of a conventional phase contrast microscope without significantly impeding natural motility. A nanostructured interface or membrane separates the main "chemotaxis" channel from two "feeder" channels that allow spatiotemporal modulation of the chemical environment within the main channel with negligible hydrodynamic disturbance of the microbial population (Figure 1).

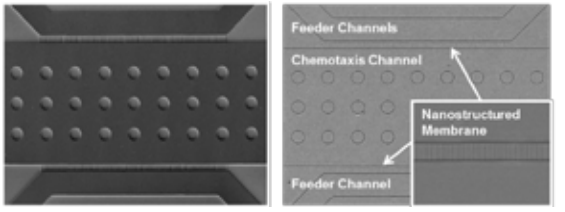



Figure 1. (A) A silicon and polymer master is formed using a combination of electron beam lithography, anisotropic silicon etching and crosslinkable polymer. (B) This master can be replicated in silicone to form a microfluidic chamber with nanostructured membranes that allow material exchange during real-time imaging of the microbial response

The dynamics of material exchange within the system has been characterized using fluorescence microscopy. Complete modulation of the local chemical environment within the device can occur within a ten-minute cycle, allowing changes in microbial motility to be monitored over time. Additionally, administration of multiple reagents from separate feeder channels enables the creation of chemical gradients in which population dynamics can be monitored. Proof-of-concept studies have been carried out, examining differences in motility and chemotactic response in *A. brasilense* and three Che1 mutant strains. Differences in velocity, reversal frequency and rate of directional change were recorded and quantified under different environmental conditions. Additional studies are being conducted to examine the response of *Populus* isolates to known plant metabolites.

Imaging studies of colonization and surface adherence have been carried out using confocal fluorescence imaging and atomic force microscopy. Real-time, 3-dimensional imaging of colonization dynamics was carried out in *Populus* roots using natural isolates, transformed to express GFP. The combined autofluorescence from the plant roots and GFP expression from the isolates allowed the growth of microbial colonies within and around the roots to be tracked over time. Atomic force microscopy was used to track the evolution of microbial biofilms from *Populus* isolate with even greater resolution. AFM analysis enabled the observation of pili formation and the evolution of distinct microbial morphologies over the course of the biofilm formation.

Taken together, the use of emerging technologies for imaging and the creation of model systems allow the observation and quantification of microbial responses to specific changes in their environment at scales that are unprecedented in natural systems. Moving forward with these technologies, we look towards the observation of complex microbial communities to better understand community dynamics across the plant-microbe interface.

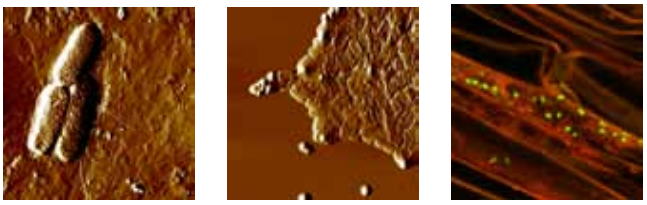


Figure 2. (A & B) Protocols for preserving pili and ultrastructural components of bacterial isolates for atomic force microscopy have been developed. Imaging on colonies of GM30 and YR343 strains on biofilms grown for 4,8,16 and 24 hours on mica was executed to better understand the expression of pili and fimbriae, as well as overall biofilm morphology. (C) *Populus* roots were fixed in 4% paraformaldehyde for detection of associated bacteria by fluorescence in situ hybridization (FISH). Bacterial probe EUB338 was labeled with Alexa488 (green). Plant roots are detected by autofluorescence (red).

# 214

## Plant-Microbe Interfaces: Collaboration Platform for Scientific Communication, Management, Information Storage and Sharing

**Guruprasad Kora**[2]* (koragh@ornl.gov), Michael Leuze,[2] Tatiana Karpinets,[1] Mustafa Syed,[1] Susan Holladay,[1] **Edward C. Uberbacher**,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division and [2]Computer Science and Mathematics, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

**Project Goals: The goal of the Plant-Microbe Interfaces science focus area 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 an initial test 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 will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.**

Plant-Microbe Interfaces (PMI) knowledgebase represents a unique platform for biologists, analysts and data collectors to share, collaborate and analyze data from a single point of access. Its goal is to consolidate all team related low-level scientific data as well as document-level project related information into an interactive computational environment. The PMI knowledgebase environment offers exceptional computational and collaborative capabilities to biologists with little or no computational background. The PMI portal is currently serving 80 members of the Department of Energy (DOE) funded Plant Microbe Interfaces scientific focus area. The PMI knowledgebase comprises of three integrated components: a data sharing and collaboration platform, a Laboratory Information Management System (LIMS) interface, and a content management system with built-in querying engine. Each of these components is presented with an easy to use web interface with appropriate security infrastructure built-in.

The PMI portal's collaborative platform was designed from the ground up, with scientific team collaboration and data sharing aspects in mind. It utilizes modern collaboration and social networking features that provide structure-less social utility tools that connect and facilitate a group of like-minded co-workers to share, collaborate and discuss on a given scientific task. Furthermore, the platform enables social features such as "Scientific Walls" to provide single point discussion threads to facilitate better inter-group interaction. This approach towards team collaboration turns out to be the best technique to assimilate and disperse data-driven knowledgebases.

The PMI portal's LIMS interface enables users to access the underlying data management layer with an easy to use and intuitive web interface. The interface seamlessly connects to laboratory-wide LIMS environment and makes day-to-day tasks like raw-data fetching and data summarization extremely efficient and easy. Furthermore, the portal provides a dynamic data analytics environment that facilitates users to perform standard statistical analysis on the LIMS stored data.

The portals content management layer handles document-level data across portal users. It provides a single unified repository to manage any content type—documents, images, data sheets, archives etc. The content management system has an intelligent reference engine that enables real-time content tagging with project entities like goals, project-wide events, and system participants, thus enabling intelligent tracking of document life-cycle. The content management system is fine-tuned towards scientific data management and retrieval processes as it provides document preview features to scientific data types directly within your browser, without having to download them. The system provides a powerful search capability that enables in-document and free-text searching across all managed content. An alert system works in conjunction with the document management system to enable real-time tracking of content across the system.

PMI knowledgebase and Portal can be accessed at *http://pmi.ornl.gov*

# 215

## Plant-Microbe Interfaces: Exploring Mutualistic and Parasitic Symbiotic Relationships of Microbes With Plants Using PMI Knowledgebase Tools

**Tatiana Karpinets**[1]* (karpinetstv@ornl.gov), Byung Park,[2] Michael Leuze,[2] Guruprasad Kora,[2] Mustafa Syed,[1] Dale Pelletier,[1] Christopher Schadt,[1] **Edward Uberbacher**,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division and [2]Computer Science and Mathematics, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

**Project Goals: The goal of the Plant-Microbe Interfaces science focus area 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 an initial test 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 will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.**

Symbiotic interactions between microbial organisms and plants can be mutualistic, when it is beneficial for both of them; parasitic, when the microbial organism benefits at the expense of the plant; or commensal, when the microbial organism benefits without damaging the plant. Although some cases of plant-microbe interactions are ambiguous and cannot be strictly classified as beneficial, parasitic, or commensal, many microbial organisms are well-known pathogens of plants and cause devastating diseases to their host. The other interactions, like mycorrhizal associations between plant roots and fungi or associations of nitrogen-fixing rhizobium bacteria with legumes, are well-known examples of resource-resource based mutualism, when one type of resource produced by the microorganism is traded for another resource produced by the plant. The distinct mutualistic and parasitic phenotypes of microorganisms make their genomes a valuable target for comparative analyses. Each of these symbiotic relationships involves adaptation and evolution of microbes and leads to appropriate changes, not only in the phenotypes of the organisms, but also in their genomes. Thus, specific molecular functions, metabolic pathways and biological processes should exist in microbial genomes that underlie the mutualistic or parasitic nature of symbionts and, likely, distinguish one phenotype from the other.

In this study we used a comparative analysis of bacterial genomes representing different types of plant-bacterial symbionts to explore genomic features underlying mutualistic and parasitic bacterial phenotypes. These phenotypes will be referred as plant pathogens and plant endophytes. An initial challenge of the study was to compile a comprehensive set of sequenced pathogens and endophytes for the comparison. This was addressed by developing a novel tool, the "Genes/genomes On-Line explorer" (GOLexplorer) as a part of the PMI knowledgebase. The tool allows one to infer confident relationships between characteristics of genes or organisms if their classifications in terms of these characteristics are available for a set of the objects. If the object is an organism, for example, the characteristics can include a type of symbiotic relationship, a type of metabolism, its host, a taxonomic group, or preferred temperature range. The developed tool uses the association rule-learning algorithm to find confident relationships between such characteristics. We have used GOLexplorer to discover associations between characteristics of bacterial organisms available in the GOLD database and to compile representative lists of plant pathogens and endophytes for further genome comparative analyses.

Our initial selection of plant symbionts using the GOLexplorer identified 28 sequenced endophytes and 36 pathogens with no significant differences between the groups in genome sizes, GC contents and taxonomic classification of the organisms at the level of phylum and order. In both groups most organisms (~80%) belonged to the phylum of proteobacteria. At the level of order, both, pathogens and endophytes, have representatives of Acholeplasmatales, Burkholderiales, Enterobacteriales, Pseudomonadales, Rhizobiales, and Xanthomonadales. Most sequenced endophytes (90%) were also annotated by the "nitrogen fixation" phenotype, but only one pathogen had this annotation. Species in the dominating orders of pathogens often had several representatives of the same genus and even of the same species. To equal the number of organisms in both groups we have removed such duplicates. This filtering resulted in 28 pathogens and 28 endophytes for further comparative analysis at the level enzymes and metabolic pathways using the Pathway Tools available in the PMIcyc and our recently developed toolkit for prediction in the genomes of Carbohydrate-Active enzymes (CAZymes).

One interesting finding of the analysis was the presence of distinct enzymatic signatures for pathogens and endophytes in terms of enriched CAZy families in their genomes. The genomes of all endophytes were enriched in CAZymes with a constructive (biosynthetic) metabolic activity. The enzymes belonged to several families of glycosyltransferases involved in the synthesis of oligo- and polysaccharides. Genomes of pathogens had a relative abundance of CAZymes with a destructive (degrading) metabolic activity. These enzymes belonged to the family of glycosyl hydrolases, which are involved in the release of glucose from oligo- and polysac-

charides. This set of enzymes was more specific for each pathogenic organism, most likely, because different sugars can dominate in different plant hosts. These and other findings will be presented, which describe the different metabolic profiles of endophytes and pathogens.

## The Predictive Microbial Biology Consortium

# 216

### Characterization of Naturally Occurring and Model Microbial Communities

Jennifer J. Mosher,[1] James G. Moberly,[1] Christopher W. Schadt,[1] Tommy J. Phelps,[1] Mircea Podar,[1] Steven D. Brown,[1] Anthony V. Palumbo,[1] Michael W.W. Adams,[2] David A. Stahl,[3] Kristina L. Hillesland,[3] Judy D. Wall,[4] Matthew W. Fields,[5] Terry C. Hazen,[6] and Dwayne A. Elias[1]* (eliasda@ornl.gov)

[1]Oak Ridge National Laboratory; [2]University of Georgia; [3]University of Seattle; [4]University of Missouri; [5]Montana State University; and [6]Lawrence Berkeley National Laboratory

**Project Goals: Microbial community construction is ongoing with *Desulfovibrio vulgaris*, *Geobacter sulfurreducens* and *Methanococcus maripaludis* to determine carbon mineralization, energy balance and electron accepting patterns in model communities. Cultures are assessed as syntrophic and competitive communities to examine intercellular communication at several omic levels and is in coordination with *D. vulgaris/M. maripaludis* co-culture and evolution studies, stress conditions implemented in-situ at Hanford, and utilizing *D. vulgaris* mutants to determine particular gene importance on community function. All data will help construct metabolic models at different complexities with other ENIGMA groups. The aims of in-situ microbial community assessment in Hanford groundwater are to determine the temporal population succession while isolating new keystone species. Further, these approaches can be used as a proxy for in-situ tests in an effort to predict the results of different perturbations.**

Microbial community construction, cultivation and analyses are being performed using the metal-reducing bacteria *Desulfovibrio vulgaris* Hildenborough and *Geobacter sulfurreducens* PCA as well as *Methanococcus maripaludis* S2 to determine complete carbon mineralization, energy balance and electron accepting patterns in consortia communities. Mono-, co-, and tri-cultures are being assessed both as syntrophic and competitive communities to examine cell to cell communication at several omic levels and are being performed in coordination with *D. vulgaris/M. maripaludis* co-culture and evolution studies (D. Stahl and K. Hillesland).

These communities will be subjected to stress conditions being implemented *in-situ* at Hanford (T. Hazen) while also utilizing several *D. vulgaris* mutants to determine the importance of particular genes and metalloproteins on community function (with J. Wall and M. Adams). Technologies developed include multispecies microarrays, along with species specific qPCR primers and fluorescent antibodies to better understand intercellular coordination within these defined communities. All data will help construct metabolic models at different complexities with microarray data being directly comparable to existing datasets from other ENIGMA groups. Future ENIGMA efforts are likely to be able to take advantage of similar technologies and these can be developed quickly for new keystone organisms of interest.

The aims of *in-situ* microbial communities assessment in Hanford groundwater are to determine the temporal population succession while isolating new species key to the function of these populations. Recently, Hanford 100H groundwater (from T. Hazen) was inoculated into triplicate, custom designed flow through reactors and incubated (30°C) for 95 days with a 100 hour generation rate. Protein, gas and liquid metabolite quantification, and 16S rDNA identification of the microbial community members were highly reproducible with the final community dominated by the genera *Pelosinus*, *Acetobacterium*, *Methanobacterium* and *Methanosarcina*. Six new genera and seven new species of sulfate- and Fe(III)- reducing bacteria were isolated including three new *Pelosinus* spp. All isolates are currently being assessed for Fe(III), U(VI), and Cr(VII) reduction. In collaboration with Mike Adams (U. Georgia) preliminary results show that cell uptake rates of V, Fe, Co, Se, W, and Mo are coordinated to the temporal succession and recession of different community groups. The follow-on experiment is to repeat these conditions with and without *in-situ* Cr(VII) levels to determine the Cr influence on planktonic community structure, in coordination with biofilm studies (M. Fields, Montana State U.). Short and longer-term environmental perturbations will be coordinated with recently developed ENIGMA field plan for Hanford to test hypotheses developed from the *in-situ* experiments. Such efforts and integration will generate a more comprehensive understanding of the community and reaction to perturbations, while supplying new microbial consortia and isolates to the wider ENIGMA team in order to further ENIGMA and DOE goals.

Student Oral Presentation–Monday

# 217

## Characterizing the Metalloproteomes of Model Microorganisms

W. Andrew Lancaster[1]* (lancaste@uga.edu), Angeli L. Menon,[1] Sunil Kumar,[1] Farris L. Poole,[1] Ming Dong,[2] Megan Choi,[2] Mark Biggin,[2] Haichuan Liu,[3] H. Ewa Witkowska,[3] Sunia A. Trauger,[4] Gary Siuzdak,[4] Steven M. Yannone,[2] John A. Tainer,[2,4] and **Michael W.W. Adams**[1]

[1]University of Georgia, Athens; [2]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [3]University of California, San Francisco; and [4]Scripps Research Institute, La Jolla, Calif.

**Project Goals: Metal ion co-factors afford proteins virtually unlimited catalytic potential, enable electron transfer reactions and greatly impact protein stability. Consequently, metalloproteins (MPs) play key roles in virtually all biological processes. However, predicting the types of metal that an organism utilizes in its metalloproteome from its genome sequence is currently impossible since metal coordination sites are diverse and poorly recognized. Determining the identity of MPs directly from native biomass can resolve some of these issues. We are using *Pyrococcus furiosus*, a hyperthermophilic archaeon that grows optimally at 100°C, as the model organism.**

Metal ion co-factors afford proteins virtually unlimited catalytic potential, enable electron transfer reactions and greatly impact protein stability. Consequently, metalloproteins (MPs) play key roles in virtually all biological processes. However, predicting the types of metal that an organism utilizes in its metalloproteome from its genome sequence is currently impossible since metal coordination sites are diverse and poorly recognized. Determining the identity of MPs directly from native biomass can resolve some of these issues. We are using *Pyrococcus furiosus*, a hyperthermophilic archaeon that grows optimally at 100°C, as the model organism. Large scale fractionation of native biomass using non-denaturing, sequential liquid chromatography (26 columns) coupled with high-throughput tandem mass spectrometry (HT-MS) to separate and identify proteins led to the identification of ~80% (967) of the cytoplasmic proteins in *P. furiosus*. By coupling native biomass fractionation with inductively coupled plasma mass spectrometry (ICP-MS), a robust, metal-based approach was developed to determine metals an organism assimilates on a given growth medium and identify metalloproteins on a genome- wide scale. Of 343 metal peaks in chromatography fractions, 158 did not match any predicted metalloprotein. Unassigned peaks included metals that *P. furiosus* was known to utilize (cobalt, iron, nickel, tungsten and zinc; 83 peaks) plus metals the organism was not thought to assimilate (lead, manganese, molybdenum, uranium and vanadium; 75 peaks). By shifting the focus from classical protein-based purification to metal-based identification, eight of 158 unexpected metal peaks were purified yielding four novel nickel and molybdenum-containing proteins, whereas four proteins contained sub-stoichiometric amounts of misincorporated lead and uranium. Analyses of two additional microorganisms (*Escherichia coli* and *Sulfolobus solfataricus*) revealed species-specific assimilation of yet more unexpected metals. Metalloproteomes are therefore much more extensive and diverse than previously recognized, and promise to provide key insights for cell biology, microbial growth and toxicity mechanisms (Cvetkovic et al., 2010).

Computational analysis of the large parallel metal and protein dataset (2589 column fractions) yielded predictions of novel metalloproteins in *P. furiosus* (Lancaster et al., 2011). The data are available at http://enigma.bmb.uga.edu/IMPACT. Homologous recombinant production of *P. furiosus* MPs for structural analyses is currently underway. The methodology to identify and predict MPs on a global scale can be adapted and applied to any organism and also provides a road map for the (partial) purification of native forms of novel MPs. We are currently applying this technology to *Desulfovibrio vulgaris*. Large-scale growth (600 L) of the organism and metal determinations using existing chromatography column fractions and correlations with iTRAQ/MS data are currently in progress.

This work was conducted as part of the ENIGMA project and 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.

**References**

1. Cvetkovic et al. (2010) Microbial metalloproteomes are largely uncharacterized. *Nature* 466, 779
2. Lancaster et al. (2011) A computational framework for proteome-wide pursuit and prediction of metalloproteins using ICP-MS and MS/MS data. *BMC Bioformatics* (in press)

# 218

## Understanding the RNA Landscape in the Microbial World

**John Tainer** (jat@scripps.edu) and Robert Rambo*

Lawrence Berkeley National Laboratory, Berkeley, Calif.

**Project Goals: ENIGMA scientists seek to understand in situ microbial activity and community dynamics through detailed assessment of molecular function from proteins to populations. By studying communities with activities of interest to DOE mission we hope to reveal the mechanistic basis for those activities and their support in a changeable and uncertain environment. ENIGMA has 4 main aims**

- **Measurement and analysis of environmental activity, composition, structure, and strategies of microbial communities in situ**
- **Use controlled laboratory consortia to identify essential microbial contributions to environmental activities, identify specific and selected interactions, and isolate keystone organisms/processes**

- **Efficiently advance these environmental microorganisms to model organism status and map their molecular functions to community phenotypes and environmental activities.**
- **Development of the LBNL Systems Environmental Microbiology Workbench and Knowledge Framework**

RNA has many diverse roles in microbial biology including direct involvements in transcriptional regulatory elements, RNA modifying enzymes, protein synthesis, intracellular protein trafficking and microbial defense. Therefore, as we aim to establish the microbial biological network for bioengineering, the network must involve a description that includes the role of RNA. Riboswitches such as the S-adenosyl methionine (SAM) riboswitch senses the metabolic environment promoting transcription of an operon in low intracellular SAM conditions. Engineering these microbes for DOE applications requires a systems-biology understanding that must involve the functional annotation of non-coding RNAs (ncRNAs). Our work takes a collaborative approach by leveraging diverse experimental techniques within the native environments of *Desulfovibrio, Halobacterium NRC-1, Pyrococcus furiosus* and *Sulfolobus solfataricus.*

# 219

## Integrated Microbiological Approaches to Characterize Cr(VI)-Reducing Microbial Community at the DOE Hanford 100H Site

Romy Chakraborty* (rchakraborty@lbl.gov), Dominique Joyner, Boris A. Faybishenko, Matthew Fields, Tamas Torok, Gary L. Andersen, and Terry C. Hazen*

Lawrence Berkeley National Laboratory

**Project Goals: We have successfully used different approaches in identifying the key microbial players involved in Chromium reduction at the DOE-Hanford 100H site, and are currently developing different strategies to best understand the key metabolic processes mediated by these microbes in the field.**

In order to stimulate microbially-mediated reduction of Cr(VI), a poly-lactate compound (HRC) was injected into the chromium-contaminated aquifer at the Hanford (WA) 100H DOE site in 2004. Cr(VI) concentrations rapidly declined to below the detection limit and using high-density DNA 16S rRNA gene microarray (Phylochip), we observed the community to transition through denitrifying, iron-reducing and sulfate-reducing and methanogenic populations. Based on these results, targeted enrichments in defined anaerobic media resulted in the isolation of an iron-reducing *Geobacter metallireducens*-like isolate strain RCH3, a sulfate-reducing *Desulfovibrio vulgaris*-like strain RCH1 and a nitrate-reducing *Pseudomonas stutzeri*-like isolate RCH2 and *Sporotalea* strain 45W among several others. These isolates were capable of reducing Cr(VI) anoxically and the whole-genome sequence data for the first three is now available from JGI. OMNILOG Phenotypic microarray was used to compare isolate RCH1 with the type strain *Desulfovibrio vulgaris Hildenburough (DvH)*. The phenotypic microarray allows for high throughput screening of metabolic activity of diverse microorganisms, the panels providing assays for approximately 760 select compounds measuring metabolism of various C, N and P substrates. The high throughput BIOLOG was used for Minimum Inhibitory Concentration (MIC) determinations of environmentally relevant stressors. Further, polyclonal antibodies were raised against the functionally dominant organisms at Hanford including *Methanococcus, Desulfovibrio, Pseudomonas* and *Geobacter* spp and tagged with different fluorescent dyes to enable specific direct enumeration and visualization from environmental samples. Also, streptavidin-coupled paramagnetic beads and biotin labeled antibodies raised against surface antigens of *DvH* were used to capture these type of cells in both bioreactor grown laboratory samples and from Hanford groundwater samples. Field deployable IMS technology may greatly facilitate environmental sampling and bioremediation process monitoring and enable transcriptomics and proteomics/metabolomics-based studies directly on cells collected from the field.

We have successfully used different approaches in identifying the key microbial players involved in Chromium reduction at the Hanford 100H site, and are currently developing different strategies to best understand the key metabolic processes mediated by these microbes in the field.

# 220

## Microbial Community Dynamics from Groundwater and Surrogate Sediments During HRC® Biostimulation for Cr(VI)-Reduction

K.B. De Leon,[1,2,6] B.D. Ramsay,[2,6] D.R. Newcomer,[3] B. Faybishenko,[4,6] T.C. Hazen,[4,6] J. Zhou,[5,6] and M.W. Fields[1,2,6]* (matthew.fields@erc.montana.edu)

[1]Department of Microbiology, [2]Center for Biofilm Engineering, Montana State University; [3]Pacific Northwest National Laboratory; [4]Lawrence Berkeley National Laboratory; [5]University of Oklahoma; and [6]ENIGMA

http://enigma.lbl.gov

**Project Goals: Determine bacterial community dynamics during biostimulation for chromate reduction.**

The Hanford 100-H site is a chromium-contaminated site that has been designated by the Department of Energy Environmental Management as a field study site for *in situ* chromium reduction. In August 2004, the first injection of hydrogen release compound (HRC®) resulted in an increase of microorganisms and a reduction of soluble chromium(IV) to insoluble chromium(III). Little is understood about the microbial community composition and dynamics during stimulation. The aim of this study is to compare microbial

communities of groundwater and soil samples across time and space during a second injection of HRC® via bar-coded pyrosequencing. We have also attempted to validate the pyrosequencing approach to microbial community analysis via the comparison of species richness and diversity estimates to a corresponding clone library for the V4 and V6 regions of SSU rDNA. These results indicate that pyrosequencing data must be thoroughly filtered and that a quality score cutoff is not universal across the SSU rDNA gene likely due to differing proportions of conserved and variable regions.

A second injection occurred November 2008 at the 100-H field site, and geochemical data collected throughout the study showed an overall decrease in nitrate, sulfate, and chromium(IV). Spatial and temporal water and soil samples were collected pre-and post-injection from four wells at the field site. Soil columns constructed from stainless steel mesh were lined with nylon mesh and filled with Hanford soils from the 100-H site. The soil columns were used to represent not only the microbes flowing through the soil via groundwater, but the microbes that require a matrix in order to grow. DNA was extracted from each of the samples and SSU rDNA gene fragments was sequenced via multiplex pyrosequencing. Soil samples differed from the corresponding groundwater (even at the phyla level) and were more diverse. However, although the community composition changed during the biostimulation, the overall community diversity was not altered. Results do not indicate a large shift in dominant organisms in soil from pre- to post- injection, and this may be due to the organisms remaining dominant from the first stimulation. However, a prevalence of core genera and rare genera were observed across 26 samples while urban and rural genera were less abundant. The β-Proteobacteria were more predominant in soil samples while γ-Proteobacteria were more equivalent in both sample types. There was a shift from *Acidovorax* to *Aquaspirillum* from upstream (non-stimulated) to downstream soil both pre- and post-injection. Furthermore, while post-injection soil samples indicate a continuing dominance of *Aquaspirillum*, corresponding water samples indicate *Pseudomonas* as a dominant genus. The greatest changes during stimulation occurred in the populations of mid-dominance either between wells or across time, and these organisms could be important to consider as possible indicator species in future work. Work in progress includes continued phylogenetic structure and composition analyses and characterization of functional diversity via GeoChips.

# 221

## Adaptive Evolution and Physiology of Nascent Microbial Mutualisms

Kristina L. Hillesland,[1]* Birte Meyer,[1]* Nicolas Pinel,[1] Nicholas Elliott,[1] Marcin Joachimiak,[2] Jennifer Kuehl,[2] Adam Deutschbauer,[2] Aifen Zhou,[3] Zhili He,[3] Jizhong Zhou,[3] Dwayne Elias,[4] Terry Hazen,[2] Adam Arkin,[2] and David A. Stahl[1]* (dastahl@u.washington.edu)

[1]University of Washington, Seattle; [2]Lawrence Berkeley National Lab; [3]University of Oklahoma; and [4]Oak Ridge National Lab

http://enigma.lbl.gov
http://stahl.ce.washington.edu/

**Project Goals: ENIGMA scientists seek to understand in situ microbial activity and community dynamics through detailed assessment of molecular function from proteins to populations. By studying communities with activities of interest to DOE mission we hope to reveal the mechanistic basis for those activities and their support in a changeable and uncertain environment. In support of those objectives the studies presented here are designed to:**

1. **extend understanding of genetic and metabolic networks sustaining and stabilizing natural microbial communities by characterizing different synthetic assemblies of species functioning in a simple two-tier food web of general environmental significance,**
2. **predict end-products of natural selection occurring in a community context,**
3. **develop a bank of genetic variants with known ecological history and evolutionary relationships to enable comprehensive genotype-phenotype map, and**
4. **identify mechanisms of specificity in interactions between species.**

A goal of DOE and ENIGMA is to understand and ultimately predict microbial community responses to environmental change. Key to achieving that goal will be determining the genetic process by which the environment affects population and community characteristics. We studied a simple two-tier microbial food chain composed of *Desulfovibrio* species and hydrogenotrophic methanogens cooperating syntrophically to degrade lactate through the obligate exchange of hydrogen. First, growth and metabolism rates in several pairings of different *Desulfovibrio* and methanogen species were compared and biomass was collected to explore the genetic basis of phenotypic variation among these pairings by microarray analysis. Second, the genetic basis of fitness improvement in an evolving syntrophy between *D. vulgaris* Hildenborough and *M. maripaludis* was explored by genome resequencing and phenotypic comparisons. Evolutionary changes after more than 300 and generations of cooperative growth include significantly increased stability, yield, and growth rate. Illumina sequencing of coculture U9 at 300 generations identified a few molecular differences between both evolved species and their common ancestors. A conserved hypothetical protein (DVU_0799)

in *D. vulgaris* with sequence similarities to an outer membrane porin had two mutations (at population frequencies of 100 and 60%). These two mutations change acidic or polar amino acids to non-polar amino acids, and therefore likely affect function. Several other populations that evolved independently but in the same conditions also substituted mutations or in-frame deletions in the same 200 bp region of this gene within the first 300 generations of evolution. Together these results suggest that DVU_0799 has a large beneficial effect on fitness in the evolution environment. *D. vulgaris* clones containing one or both of these mutations could all improve coculture growth, but the magnitude of the effects varied depending on the presence of other mutations. Experiments comparing gene expression provided evidence that a *M. maripaludis* clone from U9 had a differential effect on gene expression in evolved versus ancestral *D. vulgaris*, suggesting that adaptive changes caused specific genetic interactions between evolved *M. maripaludis* and *D. vulgaris*. With continued evolution to 1000 generations, growth rates of evolving cocultures improved substantially. An isolation-independent genome-wide characterization of 12 of these 1000 generation communities using the SOLiD 3 platform identified an average of 10 and 198 mutations at frequencies of ≥70 or ≥25%, respectively, in the 24 species populations. Surprisingly, independent non-/missense mutations were detected frequently in sulfate reduction genes in different *D. vulgaris* lineages. The loss of sulfate reduction capacity by *Desulfovibrio vulgaris*—an organism defined by this characteristic physiology—in multiple lineages has significant implications for better understanding adaptive processes leading to more efficient use of available free energy by microbial communities. Together these data show that some initial evolutionary responses of *D. vulgaris* and *M. maripaludis* to a new, mutualistic environment are repeatable, and that they may affect the interactions between these species. It may thus be feasible to predict some evolutionary responses of species of interest to DOE to environmental change, even when these species are evolving in a community.

# 222

## Parallel Evolution of Transcriptome Structure During Genome Reorganization

Sung Ho Yoon[1]* (syoon@systemsbiology.org), David J. Reiss,[1] J. Christopher Bare,[1] Dan Tenenbaum,[1] Min Pan,[1] Joseph Slagel,[1] Sujung Lim,[2] Murray Hackett,[3] Angeli Lal Menon,[4] Michael W.W. Adams,[4] Adam Barnebey,[6] Steven M. Yannone,[6] John A. Leigh,[2] and **Nitin S. Baliga**[1] (nbaliga@systemsbiology.org)

**PI: Nitin S. Baliga**[1] (nbaliga@systemsbiology.org)
**Co-PIs:** John A. Leigh,[2] Murray Hackett,[3] William Whitman,[5] Paul Adams,[7] Adam Arkin,[7] Terry Hazen,[7] Michael W.W. Adams,[4] Greg Hura,[7] Steven M. Yannone,[6] Stephen Holbrook,[7] Gary Siuzdak,[8] and John A. Tainer[7]

[1]Institute for Systems Biology, Seattle, Wash.; [2]Dept. of Microbiology and [3]Dept. of Chemical Engineering, Univ. of Washington, Seattle; [4]Dept. of Biochemistry and Molecular Biology, [5]Univ. of Georgia, Athens; [6]Life Sciences Division, [7]Lawrence Berkeley National Laboratory, Berkeley, Calif.; and [8]Scripps Research Institute, La Jolla, Calif.

http://baliga.systemsbiology.net/enigma/
http://baliga.systemsbiology.net/drupal/content/h2-regulation
http://enigma.lbl.gov

**Project Goals:**

1. **Use transcriptomics, proteomics, and metabolomics to study the systems biology of $H_2$ metabolism, formate metabolism, nitrogen fixation, and carbon assimilation in *Methanococcus maripaludis*.**
2. **Determine the mechanism of $H_2$ sensing and transcriptional regulation by $H_2$.**

Genome streamlining by assembly of genes into operons ("operonization") is instrumental in the continual adaptation of microbes to their environmental niche. However, the random genome reorganization events that drive operonization are also the roots of instability for existing operons. We have determined that there exists a statistically significant trend that correlates degree of operonization in archaea to their phylogenetic lineage. We have further characterized how microbes deal with operon instability by mapping and comparing transcriptome structures of four phylogenetically diverse extremophiles that span the range of operon stabilities observed across archaeal lineages: a photoheterotrophic halophile (*Halobacterium salinarum* NRC-1), a hydrogenotrophic methanogen (*Methanococcus maripaludis* S2), an acidophilic and aerobic thermophile (*Sulfolobus solfataricus* P2), and an anaerobic hyperthermophile (*Pyrococcus furiosus DSM* 3638). We demonstrate how the evolution of transcriptional elements (promoters and terminators) generates new operons, restores the coordinated regulation of translocated, inverted, and newly acquired genes, and introduces completely novel regulation for even some of the most conserved operonic genes such as those

encoding subunits of the ribosome. The inverse correlation ($r$ = -0.91) between the proportion of operons with such internally located transcriptional elements and the number of conserved operons in each of the four archaea reveals an unprecedented view into varying stages of the operonization process. Importantly, our integrated analysis has revealed that organism adapted to higher growth temperatures have lower tolerance for genome reorganization events that disrupt operon structures.

This work was supported by U.S. Department of Energy, Award No. DE-FG02-07ER64327 and DG-FG02-08ER64685 (to Nitin S. Baliga); the Office of Science (BER), U.S. Department of Energy, Award No. DE-FG02-08ER64685 (to John A. Leigh); the Office of Science (BES), U.S. Department of Energy, Award No. DE-FG05-95ER20175 (to Michael W. W. Adams). This work conducted by ENIGMA 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.

# 223

## Systems Biology of Halophiles

Pavana Anur,[1] Justin Ashworth,[1] J. Christopher Bare,[1] Karlyn Beer,[1,4] Manjula Bharadwaj,[1] Aaron Brooks,[1,4] Elijah Christensen,[1] Aimee Desaki,[1] Danielle Durudas,[1] Joseph Horsman,[1] Amardeep Kaur,[1] Fang Yin Lo,[1,4] Bruz Marzolf,[1] Mónica V. Orellana,[1] Min Pan,[1] Wyming L. Pang,[1] Amanda Pease,[1] David J. Reiss,[1] David Rodriguez,[1] Joseph Slagel,[1] Dan Tenenbaum,[1] Serdar Turkarslan,[1] Kenia Whitehead,[1] Elisabeth Wurtman,[1] Sung Ho Yoon,[1] Sujung Lim,[2] **John A. Leigh**,[2] **Murray Hackett**,[3] Angeli Lal Menon,[5] Aleksander Cvetkovic,[6] **Michael W.W. Adams**,[5] Adam Barnebey,[6] Stephen Holbrook,[6] Trent Northen,[6] **John A. Tainer**,[6] **Steven M. Yannone**,[6] **Bonnie Baxter**,[7] **Sunia Trauger**,[8] **Gary Siuzdak**,[8] and **Nitin S. Baliga**[1,2,4]* (nbaliga@systemsbiology.org)

**PI:** Nitin S. Baliga[1,2,4]* (nbaliga@systemsbiology.org)
**Co-PIs:** John A. Leigh,[2] Murray Hackett,[3] Michael W.W. Adams,[5] Trent Northen,[6] John A. Tainer,[6] Steven M. Yannone,[6] Bonnie Baxter,[7] Sunia Trauger,[8] and Gary Siuzdak[8]

[1]Institute for Systems Biology, Seattle, Wash.; [2]Dept. of Microbiology, University of Washington, Seattle; [3]Dept. of Chemical Engineering, University of Washington, Seattle; [4]Molecular and Cellular Biology Program, University of Washington, Seattle; [5]Dept. of Biochemistry and Molecular Biology, University of Georgia, Athens; [6]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; [7]Westminster College, Salt Lake City, Utah; and [8]The Scripps Research Institute, La Jolla, Calif.

http://baliga.systemsbiology.net/enigma/
http://baliga.systemsbiology.net/drupal/content/enigma
http://enigma.lbl.gov

**Project Goals: Developing a cutting edge, multi-scale systems biology framework spanning from single cell to community level interactions to enable predictive modeling of *Halobacterium salinarum* NRC-1 as a model to understand evolution and biomolecular interactions of DOE relevant organisms.**

Bioremediation of contaminated environments occurs through the collective metabolism of microbial communities. Strategies to enhance the potential of microbes to detoxify their environment require an understanding of how biological networks and community interactions govern microbial physiology. In salt-enriched environments such as the Hanford site, halophilic extremophiles are primary candidates for detoxification strategies. Here we describe the work of a broad consortium of investigators to develop tools that enable predictive modeling of *Halobacterium salinarum* NRC-1 physiology across biological and evolutionary scales. Starting from a fully annotated genome sequence and abundant systems-level measurements of the transcriptome and proteome, we describe our advances spanning from single-cell modeling of gene regulatory dynamics in subcircuits to interspecies modification of population-level behavior. All of our high-throughput technologies, network modeling algorithms, and software tools have been developed within a framework that is generalizable to other systems. This puts us in a unique position to apply these methods to other species of interest, such as DvH, and to suggest how synthetic modification of microbial physiology and community structure may complement current bioremediation efforts.

This work was supported by U.S. Department of Energy, ENIGMA (Environment and Networks Integrated with Genomes and Molecular Assemblies) Component 3, Award No. DE-FG02-07ER64327.

ENIGMA is a Lawrence Berkeley National Laboratory Scientific Focus Area Program supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics: GTL Foundational Science through contract DE-AC02-05CH11231.

# 224

## Bypassing Signal Activation in the System Wide Mapping of Genes Regulated by Response Regulators

Lara Rajeev, Eric G. Luning, Paramvir S. Dehal, Morgan N. Price, **Adam P. Arkin**, and **Aindrila Mukhopadhyay*** (amukhopadhyay@lbl.gov)

Physical Biosciences Division, Lawrence Berkeley National Laboratory

**Project Goals**

1. **To map the network of genes that are transcriptionally regulated by two component signal transduction systems in *Desulfovibrio vulgaris* Hildenborough.**
2. **To use experimentally validated binding motifs from *D. vulgaris* to predict functions of two component systems in related sulfate reducing bacteria.**

Two component regulatory systems, comprised of sensor histidine kinases and response regulators, are central to the regulation of stress responses in bacteria. Environmental bacteria especially encode large numbers of putative two component systems and the genes regulated by these systems represent the regulatory networks that impact important natural phenomena such as metal, sulfur, nitrogen and carbon cycling. However, due to lack of knowledge regarding the environmental cues that activate signal transduction, and paucity of methods for high throughput genetic manipulation, these valuable networks remain largely unmapped in most bacteria. We used an *in vitro* array-based DAP-chip (DNA Affinity Purified-chip) method to systematically map the genes regulated by all DNA binding response regulators in the model sulfate reducing bacterium, *Desulfovibrio vulgaris* Hildenborough. Our results from the DAP-chip measurements show at least 200 genes, representing approximately 84 operons, to be regulated by 24 response regulators in *D. vulgaris* Hildenborough, of which only one has characterized orthologs. Our results have allowed us to identify the response regulators involved in the regulation of flagella and pili assembly, lactate utilization, exopolysaccharide synthesis, lipid biosynthesis, and in the responses to low potassium, phosphate starvation and nitrite stresses among others. Gene sets regulated by multiple response regulators forming regulatory networks were also discovered. Finally, using the identified gene sets and orthologs in closely related bacteria, we predicted and experimentally verified binding motifs for 15 of these response regulators. These functional predictions may be applied to related species as well, since the binding site motifs appear conserved for several response regulators.

Funding: This research is funded by ENIGMA, a Scientific Focus Area Program supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL Foundational Science through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

# 225

## Development of Metagenomic Technologies for Analyzing Microbial Communities

Qichao Tu,[1] Ye Deng,[1] Zhili He,[1] Hao Yu,[1] Yujia Qin,[1] Aifen Zhou,[1] Jianping Xie,[1] Zhenmei Lu,[1] James Voordeckers,[1] Yongjin Lee,[1] Kai Xue,[1] Joy Van Nostrand,[1] Liyou Wu,[1] Yihuei Jiang,[1] **Terry C. Hazen,**[2] **Paul Adams,**[2] and **Jizhong Zhou**[1,2*] (jzhou@ou.edu)

[1]The University of Oklahoma, Norman, Okla.; and [2]Lawrence Berkeley National Laboratory, Berkeley, Calif.

http://enigma.lbl.gov

**Project Goals: ENIGMA scientists seek to understand in situ microbial activity and community dynamics through detailed assessment of molecular function from proteins to populations. By studying communities with activities of interest to DOE mission we hope to reveal the mechanistic basis for those activities and their support in a changeable and uncertain environment. ENIGMA has 4 main aims**

- **Measurement and analysis of environmental activity, composition, structure, and strategies of microbial communities in situ**
- **Use controlled laboratory consortia to identify essential microbial contributions to environmental activities, identify specific and selected interactions, and isolate keystone organisms/processes**
- **Efficiently advance these environmental microorganisms to model organism status and map their molecular functions to community phenotypes and environmental activities.**
- **Development of the LBNL Systems Environmental Microbiology Workbench and Knowledge Framework**

Understanding the composition, structure, and interactions of microbial communities in natural environments over time and space is crucial in microbial ecology. We have developed various metagenomics technologies to characterize microbial community structure. First, based on previous GeoChips, we have developed GeoChip 4.0, a more comprehensive GeoChip to facilitate the analysis of microbial communities from a variety of habitats. GeoChip 4.0 contains 120,054 distinct probes, covering 200,393 genes involved in different functional processes important to biogeochemistry, ecology, environmental sciences and human health. Among these, 36,062 probes are specifically designed for the human microbiome, and cover 47,979 genes in 139 functional gene families involved in 19 functional processes. In addition to updating functional gene families from previous versions of GeoChip with the latest NCBI protein repository, 118 new gene families, belonging to bacteriaphage, stress, and virulence, have been added to GeoChip 4.0 to target more microbially mediated functional processes. As a new version, GeoChip 4.0 was developed on the NimbleGen 12x135K platform so that each chip contains 12 arrays, making it possible to hybridize 12 samples under nearly identical conditions at the same time. Computational evaluation of probe specificity indicated that all designed probes were highly specific to their corresponding targets. Experimental evaluation of specificity, sensitivity and quantification using artificial and environmental samples showed GeoChip 4.0 to be a highly specific, sensitive and quantitative tool for microbial community analysis. GeoChip 4.0 has been used to analyze environmental samples from oil spill sites, soil, and human feces and proven to be a rapid and powerful tool in the study of microbial ecology. Also, a random matrix theory-based (RMT) conceptual framework for identifying functional molecular ecological networks was developed with the high throughput functional gene array hybridization data. Our results indicated that RMT is a powerful method to identify functional molecular ecological networks in microbial communities. Elucidating network interactions in microbial communities and their responses to environmental changes is fundamental in research in microbial ecology, systems microbiology, and global change. In addition, amplicon sequencing approaches have been widely used in microbial ecology, but we have found that the reproducibility and quantitative capability are quite low, primarily due to random sampling. Various approaches have been developed to

predict and minimize the artifacts associated with random sampling processes. This study will have substantial impacts on microbial ecology because this problem is associated with all current sequencing-based metagenomic studies important to energy production, climate change, environmental management, industry, agriculture, and human health. Our future work will focus on continuously updating the GeoChip with more functional processes and more microorganisms covered, developing data analysis pipelines, and using GeoChip data for molecular ecological network analysis to allow more rapid comprehensive analyses of microbial community composition, structure and functions.

This work conducted by ENIGMA 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

# 226

## Microfluidic Tools for Single-Cell Genomic Analysis of Environmental Bacteria

Peng Liu,[1] Robert J. Meagher,[1] Yooli K. Light,[1] Suzan Yilmaz,[1] Romy Chakraborty,[2] **Adam P. Arkin**,[3] **Terry C. Hazen**,[2] and **Anup K. Singh**[1]* (aksingh@sandia.gov)

[1]Biotechnology and Bioengineering Department, Sandia National Laboratories, Livermore, Calif.; [2]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; and [3]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.

**Project Goals: The goal of this project is to develop technologies for genomic analysis of single bacterium found at Hanford site to allow us 1) elucidate the genetic diversity of as-yet uncultivated microorganisms and 2) link function to species, a feat not achievable by current metagenomic techniques.**

Current metagenomic techniques (e.g., microarray or 16s rRNA sequencing) relying on pooled nucleic acids from lysed bacteria can independently measure metabolic activity and the species present, but can not link the activity deterministically to the species . We are developing high-throughput tools for studying bacteria one cell at a time, allowing us to unravel the complicated dynamics of population, gene expression, and metabolic function in mixed microbial communities. Our approach includes FISH-based identification of desired species, enrichment by cell sorting, followed by single-cell encapsulation, whole genome amplification and sequencing. Encapsulation of bacteria in pico-liter plugs in particular allows us to scale down conventional (microliter-volume) assays, such as WGA, into much smaller reaction volumes better suited to the size of an individual microbe. By dramatically reducing the reaction volume, the effective concentration of template is increased, reducing amplification artifacts that often arise in single-cell reactions carried out at a conventional scale. These technologies are being used to analyze water samples from Hanford site.

This work conducted by ENIGMA 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.

# 227

## High-Throughput Mutagenesis Strategies for Non-Model Microorganisms

Adam Deutschbauer[1]* (AMDeutschbauer@lbl.gov), Grant M. Zane,[2] Samuel Fels,[2] Hannah L. Korte,[2] Morgan N. Price,[1] Jennifer Kuehl,[1] Jason Baumohl,[1] **Adam P. Arkin,[1]** and **Judy D. Wall[2]**

[1]Physical Biosciences Division, Lawrence Berkeley National Laboratory; and [2]Department of Biochemistry, University of Missouri

https://sites.google.com/a/lbl.gov/enigma/

**Project Goals: Our aim it to achieve a deep evidence-based annotation of DOE relevant micoorganisms. To accomplish this we are developing large-scale mutagenesis and phenotyping strategies that are applicable to virtually any microorganism. Our data will be used to fill metabolic holes, uncover functions for hundreds of hypothetical proteins, and to discover novel functional relationships across the entire genome.**

Most genes in ENIGMA-relevant bacteria are poorly annotated and many are hypothetical. To address this challenge, it is imperative that flexible, rapid, and inexpensive experimental approaches are developed to assign gene function on a global scale. Here we describe three efforts to annotate gene function in sulfate-reducing and metal-reducing bacteria using high-throughput mutagenesis and phenotyping strategies: (1) directed, markerless genetic modification in *Desulfovibrio vulgaris* Hildenborough (*Dv*H), (2) parallel analysis of transposon mutants using TagModules, and (3) parallel analysis of transposon mutants by HITS (High-Throughput Insertion Tracking by Deep Sequencing). In collaboration with the Chhabra group, we have optimized construct methodologies and can now systematically generate defined *Dv*H mutants using the markerless approach. Our *Dv*H markerless methodology permits the construction of multiple genetic modifications opening the door for the systematic study of genetic interactions in a sulfate-reducing bacterium. Additionally, the markerless system holds great promise for the functional characterization of polymorphisms in evolved lines of *Dv*H (in collaboration with the Zhou and Stahl labs). Despite these advances, however, it is currently not feasible to pursue whole-genome targeted deletion libraries in all ENIGMA microorganisms. Therefore, to augment our markerless deletion approach, we have pursed transposon mutagenesis with the TagModule collection. Each Gateway-cloned TagModule contains two unique 20 bp DNA tags that permit strain pooling and parallel analysis of tag abundance. We combined the TagModules with transposon mutagenesis to create a library of ~50,000 sequence-verified and tagged mutants in *Shewanella oneidensis* MR-1 (~28K

mutants), *Desulfovibrio alaskensis* G20 (~15K mutants), and *Dv*H (~7K mutants) which, as archived single strains, serve as a rich resource for all ENIGMA collaborators. Furthermore, the presence of the TagModules permits the pooling and parallel analysis of strain fitness for ~4000 mutants by a highly quantitative, inexpensive assay. We used this pooled assay to probe the fitness of 3355 *S. oneidensis* MR-1 genes (~90% of the nonessential genome) in over 100 diverse growth conditions including different growth substrates, alternative electron acceptors, stresses, survival, and motility. We show that the pooled assay has excellent biological consistency, and relative defects as small as a 2% reduction in growth rate per generation can be assayed reliably. We find that ~70% of genes have a pattern of fitness that is significantly different from random including hundreds of hypothetical genes, and ~37% of genes have a strong enough signal to show strong biological correlations. Using fitness patterns, we were able to propose specific molecular functions for 28 genes or operons that lacked specific annotations or had incorrect annotations including a previously unknown acetylornithine deacetylase. While the TagModule approach described above was accomplished by a single laboratory and can be generally applied to create a large-scale gene-phenotype map in most microorganisms, there is still an upfront investment required to generate the initial mutant strains. To accelerate the analysis of large transposon libraries in additional ENIGMA microorganisms, we are complementing the TagModule approach with the HITS method that does not require the up-front effort to archive single mutants. Preliminary work on developing HITS in *Dv*H is presented.

# 228

## Subcellular Localization of Proteins in the Anaerobic Sulfate Reducer *Desulfovibrio vulgaris* via SNAP-Tag Labeling and Photoconversion

A. Gorur,[1] C.M. Leung,[1] S. Chhabra,[1] T. Juba,[1] A. Tauscher,[1] S. Reveco,[1] J.P. Remis,[1] B. Lam,[1] J.T. Geller,[1] T.C. Hazen,[1] M. Biggin,[1] J.M. Chandonia,[1] K.H. Downing,[1] J.Wall,[1,2] and M. Auer[1]* (mauer@lbl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley Calif.; and [2]University of Missouri, Columbia

**Project Goals: Protein localization and expression studies in planktonic cells and biofilms of *Desulfovibrio vulgaris* Hildenborough (DvH) under baseline and environmentally relevant stress conditions**

A systems biology understanding of microbes in a planktonic state and in biofilms requires the mapping of spatio-temporal distributions of macromolecules correlated to microbial activity. As part of the DOE-funded ENIGMA program, we study protein expression and localization in the anaerobic soil bacterium *Desulfovibrio vulgaris Hildenborough* (DvH),which plays a prominent role in bioremediation of DOE legacy sites by reducing and therefore immobilizing radionuclide and other toxic heavy metals in plumes therefore preventing these metals from reaching human water supplies. . Our goal is to study protein abundance and localization at the optical level as well as the EM level using cryo-tomography of labeled, photoconverted and vitrified whole-mounts as well as FIB/SEM of resin-embedded samples.

We have chosen the commercially available AGT-tag to label proteins as—unlike GFP and derivatives—it allows labeling under anaerobic conditions. This system is based on a modified O6-alkylguanine-DNA alkyltransferase (AGT) tag that undergoes a dead–end chemical reaction with a modified O6-benzylguanine (BG) derivative. This SNAP label has been conjugated to a large number of fluorophores and other biochemically functional groups allowing flexibility in experimental design.

The tagged strains that are generated using SLIC and Gateway approaches are labeled with commercially available SNAP fluorophores. After extensive optimization we have obtained robust protocols that are virtually background-free and that allow high-throughput imaging and protein expression quantification. Using deconvolution microscopy we have studied ~20 tagged strains and found a significant number to display discrete non-uniform localization patterns. For example, we found ParA, Mot-A and Mot A-1 to localize exclusively to the poles, while others such as Lyt R, FtsH, GlnA, ModA, FlgE and UvrB localize both to the poles and to secondary regions within the cell. Proteins showing a patchy or spotty distribution along the length of the cell include hup-3 and PyrB. As expected the majority of proteins display uniform distribution. We have further developed labeling and photoconversion approaches that allow visualization of protein location in the context of cellular ultrastructure and should allow us to examine its relationship to extracellular metal reduction activity that we discovered to be localized to discrete sites on the outer membrane surface. Only a subset of cells in planktonic state or in biofilms showed metal deposits on the cell surface suggesting that despite seeing the same microenvironment cells differ in their protein inventory and possibly metabolic state.

Acknowledgements/References**:** This work conducted by ENIGMA 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

## High Throughput Identification of Protein Complexes from *Desulfovibrio vulgaris* by a Tandem Affinity Purification Pipeline

**Gareth P. Butland**[1]* (GPButland@lbl.gov), **Swapnil R. Chhabra,**[1] Barbara Gold,[1] Nancy L. Liu,[1] Sonia Reveco,[1] Tom R. Juba,[2] **Judy D. Wall,**[2] Bonita R. Lam,[1] Jil T. Geller,[1] **Terry C. Hazen,**[1] Megan Choi,[1] **Mark D. Biggin**,[1] Evelin D. Szakal,[3] Simon Allen,[3] Haichuan Liu,[3] **H. Ewa Witkowska**,[3] and **John-Marc Chandonia**[1]

[1]Lawrence Berkeley National Lab, Berkeley, Calif.; [2]Univ. of Missouri, Columbia; and [3]Univ. of California, San Francisco

**Project Goals: see below**

ENIGMA's goal is to understand, at a molecular systems level, the bacterial soil communities at DOE sites contaminated with heavy metals or radionuclides. Sulfate reducing bacteria (SRBs) are key members of these communities and can reduce many of the contaminating elements to an insoluble form. Environmental change or human intervention will alter the chemical environment in the subsoil, which in turn affects which species predominate as well as microbial physiology. Therefore, it will be critical to learn how such changes affect SRBs and their interaction with other members of the community and the biogeochemistry. We have chosen *Desulfovibrio vulgaris* to address these questions in molecular detail as it is one of the dominant sulfate reducers found at DOE sites.

Most cellular processes are mediated by multiple proteins interacting with each other in the form of multi-protein complexes and not by individual proteins acting in isolation. In order to accurately model cellular processes in this SRB and how they respond to stress, our goal is to develop a comprehensive knowledgebase of protein complexes and protein-protein interactions in *D. vulgaris* using high throughput tandem affinity purification (TAP). Our approach utilizes the Sequence and Ligation Independent Cloning (SLIC) technique to generate custom suicide constructs in high throughput. Utilizing SLIC, we have achieved success rates for suicide construct generation of greater than 85% and following introduction of constructs into *D. vulgaris* ~80 % of isolates were found to express a TAP-tagged fusion protein by IP-western. Currently, we have generated 687 TAP-tagged *D. vulgaris* strains and for the last six months have been able to generate 50 new TAP-tagged strains per month.

To date, 357 unique *D. vulgaris* strains containing correctly integrated TAP-tagged chromosomal fusions have been subject to TAP analysis and the composition of purified eluates analyzed by mass spectrometry. In 291 of these analyses, the bait was verified to be present by gel-free mass spectroscopy. In these experiments, a total of 5,944 interactions were detected with 1,060 distinct prey proteins. Using curated gold standard datasets, we filtered out ubiquitous proteins and other likely false positives, resulting in a set of 293 high-confidence interactions between 89 baits and 246 preys. 38 interactions have been reciprocally confirmed, using strains in which the original prey protein was tagged and used as bait. Detected high-confidence interactions cover a range of biological processes including energy conservation (Hydrogenase(s), Dissimilatory Sulfite Reductase), protein secretion (YajC-HflCK complex), protein folding (DnaK-DnaJ-DafA complex) and cofactor biosynthesis (Heme and FeS clusters) and include both novel and previously predicted interactions. ENIGMA has also identified a large number of protein-protein interactions in *D. vulgaris* using a tagless approach, and we are integrating the analyses of these data with each other and with other large-scale ENIGMA datasets (e.g., fitness and gene expression data) in order to increase the number of high-confidence interactions. Throughout the project, we have removed many bottlenecks associated with working with *D. vulgaris* and this has enabled us to obtain throughput statistics, data quality and success rates similar to those previously reported for *E. coli*. Our rate of TAP analysis had been limited by the larger culture volumes required for *D. vulgaris* compared to *E. coli*. Improvements in biomass yield, purification processes and mass spectrometry technology have recently enabled a shift to processing 20 strains per week. We are now in a position to conduct a system- wide analysis of all stable protein-protein interactions in *D. vulgaris* and to target how these change in response to stresses typically occurring in the subsoil of contaminated sites for a select set of stress response genes.

This work conducted by ENIGMA 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.

# 230

## Accurate, High-Throughput Identification of Stable Protein Complexes in *Desulfovibrio vulgaris* using a Tagless Strategy

John-Marc Chandonia[1,3]* (JMChandonia@lbl.gov), Ming Dong,[1] Maxim Shatsky,[1,3] Haichuan Liu,[2] Lee Yang,[1] Terry C. Hazen,[1] Jil T. Geller,[1] Megan Choi,[1] Evelin D. Szakal,[2] Simon Allen,[2] Steven E. Brenner,[1,3] Steven C. Hall,[2] Susan J. Fisher,[1,2] Sunil Kumar,[4] Farris L. Poole,[4] Michael Adams,[4] Jian Jin,[1] H. Ewa Witkowska,[2] Adam P. Arkin,[1,3] and **Mark D. Biggin**[1]

[1]Lawrence Berkeley National Laboratory; [2]University of California, San Francisco; [3]University of California, Berkeley; and [4]University of Georgia

http://enigma.lbl.gov

**Project Goals: We describe a novel "tagless" method for identification of stable, soluble protein complexes that is general to all cultureable microbes and does not require genetic manipulation of the organism. Our strategy is based on the premise that the great majority of such com-**

**plexes will survive intact through a series of orthogonal chromatographic steps, with complex components having correlated elution profiles. We demonstrate the effectiveness of this method in *D. vulgaris*.**

*Desulfovibrio vulgaris* has been selected as a model bacterium for intensive study by ENIGMA because it can reduce heavy metals and radionuclide contaminants present in the soil at many DOE sites, rendering the contaminants insoluble. ENIGMA seeks to model, at a molecular systems level, how this and similar bacteria respond to natural and human induced changes in their environment and how this alters their ability to stabilize contaminants in the soil. A major component our strategy is to develop and use high throughput pipelines to purify and identify protein complexes and to structurally characterize them by EM. Most cellular processes are mediated by multiple proteins interacting with each other in the form of multi-protein complexes and not by individual proteins acting in isolation. Thus, for systems modeling it is critical to characterize protein complexes genome-wide and determine how their composition and structures change with the environment.

We describe a novel "tagless" method for identification of stable, soluble protein complexes that is general to all cultureable microbes and does not require genetic manipulation of the organism. Our strategy is based on the premise that the great majority of such complexes will survive intact through a series of orthogonal chromatographic steps, with complex components having correlated elution profiles. A major challenge is the potential for false positives (FP) caused by co-elution of proteins that are not part of a complex. Approximately 10 g soluble protein from a crude *D. vulgaris* extract have been separated using ammonium sulphate precipitation and a series of three highly parallel chromatographic steps. For the last step, 306 size exclusion columns have been run, yielding 6,859 fractions. The elution profiles of each protein across each of these columns have been measured with the aid of mass spectrometry and iTRAQ reagents (Dong et al., 2008, J Proteome Res. 7:1836-49) leading to the identification of 1,444 proteins (~40% of the proteome). For every region of elution space where two proteins overlap, Pearson correlation coefficients have been calculated between vectors of normalized relative protein amounts estimated using iTRAQ. These data were used to train a random forest classifier to identify true interactions in a manually curated gold standard (GS) set. We compare the discriminating power of these proteomic data to that of other high-throughput data, such as correlation of gene expression profiles. Our method is able to identify 66% of GS interactions present in our proteomic data at a 0% FP rate. Using the same thresholds results in the prediction of 854 novel interactions. Thus, this strategy is effective at identifying a subset of stable inter-protein interactions in a bacterial proteome at high precision.

In addition, we selected 16 complexes identified by the above fractionation strategy with molecular weights 400 - 1,000 kDa and provided them to the ENIGMA single-particle EM group. This resulted in the structures of 7 complexes being solved and showed that there are a surprisingly large number of differences in the quaternary structures of *D. vulgaris* complexes isolated from compared to those of homologous proteins from other microbes (Han et al., 2009, PNAS 106, 16580); see Han et al, poster.

We have also begun to measure the metal content of each of the size exclusion column fractions and compare the results to our iTRAQ quantification of polypeptides to identify metalloproteins; see Menon et al poster.

In the future, we plan to complete our analysis of the full *D. vulgaris* proteome and metalloproteome and study how interactions change under stress conditions that mimic those that commonly occur in contaminated soils. We will also extend our analyses to include other high-throughput datasets produced by ENIGMA, e.g., the protein interactions we have discovered using parallel Tandem Affinity Purification data from *D. vulgaris*; see Butland et al poster.

Funding: ENIGMA is a Lawrence Berkeley National Laboratory Scientific Focus Area Program supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL Foundational Science through contract DE-AC02-05CH11231

# 231

## EM Structural Survey of Large Protein Complexes in *Desulfovibrio vulgaris* and EM High Throughput Pipeline Development

Bong-Gyoon Han[1]* (BGHan@lbl.gov), Ming Dong,[1] Maxim Shatsky,[1,2] Pablo Arbelaez,[2] Jitendra Malik,[2] Dieter Typke,[1] Ross Walton,[1] Amos Song,[1] Steven Yannone,[1] Kenneth H. Downing,[1] Mark D. Biggin,[1] and Robert M. Glaeser[1]

[1]Lawrence Berkeley National Laboratory, Berkeley Calif.; and [2]University of California, Berkeley

**Project Goals: ENIGMA is conducting a systems level characterization of *Desulfovibrio vulgaris* to understand the role of sulfate reducing bacteria in reducing metals in contaminated DOE sites. As part of that effort, we are establishing methods to structurally characterize multi-protein complexes by single-particle electron microscopy.**

Protein samples were purified by a tagless strategy to carry out an unbiased survey of the stable, most abundant multi-protein complexes in *Desulfovibrio vulgaris* Hildenborough (*Dv*H) that are larger than Mr ~400 kD. The quaternary structures for 8 of the 16 complexes purified during this work were determined by single-particle reconstruction of negatively stained specimens (Han et al., 2009, PNAS 106, 16580), a success rate about 10 times greater than that of previous "proteomic" screens. In addition, the subunit compositions and stoichiometries of the remaining complexes were determined by biochemical methods. Our results show that the structures of large protein complexes vary to a surprising extent from one microorganism to another. Except for GroEL and the 70S ribosome, none of the 13 remaining complexes with known orthologs have quaternary

structures that are fully conserved. This result indicates that the interaction interfaces within large, macromolecular complexes are much more variable than has generally been appreciated. As a consequence, we suggest that relying solely on quaternary structures for homologous proteins may not be sufficient to properly understand their role in another cell of interest. The diversity of subunit stoichiometries and quaternary structures of multi-protein complexes that has been observed in our experiments with *Dv*H is relevant to understanding how different bacteria optimize the kinetics and performance of their respective biochemical networks.

Conventional single particle EM methods have not previously been able to solve protein structures rapidly enough to handle the sheer volume of protein samples produced by ENIGMA . Therefore, we have reduced the data processing time two fold by automating data collection. To further increase throughput we are implementing automated data analysis, model building from the low tilt angle image pairs and the engineering of new support-film technologies for EM sample preparation. The latter is driven by the need, encountered within this high-throughput project, for technologies that do not require sample-dependent optimization and are more likely to preserve quaternary structure in a conformationally homogeneous state.

# 232

## High Throughput Production and Analysis of Genetically Engineered *Desulfovibrio vulgaris* Strains via Homologous Recombination

**PIs participating:** Swapnil Chhabra[1]* (srchhabra@lbl.gov), Manfred Auer,[1] Gareth Butland,[1] John Marc Chandonia,[1] Terry Hazen,[1] Judy Wall,[2] Ewa Witkowska,[2] Dwayne Elias,[2] Michael Adams,[3] Matthew Fields,[4] Jan Liphardt,[1] Greg Hura,[1] and David Stahl[5]

[1]Lawrence Berkeley National Laboratory; [2]University of Missouri; [3]University of Georgia; [4]Montana State University; and [5]University of Washington

The primary focus of ENIGMA is to understand, at a basic systems level, bacterial communities in the sub soil of DOE sites contaminated with heavy metals or radionuclides. Sulfate reducing bacteria play a critical role in these communities and can directly reduce many of the contaminants to an insoluble form. Conditions in the sub soil are not static, however. Environmental or human intervention can alter oxygen, nitrate or salinity levels etc, which in turn affects which species predominate and details of their physiology. Understanding how microbes respond to such changes and how sulfate reducers interact with the other members of the community is critical if we are to model how communities cope with such change and interact with the biogeochemistry. We have selected *Desulfovibrio vulgaris,* one of the dominant sulfate reducers found at DOE sites, to understand these processes in molecular detail. For our work, we require many genetically engineered strains in which either affinity tag DNA sequences are introduced into genes, or the activity of the gene is altered by targeted mutation. The ability to modify genomes by making such locus-specific chromosomal alterations in a high-throughput and cost-effective manner has been successfully applied in yeast and *Escherichia coli*, but prior to our work, it was extraordinarily difficult to modify even a few genes in *D. vulgaris* in this way. Indeed, a diverse range of other bacteria of importance to DOE's mission have been similarly difficult to modify.

Therefore, we have developed a method for high-throughput targeted manipulation of genes in *D. vulgaris* that is both inexpensive and flexible due to the use of interchangeable "parts" for making different kinds of chromosomal modifications, including gene deletions, tagged genes for the study of protein-protein interactions and protein localization to name a few. This systematic approach for chromosomal modification can be applied to a wide range of bacteria with minimal need for methodological alteration and relies on the facile construction of suicide vectors through the use of high-throughput methods, including Sequence and Ligation Independent Cloning (SLIC), heretofore used for plasmid-based (rather than chromosomally based) metabolic engineering and heterologous protein expression studies. Our procedures generate tagged genes or marker exchange deletions by double homologous recombination events. For tagged genes, this protocol ensures that a single copy of the gene with the tag is produced from its natural promoter and, in most cases, without polarity. For deletion construction, sequences of the targetted gene are removed from the cell, preventing rearrangements that could restore a functional gene.

Prior to our work, it was only possible to produce a handful of homologous recombination targeted mutations in *D. vulgaris* per year, with the attempts failing for most genes. With our new strategy, we have been able to produce 50 strains per month for the last 6 months and a total of ~762 strains over all. Importantly, over 79% of attempts to modify genes have been successful. Thus, we can now target most genes of importance in various stress responses and are in a position to conduct a full genome-wide analysis. We have engineered 687 strains for Tandem Affinity Purification, which are being successfully used for protein/protein interactions analyses and structural characterization of proteins. A further 75 strains have been constructed for mapping the location of protein complexes within the cell by either deconvolution or single molecule microscopy. In addition, our method will help fill gaps in metabolic pathways and greatly assist in the functional annotation of unknown genes. Our goals for the next few years are to produce strains for the mapping the full *D. vulgaris* interactome and localizome, to assist further annotation of the genome, and to initiate examination of protein complex remodeling in response to environmental stresses.

# 233

## High-Throughput Pipeline for the Purification and Identification of *Desulfovibrio vulgaris* Membrane Protein Complexes

Peter J. Walian[1]* (PJWalian@lbl.gov), Simon Allen,[2] Lucy Zeng,[1] Evelin D. Szakal,[2] Haichuan Liu,[2] Steven C. Hall,[2] Susan J. Fisher,[1,2] Ralph Santos,[1] Bonita Lam,[1] Jil T. Geller,[1] **Terry C. Hazen,[1] John-Marc Chandonia,[1] H. Ewa Witkowska,[2] Mark D. Biggin**,[1] and **Bing K. Jap[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif.; and [2]University of California, San Francisco

http://enigma.lbl.gov

**Project Goals: To develop and apply a pipeline for the high-throughput isolation and identification of *Desulfovibrio vulgaris* membrane protein complexes in cultures grown under standard conditions, and to characterize changes in these complexes brought about by environmentally relevant stressors.**

The ability of the Gram-negative sulfur-reducing bacterium *Desulfovibrio vulgaris* to reduce heavy metals makes it an ideal candidate for studying how bacteria can influence the biogeochemistry in the subsoil at DOE contaminated sites. Knowledge of this organism's molecular level responses to environmental changes will be essential for accurately modeling its stress response pathways and understanding how it behaves under a range of chemical environments. To obtain this knowledge we have developed a high-throughput pipeline for the isolation and identification of untagged membrane protein complexes. Membrane proteins are especially of interest, as they serve at the interface of cell-cell communications and coordinate interactions with the extracellular environment. It is widely recognized, however, that these proteins are particularly challenging to purify and characterize, and that the use of an inappropriate detergent or detergent-to-protein ratio, for example, can lead to the disruption of complexes or the formation of non-biologically relevant aggregates. The relatively low yield of *D. vulgaris* cells per liter of culture (about one-tenth that of *E. coli*) presents an additional challenge. To address these challenges, we have developed a unique processing pipeline that features mild, but effective, detergent solubilization, liquid chromatography and native electrophoresis methods. Large-scale cultures up to 100 liters in size have been processed. This "tagless" strategy can provide uniquely global views of changes in membrane protein populations from cultures grown under a variety of conditions. Additionally, a distinct advantage of the tagless approach is that it does not require genetic manipulation which can invoke a steep and time-consuming learning curve when tackling organisms where there is minimal previous genetic experience. Thus, our method should be general to a wide array of microbes of interest to the DOE but for which to date facile genetic manipulation is not possible.

Our tagless study of the membrane protein complexes of the *D. vulgaris* outer-membrane is complete and we are at an advanced stage of data collection for the inner-membrane component with over 1000 samples being analyzed by mass spectrometry per growth condition. Outer-membrane proteins provide the front line of defense for Gram-negative bacteria such as *D. vulgaris* and are expected to readily yield a variety of membrane protein changes in response to environmental stressors. We have identified 70 outer-membrane proteins derived from cells grown to mid-log phase under standard conditions representing a coverage of 82% of those predicted to be encoded by the genome. Of these, at least 50 appear to be in some form of complex (homo- or heteromeric). This list will serve as our reference dataset with which we can assess the nature of a cell's membrane protein-based response to a variety of stresses. Processing of outer-membrane proteins from stressed cultures (to include growth to stationary phase, and growth under elevated levels of nitrate or NaCl) is now underway. Exciting preliminary results provide a global view of a large number of stress-induced changes in outer-membrane proteins and demonstrate the potential of this approach. Taken together, these results indicate that this tagless technique will be an effective tool for revealing changes in *D. vulgaris* membrane proteins arising from environmental stress. We propose to extend our studies to additional stresses that have also been examined by other system wide methods within ENIGMA, and to conduct a comprehensive characterization of the hundreds of inner-membrane proteins encoded by *D. vulgaris*.

# 234

## ENIGMA-MS: Using Stable Isotopes and Novel Metabolomic Technologies to Advance Our Understanding of Microbial Metabolism

Richard Baran[1]* (RBaran@lbl.gov), Peter Benke,[1] Edward Baidoo,[1] **Jay Keasling,[1] H. Gary Siuzdak,[2]** and **Trent Northen**[1]

[1]Lawrence Berkeley National Lab, Berkeley Calif.; and [2]The Scripps Center for Mass Spectrometry, La Jolla, Calif.

http://enigma.lbl.gov

**Project Goals: A microorganism-based approach to address DOE mission goals in remediation, carbon sequestration and energy production will require quantitative understanding of biological complexity at multiple scales — from molecular networks of individual species to the dynamic inter-species interactions within the communities in which they reside. The broad goals of ENIGMA are to understand biological complexity by discovering and predictively modeling interactions within microbial and community processes that drive complex geochemistries in key environments. In doing so we expect to define biological principles governing selection of microbial community function and composition in given environments.**

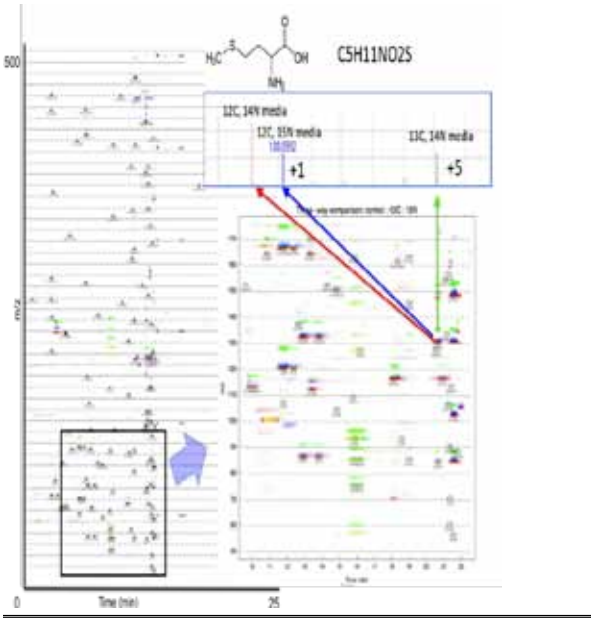



Figure 1. Stable Isotope Labeling for metabolite profiling

Metabolite profiling using mass spectrometry provides an attractive approach for the interrogation of cellular metabolic capabilities. Untargeted metabolite profiling using ElectroSpray Ionization (ESI) has the potential to identify numerous novel metabolites, however, de novo identification of metabolites from spectral features remains a challenge given the large number of experimental artifacts. The ENIGMA MS group has developed and reported an integrated workflow for metabolite identification using uniform stable isotope labeling. Metabolite profiling of cell and growth media extracts of unlabeled control, $^{15}N$, and $^{13}C$-labeled cultures of the non-filamentous cyanobacterium, *Synechococcus sp. PCC 7002* was performed using normal phase liquid chromatography coupled to mass spectrometry (LC-MS). Visualization of three-way comparisons of raw datasets highlighted characteristic labeling patterns for metabolites of biological origin allowing exhaustive identification of corresponding spectral features (Fig. 1). Additionally, unambiguous assignment of empirical formulas was greatly facilitated by the use of stable isotope labeling. Empirical formulas of metabolites responsible for redundant spectral features were determined and fragmentation (MS/MS) spectra for these metabolites were collected. Analysis of acquired MS/MS spectra against spectral database records led to the identification of a number of metabolites absent not only from the reconstructed draft metabolic network of *Synechococcus sp. PCC 7002*, but not included in databases of metabolism.

This work has recently been extended with systematic screening of consumed or excreted metabolites using metabolite profiling of growth media from microbial cultures (referred to as metabolic footprinting). We performed a systematic evaluation of exchange of metabolites between a *Synechococcus sp. PCC 7002* and different growth media using metabolomics. It was found that 102 out of 202 detected metabolites were exchanged significantly. This metabolic footprinting approach is being extended to study interactions between different organisms. In addition, the presence of membrane transport activities for specific metabolites was highlighted and can enable the search for corresponding transport proteins.

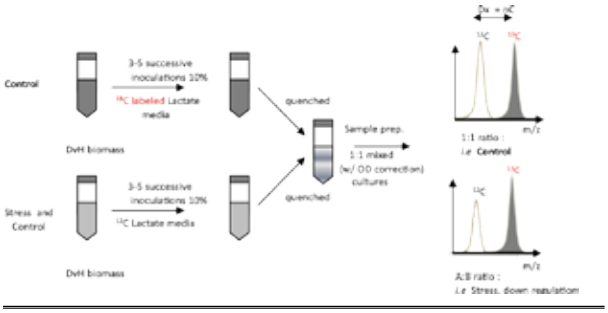



Figure 2. Stable Isotope Labeling based quantification

In addition to facilitating identification of compounds and their exchange with the environment, growth on stable istopic media is used to quantify microbial responses to environmental stresses. Since the quantification of a large number of metabolites is impractical and time consuming, our laboratory has designed an isotope dilution mass spectrometry (IDMS) strategy (Fig. 2) that improves upon precision (as well as on identification, as previously described) and hence simplifies quantification of microbial responses to stress. This strategy involves the mixing of *D. vulgaris* cultures grown side-by-side on the unlabeled (stress) and $^{13}C$ uniformly labeled (control) form of the same carbon source (lactate). Consequently, the comparison of the $^{12}C$ and $^{13}C$ fully-labeled metabolites is representative of stressed versus control biomass. The mixing of the quenched cultures prior further sample preparation enhances the precision of the measurement, which is imperative to quantification with high degree of confidence. Currently, we have achieved ~80% labeling efficiency after three consecutive inoculums (e.g. initiate the feeding of labeled lactate after the cells have entered a period of starvation). However, it is not necessary to achieve 100% $^{13}C$ conversion, if reproducible labeling efficiencies for all metabolite of interest can be achieved.

# 235

## ENIGMA-MS: Protein Identification, Quantitation and Structural Characterization

Simon Allen,[1] Megan Choi,[2] Haichuan Liu,[1] Christopher Petzold,[2] Max Shatsky,[2] Evelin D. Szakal,[1] Sunia A. Trauger,[3] Lee Yang,[2] Ming Dong,[2] Peter Walian,[2] Steve S. Yanonne,[2] Lucy Zeng,[2] Rich Niles,[1] Susan J. Fisher,[1] Steven C. Hall,[1] **John-Marc Chandonia**,[2] **Jian Jin**,[2] **Gareth Butland**,[2] **Bing Jap**,[2] Paul Benton,[2] Benjamin Bowen,[2] Farris Poole,[2] **John A. Tainer**,[2,3] **Mark Biggin**,[2] **Trent Northen**,[2] **Michael Adams**,[4] Aindrila Mukhopadhyay,[2] **Gary Siuzdak**,[2,3] and **H. Ewa Witkowska**[1]* (witkowsk@cgl.ucsf.edu)

[1]University of California, San Francisco; [2]Lawrence Berkeley National Lab; [3]The Scripps Research Institute; and [4]University of Georgia

**Project Goals: The overarching goal of the MS Proteomics component of ENIGMA is to develop and provide MS tools for a comprehensive characterization of proteomes and interactormes of bacteria, with a focus on addressing key challenges that are presented by the bioremediation needs at metal-contaminated sites, e.g., Hanford. To this end, high throughput pipelines for fractionation and identification of protein complexes were introduced and successfully executed for DvH, P. furiosus and S. solfataricus. In addition, the MS-based assays aimed at detailed characterization of protein interactions with protein and/or non-proteinaceous partners are being developed, e.g., metalloproteomics, identification of proteins interacting with specific ligands and analysis of intact protein complexes, the latter to establish their stoichiometries and architecture. Ultimately, in collaboration with other components of ENIGMA, we plan to develop multi-omic platforms capable of integrated analysis of different factors.**

Understanding protein – function relationships and protein interactions with other cell components is critical to ENIGMA's experimental goals. Robust mass spectrometry (MS) platforms provide the primary analytical techniques for identification, quantitation and characterization of bacterial proteomes. Our group has implemented novel experimental workflows, introduced automated high-throughput routines to enhance MS data acquisition, developed new technologies to increase sensitivity and broaden applicability of MS techniques to protein-associated species, and built computational and bioinformatics tools for data analysis and interpretation. Protein identification is performed routinely using gel- and solution-based liquid chromatography (LC) MS workflows. A number of specialized technologies that apply the power of MS for detection and identification of protein-protein (tagless[1] and qTagless[2]), protein-metal (metalloproteomics)[3] and protein-ligand[4] interactions were developed and applied to various bacterial species. The latter technologies represent integration of metabolomics and proteomics strategies.

Two major workflows are currently available for protein identification (ID): gel- and solution-based LC MS. In the gel-based workflow, final protein separation is performed using electrophoresis (SDS PAGE), visible bands (stained with Coomassie Blue or silver) are excised, proteins are in-gel digested with trypsin in a 96-well format by a robotic platform, and resulting peptides are analyzed via 1D (low pH reversed phase) nanoLC ESI MS, primarily using a Thermo LTQ mass spectrometer operated in a data-dependent mode. Batch data analysis is performed using the Mascot search engine. A variant of the gel-based LC workflow that employs gel-free electrophoresis for protein separation was developed at LBNL: it offers the ability to further automate the process of sample preparation for MS[5].

The gel-based LC workflow is routinely used in a number of ENIGMA projects. Specifically, it has been applied in conjunction with metalloproteomic studies to map metal-binding protein complexes in bacterial cell lysates of *Pyrococcus furiosus*. Diethylaminoethyl anion exchange (DEAE) column chromatography-fractionated proteins were analyzed to derive protein IDs and in parallel, with inductively coupled plasma MS to detect and quantitate metal ions present in protein complex-containing fractions. Using this platform, over 7000 samples have been analyzed to identify protein complexes and unique metalloproteins. Gel-based LC MS also serves as a primary workflow for identification of outer- and inner-membrane protein complexes in *Desulfovibrio vulgaris* Hildenborough (*Dv*H). Fractionation of inner and outer membrane complexes in *Dv*H employs a combination of chromatographic and electrophoretic (Blue Gel) steps, with a final separation of protein complex components using 1D SDS PAGE that is followed by MS protein identification using either MALDI or ESI-based methods. To date, ~4000 membrane samples were analyzed. The first survey of protein complexes present in the outer membrane of *Dv*H is currently being prepared for publication. Furthermore, the gel-based LC workflow also serves as an adjunct technique in analysis of pulldowns derived from tandem affinity purification (TAP) experiments.

For solution-based LC MS workflows, tryptic digestion is performed using either classical solution digestion protocols or on a PVDF membrane in a 96-well format. Peptides are analyzed either by 1D or 2D LC MS, the latter utilizing cation exchange or reversed phase separation at basic pH as the first dimension. Solution-based 1D LC MS is also used for analysis of protein mixtures derived from the TAP (>380 baits analyzed so far) and qTagless strategy workflows for the identification of the soluble *Dv*H protein complex components. In addition, 2D LC methods are used for an exhaustive proteomics surveys of soluble and membrane compartments of a bacterial cell.

Different approaches to a tagless, *i.e.*, non-targeted, analysis of protein complexes in bacteria have been utilized and customized for soluble and membrane compartments in different types of bacteria. The protein complex mapping of *P. furiosus* and *S. solfataricus* employed a multidimensional separation of biomass under native conditions and monitoring the protein content of chromatographic fractions by a

combination of 1D SDS PAGE and LC MS utilizing the Thermo LTQ mass spectrometer. To analyze soluble protein complexes in *Dv*H, a quantitative qTagless strategy was developed using MS tools to track protein elution through the final steps of a multi-dimensional protein separation space. To this end, dense sampling of protein fractions collected at the size exclusion step of protein fractionation was performed and relative concentrations of each polypeptide across the separation column were measured with the aid of iTRAQ reagents[2]. The derived polypeptide elution profiles were subjected to computational analysis to assign probabilities for "true" protein complex components vs. "opportunistic" coeluters. Currently, all MS analyses for the qTagless strategy utilize a LC MALDI MS/MS workflow (AB 4800 TOF/TOF mass spectrometer). To enhance throughput of the qTagless strategy, a miniaturized protocol for protein digestion and peptide labeling with iTRAQ reagents in the 96-well PVDF membrane plate format was introduced. In addition, automated iterative MS/MS acquisition (IMMA) software was developed to increase the efficiency of protein identification in LC MALDI MS/MS workflows.

The current proteomics workflows that focus on elucidation of changes in the repertoire and in the level of protein expression are not well suited to tackle the subtle, often non-stoichiometric alterations in protein structure, *e.g.*, post-translational modifications (PTMs). At the same time, genomic and transcriptional analyses provide little help in discerning their presence and localization. While a number of approaches will be necessary to provide truly comprehensive protein characterization, we propose to focus on the following areas: (1) targeted and quantitative analysis of phosphorylation and glycosylation as likely drivers of protein function, (2) analysis of intact proteins and protein assemblages as an entry to protein population studies, and (3) integration with stable isotope probes to understand the dynamics of protein expression and link with metabolic capabilities. To this end, the results of pilot analyses of intact protein complexes by native MS are very encouraging. In a published study of PTMs, trimethylation was observed in a number of proteins engaged in sulfate reduction in *Dv*H[6]. Accordingly, there is a need for new technologies that will enable multifaceted characterization of proteins of interest and ultimately, entire proteomes to capture microheterogeneity of structures that can be linked to function. Development of MS tools for targeted PTM discovery, analysis of intact proteins and their interactions with protein and non-protein (*e.g.*, metals, ligands) partners will be prioritized for development by the ENIGMA-MS group.

**References**


1. Menon et al., 2009. *Mol. Cell. Proteomics* 2009, 8:735-751.
2. Dong et al., 2008. *J Proteome Res.* 7:1836-49.
3. Cvetkovic et al., 2010. *A Nature* 466:779-82.
4. Kalisiak et al., 2009. *J. Am. Chem. Soc.* 131:378-386.
5. Choi et al., 2010. *Electrophoresis* 31: 440-7.
6. Gaucher et al., 2008. *J Proteome Res.* 7: 2320-31.


# 236

## Deconvoluting Signal From Noise: Deciphering Biological Functions and Interactions


Ben Bowen[1]* (bpbowen@lbl.gov), Marcin Joachimiak,[1] David J. Reiss,[3] Morgan Price,[1] John-Marc Chandonia,[1] Paramvir Dehal,[1] Gary Siuzdak,[4] Trent Northen,[1] Adam Arkin,[1] and Nitin Baliga[3]

[1]Lawrence Berkeley National Lab; [2]University of California - Berkeley; [3]Institute for Systems Biology; and [4]The Scripps Research Institute


**Project Goals: Overview of the computational tools within the ENIGMA SFA**

ENIGMA is at the forefront of systems biology of microbes and their communities. In systems biology, computation has a crucial role in processing large amounts of data to construct a quantitative and predictive understanding of biological function at multiple scales. Initially, our algorithms analyze raw data from a diverse array of high-throughput technologies such as mass spectrometry, sequencing, and high density microarrays to yield quantitative measurements of sequence variations, transcripts, proteins, and metabolites. At an intermediate level, our algorithms integrate these data to find statistically significant patterns over multiple dimensions of environmental space and time. These patterns reveal biodiversity in a community, genome organization in a microbe, transcriptome structure and regulation, protein-protein, regulons, and novel metabolic capabilities. One level up, we are inferring organizational principles that relate the behavior of organisms in a community, and the functioning of regulons and protein complexes within metabolic and regulatory networks. We illustrate examples of efforts within ENIGMA that span this continuum of algorithm development across multiple scales of systems biology:

**TIER 1: PROCESSING AND ANALYSIS OF RAW DATA.** The ENIGMA project has generated gene expression, gene fitness, proteomic, metabolomic, and protein-protein interaction data among others. A variety of approaches including associative biclustering have found relationships in different biological contexts including: transcription regulatory networks, protein-protein interaction networks, and metabolic pathways. These analysis have led to numerous discoveries, but they are all rooted in the correct handling of complex datasets that present technical and scientific challenges to process.

**TIER 2: STATISTICAL PATTERN IDENTIFICATION OF SPATIO-TEMPORAL PHENOMENA.** With the ongoing deluge of functional genomics data it has become advantageous to: a) simultaneously query multiple data types, b) jointly determine confidence across data layers, and c) systematically form hypothesis from multiple types of evidence. These challenges are met with approaches involving associative biclustering methods that identify a variety of types of coherent patterns in combined functional genomics

data. The method searches for associations using statistical criteria functions and estimates confidence across multiple data types. We have shown that on a model synthetic gene expression dataset our method outperforms other methods designed to identify transcription modules in gene expression data alone. The method is placed in a computational framework, which allows rapid customization and deployment for new data sets and data types. For example, results of searches incorporating data on transcription, such as gene expression microarrays, provide direct information on putative regulons. We are currently analyzing our results of associative data patterns from a large yeast data compendium. We have also developed a graphical viewer to allow interaction with the results of the associative biclustering searches and the associated data types, and we are working on applying this to a massive collection of metabolite mass spectra measurements.

**TIER 3: SYSTEMS LEVEL ANALYSIS OF METABOLISM AND REGULATION.** Gene regulatory networks (GRNs) spatiotemporally regulate cellular physiology to optimize resource utilization, maintain integrity of genetic information, and contribute towards competitive fitness of the organism under changing environmental conditions. These networks are dynamically modulated with changing environments, and the underlying mechanisms behind these changes may be learned by integrating a wide variety of experimental and genomic data. We have successfully inferred causal and predictive models for these networks in *Halobacterium salinarum* NRC-1, and are currently applying these methods to other DOE-relevant organisms.

# 237

## An ENIGMA Analysis Platform for Metabolomics: Towards Identification of Metabolites from Untargeted Experiments

**Gary Siuzdak*** (siuzdak@scripps.edu)

[4]The Scripps Research Institute

**Project Goals: Develop metabolomics based microbial tools such as XCMS metaXCMS, chemical approaches, and to facilitate identification we have developed a novel database, METLIN. The METLIN Metabolite Database is a repository of metabolite information as well as tandem mass spectrometry data. METLIN is a metabolite database for metabolomics containing over 42,000 structures, it also represents a data management system designed to assist in a broad array of metabolite research and metabolite identification by providing public access to its repository of current and comprehensive MS/MS metabolite data.**

Mass spectrometry-based untargeted metabolomics requires data preprocessing approaches to correlate specific metabolites to their biological origin. XCMS is an LC/MS-based data analysis approach incorporating novel nonlinear retention time alignment, feature detection, and feature matching. Without using internal standards, the method dynamically identifies hundreds of endogenous metabolites for use as standards, calculating a nonlinear retention time correction profile for each sample. Following retention time correction, the relative metabolite ion intensities are directly compared to identify changes in specific endogenous metabolites.

XCMS, however, often results in the observation of hundreds to thousands of features that are differentially regulated between sample classes. A major challenge in interpreting the data is distinguishing metabolites that are causally associated with the phenotype of interest from those that are unrelated but altered in downstream pathways as an effect. To facilitate this distinction, here we describe new software called metaXCMS for performing second-order ("meta") analysis of untargeted metabolomics data from multiple sample groups representing different models of the same pheno- type. While the current version of XCMS was designed for the direct comparison of two sample groups, metaXCMS enables meta-analysis of an unlimited number of sample classes to facilitate prioritization of the data and increase the probability of identifying metabolites causally related to the phenotype of interest. metaXCMS is used to import XCMS results that are subsequently filtered, realigned, and ultimately compared to identify shared metabolites that are up- or down-regulated across all sample groups. We demonstrate the software's utility with halobacterium mutants. metaXCMS is freely available at http://metlin.scripps.edu/metaxcms/.

To further facilitate the assignment of unknown mass spectral features, we have also demonstrated that profiling can be performed on cultures uniformly labeled with stable isotopes of nitrogen ($^{15}N$) or carbon ($^{13}C$). This makes it possible to accurately count the number of carbon and nitrogen atoms in each molecule, providing a robust means for reducing the degeneracy of chemical space and thus obtaining unique chemical formulae for features measured in untargeted metabolomics having a mass greater than 500 Da, with relative errors in measured isotopic peak intensity greater than 10%, and without the use of a chemical formula generator dependent on heuristic filtering. These chemical formula can serve as indicators for the presence of particular metabolic pathways.

In conjunction with XCMS and metaXCMS, and these chemical approaches, to facilitate identification we have developed a novel database, METLIN. The METLIN Metabolite Database is a repository of metabolite information as well as tandem mass spectrometry data. METLIN is a metabolite database for metabolomics containing over 42,000 structures, it also represents a data management system designed to assist in a broad array of metabolite research and metabolite identification by providing public access to its repository of current and comprehensive MS/MS metabolite data. An annotated list of known metabolites and their mass, chemical formula, and structure are available on the METLIN website. Each metabolite is linked to outside resources such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) for further reference and inquiry. MS/MS data is also available on many of the metabolites. The

list is expanding continuously as more metabolite information is being deposited and discovered.

# 238

## ENIGMA Knowledgebase: MicrobesOnline, Gaggle and RegTransBase

Paramvir S. Dehal[1]* (psdehal@lbl.gov), J. Chris Bare,[2] Pavel S. Novichkov,[1] John-Marc Chandonia,[1] David Reiss,[2] Morgan N. Price,[1] Keith Keller,[1] Jason Baumohl,[1] Marcin P. Joachimiak,[1] Inna L. Dubchak,[1] Adam P. Arkin,[1,3] and Nitin S. Baliga[2]

[1]Lawrence Berkeley National Laboratory; [2]Institute of Systems Biology; and [3]University of California, Berkeley

**Project Goals: ENIGMA scientists seek to understand in situ microbial activity and community dynamics through detailed assessment of molecular function from proteins to populations. By studying communities with activities of interest to DOE mission we hope to reveal the mechanistic basis for those activities and their support in a changeable and uncertain environment. ENIGMA has 4 main aims**

- **Measurement and analysis of environmental activity, composition, structure, and strategies of microbial communities in situ**
- **Use controlled laboratory consortia to identify essential microbial contributions to environmental activities, identify specific and selected interactions, and isolate keystone organisms/processes**
- **Efficiently advance these environmental microorganisms to model organism status and map their molecular functions to community phenotypes and environmental activities.**
- **Development of the LBNL Systems Environmental Microbiology Workbench and Knowledge Framework**

The ENIGMA Knowledgebase integrates the diverse data sets generated by the project with the central goal of enabling development of computational algorithms to predict gene, microbe and microbial community function. To this end, we maintain systems for low level per experiment data capture, cross experiment data integration and data analysis. Because of the importance of relating experimental data and meta-data, we have created LIMS and relational databases for data repositories for ENGIMA experiments. This underlying data, together with relevant publicly available data sets, is then integrated into MicrobesOnline, Gaggle and RegTransBase to enable the computation of predictive models of metabolism, gene regulation and cell response to environmental stimuli.

**MicrobesOnline:** The MicrobesOnline database (http://www.microbesonline.org) currently holds over 3000 microbial genomes and is updated semi-annually, providing an important comparative and functional genomics resource to the community. MicrobesOnline continues to provide an interface for genome annotation, which like all the tools reported here, is freely available to the scientific community. MicrobesOnline allows the user to quickly access functional genomics data in comparative and evolutionary framework by providing gene homology, domains, phylogenetic trees, operon and regulons predictions together with functional data such as protein-protein interaction, microarray expression data and phenotype/genotype associations. We have developed methods, FastBLAST, FastHMM and FastTree, to enable us to deal with the many millions of gene sequences generated from metagenomics. These tools allow MicrobesOnline to provide the only comparative metagenomic data browser which features a tree browser for every gene family.

**RegTransBase:** We have built tools and resources for studying regulation in bacteria and archaea using comparative genomics approach. In addition to working on a high quality semi-manual regulon inference in a wide range of species we are building several on-line resources covering different aspects of regulation. RegTransBase, a database of regulatory interactions from literature collected by a group of experts, currently includes 5,100 annotated articles describing twelve thousand experiments. RegPrecise describes manually curated computational predictions of regulons in bacterial genomes done by comparative genomics. RegPredict is a set of highly integrated web tools for fast and accurate inference of regulons. All regulation-related resources are based on the MicrobesOnline data.

**Gaggle:** The ability to seamlessly interoperate across analysis tools, data sets and data types developed by different scientists across the world has long been a limitation for biologists. Gaggle is a framework for interoperability between bioinformatics software tools which solves this problem. Gaggle enables message passing between data sources, analysis software and visualization tools, including Cytoscape, MultiExperiment Viewer and R, in addition to web resources such as KEGG and STRING. Gaggle and the Firefox web browser extension Firegoose, allows the user to treat independently developed programs as a larger, coupled suite of tools for exploratory data analysis. In addition to visualization in externally developed tools such as MeV and Cytoscape, the Gaggle Genome Browser was developed to visualize any experimental data along the genome coordinates. Visualization of data such as microarray, RNA-seq, protein mass spectrometry and ChIP-chip/seq in the context of the genome can reveal the molecular mechanisms that regulate transcription.

# 239

## Microbial Deconstruction Proteomics: Mapping Protein Subcellular Localization in the Extremophiles *Pyrococcus furisiosus*, *Halobacterium salinarum*, and *Sulfolobus solfataricus*

Robert Rambo,[1] Adam Barnebey,[1] Michael W.W. Adams,[3] Christopher Bare,[5] Nitin S. Baliga,[5] Trent Northen,[1] Ben Bowen,[1] Sunia Trauger,[4] Gary Siuzdak,[4] John A. Tainer,[1,6] and **Steven M. Yannone**[1]

[1]Dept. of Molecular Biology, Lawrence Berkeley National Lab; [2]Center for Life in Extreme Environments, Portland State University; [3]Department of Biochemistry and Molecular Biology, University of Georgia; [4]Center for Mass Spectrometry, The Scripps Research Institute; [5]Institute for Systems Biology; and [6]Department of Biochemistry and Molecular Biology, The Scripps Research Institute

**Project Goals: This project started as part of a larger foundational science program (MAGGIE) to develop tools and technologies needed to manipulate non-model organisms and microbial communities to address DOE mission goals. Given that a large proportion of genes and proteins from organisms of interest are poorly characterized, and that all novel enzymes are uncharacterized by definition, we set out to develop universally applicable and practical technologies for mapping protein localization and protein-assemblies within any given microbe. Our goals include; 1) developing universal microbial deconstruction and fractionation processes that allow proteome-wide analysis retaining abundance, assembly, and localization information for each protein, 2) to develop a user-friendly interface for these complex and large data sets that are useful to both informaticists and bench scientists, 3) develop molecular biology for our model system (*Sulfolobus solfataricus*) to validate novel findings with our methods.**

The speed and efficiency of microbial genome sequencing has far outpaced our ability to assign functions to novel genes. To fully exploit the diversity of chemistries evolved within microbial life, and to understand interactions within communities, new types of informative datasets are needed to annotate the growing number of genes with unknown function. Subcellular localization and assembly into larger complexes are informative factors in predicting or determining protein functionality. Extremophilic enzymes and protein complexes are exceptionally stable and arguably the most tractable model system for proteome-wide isolation of macromolecular assemblies.

Here we have chosen the extremophilic microbes *Sulfolobus, Pyrococcus* and *Halobacterium* as model systems to develop a universally applicable and practical technology for mapping protein localization and protein-assemblies within any given microbe. We have applied robust biochemical fractionation techniques coupled to HTP MS/MS proteomics to assign cellular locality and physical characteristics to all proteomic identifications. Density, mass, and cellular locality were exploited to fractionate microbial proteomes for each of these highly divergent extremophiles. Constant buffer conditions matched to the widely varied intracellular condition of each organism were used to both stabilize assemblies, and establish the universal applicability of our approaches. Microbial biomass was partitioned into four primary fractions, 1) extracellular, 2) membrane, and two intracellular fractions 3), high-mass particles, and 4) low-mass particles. With this approach we observe 60% of the predicted *Sulfolobus* proteome (1783 proteins) and 305 proteins partition with >95% confidence of being exclusively in one cellular partition. Not surprisingly, the majority (184) of these reside exclusively in the membrane fraction with 30% of these being hypothetical proteins. The intracellular high-mass partition contained intact thermosome and ribosome that were characterized by small-angle x-ray scattering and EM. The small-mass partition was by far the most complex and degenerate, and contained only 22 proteins that were not observed elsewhere in the cell. These highly complex proteomic data sets are presented for simple and intuitive visual inspection using a genome browser developed within ENIGMA. The methods and technologies developed here are applicable to any organism or community of interest to DOE and provide novel proteome-wide datasets for assigning protein locality, function, and assembly states within microbes.

Because our localization proteomic data sets are novel, they require a means for validation. To address this, we have developed a new high-throughput protein expression system for *Sulfolobus*. We have built on viral based vectors to develop a PCR-based gateway-cloning vector. We are implementing our new molecular biology to validate assemblies inferred from the cellular deconstruction analyses and to validate localization and assembly states of hypothetical and annotated proteins.

*Sulfolobus solfataricus* is a single cell organism that thrives at 80°C in highly acidic volcanic springs (pH=2-3). There are very few life forms able to compete in this extreme environment, which leads to an exceptionally simple microbial community, including only viral pathogens and fewer than twenty putatively identified organisms. Such a simple community can provide an excellent platform to test hypotheses about co-evolutionary adaptations and community interactions from more complex and less malleable communities. Together, the simple nature of solfataric spring communities, our novel molecular biology, and our deconstruction data sets make *Sulfolobus* a particularly useful model system for testing co-evolution and community interaction hypotheses from more complex systems.

# 239A‡

submitted post-press

## Genome-Scale Phylogenetic Function Annotation of Large and Diverse Protein Families

Barbara E. Engelhardt,[1,6] Michael I. Jordan,[1,2] Susanna Repo,[3] Gaurav Pandey,[3] Ameet Talwalkar,[1] Jeffrey Yunes,[4] **John-Marc Chandonia**,[3,5*] and Steven E. Brenner[3,5] (brenner@compbio.berkeley.edu)

[1]EECS Department, University of California, Berkeley; [2]Department of Statistics, University of California, Berkeley; [3]Plant and Microbial Biology Department, University of California, Berkeley; [4]Department of Bioengineering, University of California, Berkeley; [5]Lawrence Berkeley National Laboratory; and [6]Computer Science Department, University of Chicago, Chicago, Ill.

**Project Goals: We are awash in proteins discovered through high-throughput sequencing projects. As only a minuscule fraction of these have been experimentally characterized, computational methods are widely used for automated annotation. Unfortunately, these predictions have littered the databases with erroneous information, for a variety of reasons including the propagation of errors and the systematic flaws in BLAST and related methods. In collaboration with Michael Jordan's group, we have developed a statistical approach to predicting protein function that uses a protein family's phylogenetic tree, as the natural structure for representing protein relationships. We overlay on this all known protein functions in the family. We use a model of function evolution to then infer the functions of all other protein functions. Even our initial implementations of this method, called SIFTER (statistical inference of function through evolutionary relationships) have performed better than other methods in widespread use. We are presently making numerous improvements to the underlying SIFTER algorithm and enhancing its ability to work on a wide range of data.**

It is now easier to discover thousands of protein sequences in a new microbial genome than it is to biochemically characterize the specific activity of a single protein of unknown function. Through metagenomic analysis, next-generation sequencing heralds unprecedented opportunities for understanding the environmental microbiota. A single experiment alone, the Global Ocean Sampling study, more than doubled the number of known protein sequence entries. However, despite this large body of new sequence information, functional annotation remains a major challenge. Molecular functions of proteins in the novel genomes continue to be discovered, in large part by homology to those experimentally characterized in model organisms.

Typically, protein function annotation involves finding homologs of a protein sequence, followed by database queries and computational techniques to predict function from the annotated homologs. These methods rely on the principle that proteins from a common ancestor may share a similar function. However, most protein families have sets of proteins with different functions and therefore traditional bioinformatics approaches are unable to reliably assign the appropriate function to unannotated proteins. Currently, protein function databases have a large proportion of erroneously annotated proteins, where the incorrect annotations were either derived using an imprecise computational technique or inferred using another incorrect annotation.[1-4]

We have proposed integrating available functional data using the evolutionary relationships of a protein family, and we implemented this method in the program SIFTER (Statistical Inference of Function Through Evolutionary Relationships). The SIFTER methodology uses a statistical graphical model to compute the probabilities of molecular functions for unannotated proteins. Currently, SIFTER takes as input a reconciled phylogeny and a set of annotations for some of the proteins in the protein family. We incorporate known information about function by computing the probability of each of the candidate functions for the proteins in the tree with available functional evidence from the GOA database. The candidate molecular functions are represented as a boolean vector, where initially the probability associated with each candidate function is a function of the set of annotations for that protein and their corresponding evidence types (e.g., experimental, electronic). From this reconciled phylogeny with sparse observations, SIFTER computes the posterior probability of each molecular function for all proteins in the family using a simple statistical model of protein function evolution.

We tested the performance of SIFTER on three different protein families: AMP/adenosine deaminases, sulfotransferases and Nudix hydrolases with cross-validation experiments. SIFTER's performance was compared with three other function prediction algorithms: BLAST, GOtcha and Orthostrapper, and SIFTER was shown to outperform the other methods. In addition, on a genome-wide scale we used SIFTER to annotate the experimentally characterized proteins from *Schizosaccharomyces pombe*, based on the annotations from 26 other fungal genomes. The newest version of SIFTER implements a faster method for calculating the posterior probabilities, and this improvement, together with a more general evolutionary model make SIFTER applicable on large and functionally diverse protein families and on genome-scale function annotation.

The development of SIFTER is an ongoing project and a new version of the program is now available (manuscript under review). We are currently testing SIFTER for metagenomic sequences with the acid mine drainage datasets from Jill Banfield. In the near future, we are planning to expand our analysis to other metagenomic datasets, such as the termite gut datasets from the JGI. We also use SIFTER to annotate enzymes from chlorite dismutase and perchlorate reductase families, in order to identify species that are capable of perchlorate reduction. Furthermore, we are validating SIFTER predictions experimentally using the large and extremely diverse Nudix family of hydrolases as a test bed.

**References**

1. Brenner SE 1999 *Trends Genet*. 15 132-3
2. Galperin MY and Koonin EV 1998 *In Silico Biol*. 1 55-67
3. Jones CE, Brown AL and Baumann U 2007 *BMC Bioinformatics*. 8 170
4. Schnoes AM, Brown SD, Dodevski I and Babbit PC 2009 *PLoS Comput Biol.* 5 e1000605

This project has been funded with DOE grant BER KP 110201.

# Structural Biology, Molecular Interactions, and Protein Complexes

# 240

## Neutron Protein Crystallography Station User Facility

S. Zoe Fisher, Andrey Kovalevsky, Marat Mustyakimov, Marc-Michael Blum, Benno P. Schoenborn, Mary Jo Waltman, and **Paul Langan*** (Langan_paul@lanl.gov)

Bioscience Division, Los Alamos National Laboratory, Los Alamos, N.M.

**Project Goals: PCS is a high-performance neutron beamline that forms the core of a BER-funded experimental User capability at Los Alamos Neutron Science Center (LANSCE) for investigating the structure and dynamics of proteins, biological polymers, and membranes.**

Neutron diffraction is a powerful technique for locating hydrogen atoms, which can be hard to detect using X rays, and therefore can provide unique information about how biological macromolecules function and interact with each other and smaller molecules. This unique User capability is being used to investigate several enzymes that are important to USDA and DOE Genome Science program missions in renewable energy and the environment, with a view to understanding their detailed catalytic mechanisms. This new information is then being exploited to manipulate their performance and use. Neutron diffraction has also been crucial in revealing the structures and hydrogen bond arrangements in naturally occurring cellulose in lignocellulosic biomass and how they are rearranged by pretreatments to enhance conversion to biofuels. This information has led to the optimization of pretreatments to improve their cost-efficiency.

PCS Users have access to neutron beam time, deuteration facilities, protein expression and substrate synthesis with stable isotopes, a purification and crystallization laboratory, and software and support for data reduction and structure analysis. A HomeFlux X-ray system has been recently purchased that will allow users to collect X-ray data from the same samples used for neutron diffraction. The PCS beamline exploits the pulsed nature of spallation neutrons and a large electronic detector to efficiently collect wavelength-resolved Laue patterns using time-of-flight techniques. We encourage potential users to communicate with us before applying for beam time for technical guidance and help with proposal preparation.

For technical information about the PCS and experimental requirements contact Zoe Fisher (505) 665-4105 zfisher@lanl.gov or Paul Langan (505) 665 8125 langan_paul@lanl.gov

**Proposal Submission:** Proposals must be submitted using the process on the LANSCE website. To access the proposal submission site, go to the LANSCE home page, http://lansce.lanl.gov/. On this page click the tab "Lujan Center' and then the link 'Submit a Proposal.This will take you to the on-line submission system. Detailed instructions for preparing the proposals can be found on the proposal submission sites under "Step-by-Step Guide to Submitting an Online Proposal."

# 241

## The Berkeley Synchrotron Infrared Structural Biology (BSISB) Program Overview

**Hoi-Ying N. Holman**[1,2]* (hyholman@lbl.gov), Hans A. Bechtel,[1,3] Rafael Gomez-Sjoberg,[1,4] Zhao Hao,[1,2] Ping Hu,[1,2] Michael C. Martin,[1,3] and Peter Nico[1,2]

[1]Berkeley Synchrotron Infrared Structural Biology Program, [2]Earth Sciences Division, [3]Advanced Light Sources, and [4]Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.

**Project Goals: The Berkeley Synchrotron Infrared Structural Biology (BSISB) program is a national user facility for infrared spectromicroscopy and chemical microcharacterization of living cells.**

The Berkeley Synchrotron Infrared Structural Biology (BSISB) program is a national user facility for infrared spectromicroscopy and chemical microcharacterization of biological systems. BSISB was initiated in 2010 to maintain a forefront research facility for infrared and optical characterization of chemistry in living cells with state-of-the-art instrumentation and expertise. The BSISB program has developed an integrated microfluidic synchrotron infrared (SIR) spectromicroscopy platform, which is a technique that is ideal for tracking the chemical composition and reactions in living cells during their adaptive responses to internal or external stimuli and perturbations. The BSISB program is also developing visible (VIS) hyperspectral/fluorescence microscopy approaches for simultaneously tracking changes in cellular morphology, structure, and other biological processes such as gene expression and signaling during SIR experiments. This new BSISB development of live-cell chemical biological imaging technologies will also be aided by a new generation of microfluidics platform. Our technological research and development effort will be accelerated by BSISB participating scientists with wide ranging research projects of bioenergy, medical, and environmental studies

that are important to DOE missions. Such synergistic interactions will provide interdisciplinary expertise and scientific critical mass to meet the emerging BSISB experimental challenges.

# 242

## The National Center for X-Ray Tomography, a DOE Structural Biology Facility

Carolyn A. Larabell[1,2,3]* (CarolynLarabell@ucsf.edu) and Mark A. Le Gros[2,3]

[1]Department of Anatomy, University of California, San Francisco; and [2]Physical Biosciences Division, Lawrence Berkeley National Laboratory

http://ncxt.lbl.gov

**Project Goals: DOE Structural Biology User Facility**

The National Center for X-ray Tomography (NCXT) is a unique structural biology facility for high-resolution imaging of cells, biofilms and hydrated organic materials in general. The Center has developed soft X-ray tomography (SXT), which is a technique ideally suited to imaging subcellular architecture and organization in the native state. SXT is similar in concept to the well-established medical diagnostic technique computed axial tomography (CAT), except SXT is capable of imaging with a spatial resolution of 50 nm, or better. In SXT, cells are imaged using photons from a region of the spectrum known as the 'water window,' between the K shell absorption edges of carbon (284 eV, λ=4.4 nm) and oxygen (543eV, λ=2.3 nm). This results in quantitative, high-contrast images of intact, fully hydrated cells without the need to use contrast-enhancing agents. The method is particularly sensitive to the spatial distribution of carbon throughout microorganisms and provides information of great relevance to basic biological systems science, the production of biofuels and carbon sequestration. To image specific molecules with respect to cell structures, we have developed correlated imaging methods such as high numerical aperture cryogenic light microscopy. This multimodal approach allows fluorescently labeled molecules to be localized in the context of a high-resolution 3-D tomographic reconstruction of the cell. We will provide examples of data collected using SXT, which demonstrate that SXT is very well suited to DOE mission relevant research. The NCXT, as a DOE Structural Biology Facility, makes SXT available to national and international investigators.

# 243

## Proteomic Scale Solution X-Ray Scattering and its Implications for Structural Biology

**John Tainer** and Greg Hura* (glhura@gmail.com)

Lawrence Berkeley National Laboratory

**Project Goals: Ecosystems and Networks Integrated with Genes and Molecular Assemblies A multiscale systems approach to microbial bioremediation, carbon sequestration and energy production; from molecules to cells to communities.**

High throughput solution structural analyses by small angle X-ray scattering efficiently enables the characterization of shape and assembly for nearly any purified protein. Crystallography has provided a deep and broad survey of macromolecular structure. Shape and assembly from SAXS in combination with available structures is often enough to answer critical mechanistic questions both enhancing the value of a structure and obviating larger crystallographic projects. Moreover, SAXS is a solution based technique, sample requirement are modest and compatible with many other biophysical methods. Here we present our high throughput SAXS data collection and analysis pipeline as applied to structural genomics targets, and metabolic pathways. Our goals of metabolic engineering and understanding protein mediated reactions rely on knowing the shape and assembly state of reactive complexes under an array of conditions. Given the number of gene products involved in metabolic networks, SAXS will play an important role in characterizing the structure of each individually, in complex with partners, and in various contexts. SAXS is well positioned to bridge the rapid output of bioinformatics and the relatively slow output of high resolution structural techniques.

submitted post-press

# 243A‡

## $CO_2$ Capture by Amyloid Fibers

Dan Li*[1] (danli@mbi.ucla.edu), Hiroyasu Furukawa,[2] Omar M. Yaghi,[2] and **David Eisenberg**[1]

[1]UCLA-DOE, Institute for Genomics and Proteomics; [2]Department of Chemistry and Biochemistry, University of California-Los Angeles

**Project Goals: To sequestrate $CO_2$ from flue gases by functionalized amyloid fibers.**

Global warming accompanied with catastrophic consequences including food and water shortages, ecosystem irreversible change, and extreme weather events has generated great worries. A major cause of global warming is the increase of atmospheric greenhouse gas levels. 72% of the emitted greenhouse gases consist of carbon dioxide. Over the past 60 years, the global carbon emissions dramatically increased mainly from burning of carbon-based fossil fuels.

Capture of the $CO_2$ emitted from fossil fuel burning is of great importance for environmental protection.

Currently, aqueous monoethanolamine (MEA) is extensively applied to sequestrate $CO_2$ from flue gases. The amino group of MEA reacts with $CO_2$ forming carbamate. The reaction is rapid and requires no catalyst. However, MEA has many drawbacks. It is toxic, flammable and corrosive. MEA degrades in the presence of $O_2$ and $CO_2$ resulting in extensive amine loss and equipment corrosion as well as generating amine waste pollution. The regeneration of MEA demands high energy.

Here, we present a potential amine substitute—Lys-containing amyloid fibers. Amyloid fibers are protein aggregates with well-defined cross-β structures. As biological macromolecules, protein materials are biodegradable and environmental friendly. Protein amyloid fibers are repetitive structures. The hydrophobic core confers compactness and stability. Amyloid fibers cause a wide range of human pathologies, including Alzheimer's disease, dialysis-related amyloidosis and Parkinson disease. Attractively, the intrinsic structural properties of amyloid fibers make it possible to be functionalized as applicable materials. We have studied the $CO_2$ absorption capacity of amyloid fibers formed by Ac-VQIVYK-$NH_2$, an amyloidogenic sequence from the Alzheimer's tau protein. The dry fibers showed chemisorption of $CO_2$ by an isotherm experiment. In the VQIVYK fiber, the distance between lysines is ~4.8 Å, which may allow cooperativity of neighboring lysines. Therefore, compared to the monomeric peptide, lysine in fibers has a lower side chain pKa and is more favorable as the primary amine for carbamate formation.

# Validation of Genome Sequence Annotation

## 244

### NBC: The Naïve Bayes Classification Tool for Shotgun Metagenomics from Bacteria, Fungi, and Viruses

Gail Rosen[1]* (gailr@ece.drexel.edu), Aaron Rosenfeld,[1] Tze Yee Lim,[1] Yemin Lan,[1] and Christopher Blackwood[2]

[1]Drexel University (Electrical and Computer Engineering, Computer Science, Physics, and Biomedical Engineering Departments), Philadelphia, Penn.; and [2]Kent State University (Department of Biological Sciences), Kent, Ohio

**Project Goals: We aim to produce a taxonomic classifier that runs quickly and can classify genomes (prokaryotic, eukaryotic, and viral). The naïve Bayes classification tool is implemented on a web site for public use, http://nbc.ece.drexel.edu. The database is soon to be expanded to include fungi and viruses, in order to expand its capabilities to soil and marine studies.**

Datasets from high-throughput sequencing technologies have yielded a vast amount of data about organisms in environmental samples. Yet, it is still a challenge to assess the exact organism content in these samples because the task of taxonomic classification is too computationally complex to annotate all reads in a whole-genome shotgun dataset. An easy-to-use webserver is needed to process these reads. While many methods exist, only a few are publicly available on webservers, and out of those, most do not annotate all reads.

We introduce a webserver that implements the naïve Bayes classifier (NBC) to classify all metagenomic reads to their best taxonomic match. Results indicate that NBC can assign next-generation sequencing reads to their taxonomic classification and can find significant populations of genera that other classifiers may miss. We demonstrate that the tool can handle a complete pyrosequencing dataset in one day on a four-core server, and it gives the full lineage for each read, so that users can easily analyze the taxonomic composition of their datasets.

In addition to classification, we are working on developing a way to assess the novelty of sequences using detection theory. Using 5-fold cross-fold validation, we are able to achieve approximately 89% sensitivity and 95% specificity for species-level known/novel determination using the website top-hit likelihood scores. We are also working in conjunction with RDP to apply this technique to 16S rRNA sequences.

## 245

### Assessing an Integrated Approach to Accurate Functional Annotation of Putative Enzymes in the Methanogen *Methanosarcina acetivorans*

Lucas Showman[1]* (lshowman@iastate.edu), Ethel Apolinario,[2] Libuse Brachova,[1] Yihong Chen,[3] Zvi Kelman,[3] Zhuo Li,[3] Kevin Sowers,[2] John Orban,[3] and **Basil J. Nikolau**[1]

[1]W. M. Keck Metabolomics Research Laboratory, Iowa State University, Ames; [2]Department of Marine Biotechnology, University of Maryland, Baltimore County; and [3]Institute for Bioscience and Biotechnology Research, University of Maryland College Park, Rockville

http://carb.umbi.umd.edu/g2f

**Project Goals: We are using the methanogenic archaeon *Methanosarcina acetivorans* (MA) as a model organism to develop tools for rapid and reliable annotation and validation of protein function. The target genes are putative enzymes with detectable in vivo expression that have been selected after genome re-annotation.**

The experimental approach begins with heterologous *E. coli* expression and purification of individual MA gene products. An initial ligand-binding screen of the purified protein using NMR spectroscopy determines whether candidate compounds can physically interact with the protein and act as putative substrates or products. Where possible, this is followed up with an experiment to see if the MA gene can complement an *E. coli* strain carrying a knock-out allele at the closest homolog.

In most if not all of our case studies, we find that the initial NMR screen is indicative of whether the function assignment is correct. Therefore this represents a method that may be suitable for accurate and efficient functional annotation of partially characterized enzymes without the need for developing protein-specific assays.

# 246

## Functional Annotation of Putative Enzymes in *Methanosarcina acetivorans*

Ethel Apolinario,[1] Libuse Brachova,[3] Yihong Chen,[2] Zvi Kelman,[2] Zhuo Li,[2] Basil J. Nikolau,[3] Lucas Showman,[3] Kevin Sowers,[1] and **John Orban**[2,*] (jorban@umd.edu)

[1]Department of Marine Biotechnology, University of Maryland, Baltimore; [2]Institute for Bioscience and Biotechnology Research, University of Maryland College Park, Rockville; [3]W.M. Keck Metabolomics Research Laboratory, Iowa State University, Ames

**Project Goals: The goal of the project is to develop rapid experimental approaches for accurate annotation of putative enzymatic functions. Targets of interest range from those with tentatively assigned function to hypotheticals.**

Methane-producing organisms provide an efficient and cost-effective biofuel which is self-harvesting and can be distributed readily using infrastructure that is already in place. As with other genomes, however, accurate functional annotation in methanogens lags significantly behind the large body of sequence data, representing a sizable gap in our understanding of biology in these organisms. We are using the methanogenic archaeon, *Methanosarcina acetivorans* (MA), as a model system for developing experimental tools for rapid and reliable annotation and validation of function. The target genes are putative enzymes in MA with detectable *in vivo* expression.

Our experimental approach utilizes a combination of methods for rapid function assignment. NMR spectroscopy is used to screen for putative substrates, products, or their structural analogs. Where possible, we have followed up on function assignments by checking to see if the MA gene can complement the corresponding *E. coli* knockout. We have used this approach to both validate and correct functional assignments in MA target genes, as will be illustrated with examples. Further, insights into the functional annotation of "hypotheticals" are being obtained by integrating mass spectrometry based metabolite profiles of gene knockouts with NMR-based approaches and these will also be discussed.

# 247

## The Ribosomal Database Project: Tools and Sequences for rRNA Analysis

J.R. Cole* (colej@msu.edu), Q. Wang, B. Chai, J.A. Fish, D.M. McGarrell, and **J.M. Tiedje**

Center for Microbial Ecology, Michigan State University, East Lansing

**Project Goals: The Ribosomal Database Project (RDP; http://rdp.cme.msu.edu) offers aligned and annotated rRNA sequence data and 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 bioremediation.**

Our view of the evolutionary relationships among life forms on Earth has been revolutionized by the comparative analysis of ribosomal RNA sequences. Life is now viewed as belonging to one of three primary lines of evolutionary descent: Archaea, Bacteria and Eucarya. This shift in paradigm has not only challenged our understanding of life's origin, but also provided an intellectual framework for studying extant life—particularly the vast diversity of microorganisms. Ribosomal RNA diversity analysis using genes amplified directly from mixed DNA extracted from environments has demonstrated that the well-studied microbes described by classical microbial systematics represent only a small percentage of diversity. The use of rRNA to explore uncharacterized diversity had become such a relied-upon methodology that by 2008, 77% of all INSDC bacterial DNA sequence submissions described an rRNA sequence, and only 2% of these entries had a Latin name attached (valid or otherwise; Christen, 2008)! Examining the RDP's collection of quality rRNA sequences demonstrates that cultivated organisms represent only a fraction of observed rRNA diversity, and currently available genome sequences cover an even smaller slice of this cultivated fraction (Fig. 1). Phylogenetically informed selection of sequencing candidates, as done in the GEBA Project, can help improve genome coverage of diversity represented by cultivated organisms (Wu et al., 2009), and single cell sequencing can provide partial genome data for uncultivated organisms; but it will be years before these techniques are able to make practical progress towards tackling the immense diversity represented by the collection of rRNA sequences. In fact, it is our knowledge of rRNA diversity that is guiding these efforts.

In the current release (January 2011), RDP offers 1,498,677 aligned and annotated quality-controlled public bacterial and archaeal rRNA sequences along with tools that allow researchers to examine their own sequences in the context of the public sequences (Cole et al., 2009). In addition, 8,770 researchers have over 5.9 million private pre-publication

sequences in the *my*RDP account system. On average, the RDP website is visited by over 8,400 researchers (unique ip addresses) from more than 150 countries in over 21,000 analysis sessions each month, while Web Services (SOAP) interfaces to these RDP analysis functions process an additional 8.6 billion bases of sequence per month for high-volume users running their own analysis queues. In addition, since its release in May 2008, the RDP Pyro Pipeline has been used by over 1,800 researchers (unique e-mail addresses) to process their own high-throughput current-generation rRNA sequence data. In addition, the command-line version of the RDP Classifier, designed for users needing local high-throughput analysis, has been downloaded over 3,400 times since its release on SourceForge (Wang et al., 2007). The RDP has also been working with standards bodies, such as the Genomic Standards Consortium (GSC; http://gensc.org/gsc/) and the Terragenome Consortium (http://www.terragenome.org) to help define environmental annotation standards for rRNA and other environmental marker gene libraries, and to assure that RDP is ready for the new standards. These results have been incorporated into the new MIMARKS (Minimal Information about a MARKer gene Sequence; Nature Biotechnology, in press) covering rRNA and environmental gene sequences.

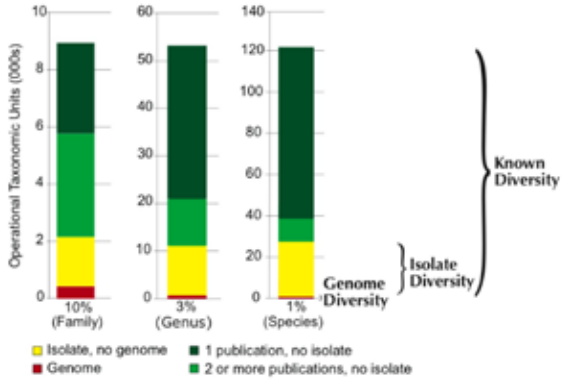



Figure 1. Clusters from a total of 668,513 high-quality near-full-length bacterial rRNA sequences, including sequences from all 2,179 bacterial genomic sequences available from GenBank RefSeq on 23 September 2010 were clustered using the RDP mcClust tool implementing a memory constrained complete-linkage algorithm. Cluster distances approximate the given taxonomic rank.

The RDP is supported by the Office of Science (BER), U.S. Department of Energy, Grant No. DE-FG02-99ER62848

## References

1. Christen, R. 2008. Global sequencing: a review of current molecular data and new methods available to assess microbial diversity. *Microbes and Environments* 23:253-268.
2. Cole, J. R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, T. Marsh, G. M. Garrity, and J. M. Tiedje. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37:D141-D145
3. Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole. 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol.* 73:5261-5267.
4. Wu, D., P. Hugenholtz, K. Mavromatis, R. Pukall, E. Dalin, et al. 2009. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 462:1056-1060.

# Knowledgebase and Computing for Systems Biology

## 248

### Web API for Annotation, Access, and Analysis of Genomic Data in the SEED Framework

**Ross Overbeek**,[1]* Robert Olson,[2,3] Terrence Disz,[2,3] Bruce Parello,[1] Rob Edwards,[4] **Christopher Henry**,[2,3] Scott Devoid,[3] and **Rick Stevens**[2,3]

[1]Fellowship for Interpretation of Genomes, Ill.; [2]University of Chicago, Ill.; [3]Argonne National Laboratory, Ill.; and [4]San Diego State University, Calif.

**Project Goals: This project aims to develop software and data infrastructure to support programmatic access to the SEED and Model SEED resources for the annotation of genomes, integration of omics data, and reconstruction of genome-scale metabolic models. Within this overarching objective, are four specific aims: (i) enhancing the computational infrastructure behind the SEED framework to improve extensibility, accessibility, and scalability; (ii) integrate extensions into the SEED data-model and software to accommodate new data types including regulatory networks, genome-scale metabolic models, structured assertions, eukaryotic genome data, and growth phenotype data; (iii) develop an application programming interface (API) to provide remote access to the SEED database and tools; (iv) apply SEED infrastructure, data, and APIs to annotate genomes and construct genome-scale metabolic models for organisms with applications to bioenergy, carbon cycle, and bioremediation.**

The SEED environment for the integration, annotation, and comparison of genomic data now includes thousands of microbial genomes and many eukaryotic genomes, all linked with a constantly updated set of curated annotations embodied in a large and growing collection of encoded subsystems and derived protein families. Additionally, the Model SEED resource has been developed to translate SEED annotations into functioning genome-scale metabolic models. Recently we have developed a set of web-services for SEED that offer programmatic access to data and tools included within the SEED environment (find documentation at http://blog.theseed.org/servers/). The services include the ability to remotely submit genomes to RAST and Model SEED for annotation and modeling; they enable users to query the SEED database for genome features, functional annotations, gene orthologs, and sequence similarities; and they enable users to apply flux balance analysis with genome-scale metabolic models to simulate cell growth in a variety of media conditions and with a variety of mutations. We highlight the powerful features of the SEED web services by demonstrating how multiple functions may be combined together to answer important questions in biology. We apply the web services for accessing genomes, sequence similarities, metabolic neighborhood, subsystems, and metabolic models to identify genes that may be associated with gap filled reactions in metabolic models. Next, we combine genome feature queries, subsystems queries, expression data queries, and annotation queries to identify commonly clustered and co-expressed sets of functional roles across all known genomes. Finally, we combine flux variability analysis, gene knockout studies, essentiality data queries, and annotation queries to study redundancy of essential metabolic functions across all known genomes.

## 249

### Enabling a Systems Biology Knowledgebase with Gaggle and Firegoose

J. Christopher Bare[1]* (cbare@systemsbiology.org), Karen Foley,[1] Tie Koide,[2] David J. Reiss,[1] Paul T. Shannon,[1] Dan Tenenbaum,[1] Steven M. Yannone,[3] Sung Ho Yoon,[1] Wei-Ju Wu,[1] and **Nitin S. Baliga**[1] (nbaliga@systemsbiology.org)

[1]Institute for Systems Biology, Seattle, Wash.; [2]Universidade de São Paulo, Ribeirão Preto, SP, Brazil; and [3]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.

http://gaggle.systemsbiology.net
http://baliga.systemsbiology.net/enigma/

**Project Goals: This project will build on previous iterations of the Gaggle framework to provide a means of interoperability between bioinformatics software and data sources that enables independently developed tools to be used together for exploratory analysis of high-throughput data. The framework will provide interoperability that is simple to implement and extensible to accommodate unforeseen applications.**

Systems biology is made possible by technological advances in instrumentation on which thousands of measurements can be made simultaneously. Parallel advances in computing are critical to interpreting high-throughput biological data and developing an understanding of the fundamental molecular mechanisms that define the cell's regulatory circuitry.

In the Baliga lab at the Institute for Systems Biology we are reverse engineering biological circuits to understand how cells adapt to changing environments. Our long term vision is to build predictive mathematical models which can be applied to bioenergy, and bioremediation. This requires measuring and analyzing several different types of biological data: DNA sequence, transcription, protein abundance and protein-DNA binding.

To this end, we have developed software tools for exploratory analysis of systems biological data and for inference of the regulatory networks that control gene expression. Network inference algorithms, cMonkey and Inferelator, discover coregulated modules of cellular functionality and their regulators based on gene expression and de novo regulatory motif detection.

The Gaggle framework and Firegoose provide interoperability between data sources, web applications and desktop software for exploration, analysis and visualization of the resulting large biological networks in the context of underlying raw data. The database integration and software interoperability enabled by Gaggle and Firegoose is going to be a key aspect of a Systems Biology Knowledgebase that takes into account the unanticipated advances in technologies for measuring new kinds of data.

Funding: This work conducted by ENIGMA 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. Prior support provided by U.S. Department of Energy, Award No. DE-FG02-07ER64327 and DG-FG02-08ER64685.

# 250

## The PhyloFacts Microbial Phylogenomic Encyclopedia: Phylogenomic Tools and Web Resources for the Systems Biology Knowledgebase

Kimmen Sjölander* (kimmen@berkeley.edu)

University of California, Berkeley

**Project Goals: PhyloFacts is designed to improve the accuracy of functional annotation of microbial genomes through evolutionary reconstruction of gene families, integrating information from protein 3D structure, biological process, pathway association, protein-protein interaction and other types of experimental data to improve both the specificity and coverage of protein "function" prediction. Key computational challenges that we will address in this project include the development of a system for whole genome functional annotation and simultaneous taxonomic and functional annotation of metagenome datasets using HMMs (hidden Markov models) placed at internal nodes of gene family trees. A key component of our system is a novel phylogenomic approach to ortholog identification: Berkeley PHOG. Two genes are orthologous if they are descended from an ancestral gene by speciation. Orthologous genes are generally assumed to share a common function, while the functions of paralogous genes (related by duplication) can be anticipated to have diverged from that of the common ancestor.**

Gene families evolve novel functions through diverse evolutionary processes, including point mutations, gene duplication, and changes in domain architecture resulting from gene fusion and fission events. While some of these evolutionary events produce relatively small changes to function (e.g., mutations at positions that are distant from an enzyme's active site), others (especially duplication events and domain architecture changes) can result in dramatic shifts in the molecular function and biological process. As a result, homology-based functional annotation, i.e., transferring the annotation of the top BLAST hit, has been shown to be associated with significant errors: upwards of 20-25% of sequences have been estimated to have errors in their functional annotations.

The combined use of evolutionary reconstruction and structural analyses helps avoid the errors in sequence functional annotation that are now known to be rampant. We take this approach, termed *structural phylogenomics* (Sjölander, "Getting started in Structural Phylogenomics", *PLoS Computational Biology*, 2010) in constructing the PhyloFacts Microbial Phylogenomic Encyclopedia (at http://phylogenomics.berkeley.edu/phylofacts/). PhyloFacts is designed to improve the accuracy of functional annotation of microbial genomes through evolutionary reconstruction of gene families, integrating information from protein 3D structure, biological process, pathway association, protein-protein interaction and other types of experimental data to improve both the specificity and coverage of protein "function" prediction.

Key computational challenges that we will address in this project include the development of a system for whole genome functional annotation and simultaneous taxonomic and functional annotation of metagenome datasets using HMMs (hidden Markov models) placed at internal nodes of gene family trees.

A key component of our system is a novel phylogenomic approach to ortholog identification: Berkeley PHOG (Datta *et al*, "Berkeley PHOG: PhyloFacts Orthology Group Prediction Web Server" *Nucleic Acids Research* 2009). Orthology is a phylogenetic term: two genes are orthologous if they are descended from an ancestral gene by speciation. Orthologous genes are generally assumed to share a common function, while the functions of paralogous genes (related by duplication) can be anticipated to have diverged from that of the common ancestor. We identify orthologs based on reconstructed evolutionary histories for proteins clustered on the basis of sharing a common domain architecture using the FlowerPower algorithm (Krishnamurthy et al, "FlowerPower: clustering proteins into domain architecture classes for phylogenomic inference of protein function", *BMC Evolutionary Biology* 2007), and also for individual PFAM domains (in which cases proteins may have different overall domain architectures). The PHOG algorithm can be tuned to user-specified taxonomic distances, allowing highly precise predictions of orthology.

We are in the process of expanding PhyloFacts coverage of key microbial genomes, with extensions to the pipeline described in (Krishnamurthy *et al*, "PhyloFacts: An online structural phylogenomic encyclopedia for protein functional and structural classification", *Genome Biology* 2006) and will provide different ways of accessing the data for online (web-based) analysis and to download the core data for use at external sites.

Grant Number: De-Sc0004916

# 251

## Integrated Approach to Modeling Transcriptional Regulatory Networks

Pavel Novichkov[1]* (PSNovichkov@lbl.gov), Dmitry Rodionov,[2] Marcin Joachimiak,[1] David J. Reiss,[3] **Adam P. Arkin**,[1] **Inna Dubchak**,[1] and **Nitin Baliga**[3]

[1]Lawrence Berkeley National Laboratory Berkeley Calif.; [2]Burnham Institute for Medical Research La Jolla, Calif.; and [3]Institute for Systems Biology Seattle, Wash.

**Project Goals: Integrate computational approaches for reconstruction of large-scale transcriptional regulatory networks and building predictive models in microbial genomes.**

One of the major challenges facing the bioinformatics community in view of advances in DNA sequencing technology and constantly growing number of complete genomes is providing effective tools to enable high-quality reconstruction of transcriptional regulatory networks (TRN) and building a predictive models for transcriptional control of physiology in microbial organisms. This challenging problem can be addressed using two complementary approaches: i) a data-driven systems approach that relies on the integration of diverse data including high-throughput gene expression and protein-protein interactions data; ii) a genomic-driven approach based on the simultaneous analysis of several closely related genomes to decipher evolutionary conserved regulatory signals. Recently we were able to demonstrate that both approaches allow us to get deep insight into transcriptional regulation of microbial organisms and build large-scale TRNs and high-quality predictive models.

Using the data-driven systems approach we have constructed a large-scale predictive model for transcriptional control of physiology in *H. salinarum NRC-1*, which represents a class of poorly studied organisms (Archaea). Using relative changes in 72 transcription factors (TFs) and 9 environmental factors (EFs) this model accurately predicts dynamic transcriptional responses of ~80% genes in 147 newly collected experiments representing completely novel genetic backgrounds and environments.

A key aspect of the data-driven systems approach is the reduction of data complexity and the automated formulation of hypothesis based on multiple lines of evidence. We have developed an associative biclustering method to identify a variety of types of coherent patterns in combined functional genomics data. The method searches for statistical associations and estimates confidence across multiple data types. Results of searches incorporating data on transcription, such as gene expression microarrays, provide direct information on putative regulons. The method is placed in a computational framework, which allows rapid customization and deployment for new data sets and data types.

Finally, using comparative genomics approach we were able to build detailed large-scale TRNs in groups of closely related genomes representing diverse bacterial taxa, including *Shewanella* (16 genomes, 82 TFs), *Desulfovibrionales* (10 genomes, 32 TFs), *Bacillales* (11 genomes, 105 TFs), *Enterobacteriales* (12 genomes, 58 TFs), etc. The typical large-scale TRN covers a representative set of metabolic pathways and biological processes. For instance, *Shewanella* TRN encodes regulation of metabolism of carbohydrates, nitrogen and aminoacids, fatty acids and nucleotides, cofactors and metals, stress response, etc. The analysis gene expression profiling data of TF mutants, which have been recently done for three transcription factors FUR, PerR and Rex in *Desulfovibrio desulfuricans G20*, shows nice agreement with regulons reconstructed by comparative genomics.

TRNs built using data-driven systems approach contains a wealth of knowledge about the putative regulatory interactions that can be further analyzed in details by comparative genomics and automatically propagated to closely related species. On the other hand, regulons reconstructed from genomics-driven approach provide detailed description of the regulon content, assigned transcription factor and TF binding sites in promoter regions of target genes, and thus creates a solid ground for building predictive models. Thus we believe that tight integration of both strategies is required to achieve significant improvement in both coverage and quality of transcriptional regulatory networks. The iterative procedure of TRN and predictive model refinement will be implemented by means of establishing the central repository of putative regulons, which will serve as a source and sink of putative regulons predicted by different approaches.

# 252

## Large-Scale Genomic Reconstruction of Transcriptional Regulatory Networks in Bacteria

Pavel Novichkov[1]* (PSNovichkov@lbl.gov), Alexey Kazakov,[1] Semen Leyn,[2] Dmitry Ravcheyev,[2] Andrei Osterman,[2] Inna Dubchak,[1] Adam Arkin,[1] and **Dmitry Rodionov**[2] (rodionov@burnham.org)

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif.; and [2]Sanford-Burnham Medical Research Institute, La Jolla, Calif.

**Project Goals: The major goals of this project are to: (1) develop integrated platform for genome-scale regulon reconstruction; (2) Infer regulatory annotations and reconstruct transcriptional networks in several groups of microbial species important for DOE mission; (3) Develop RegKnowledgebase on microbial transcriptional regulation.**

Genome-scale annotation of regulatory features of genes and reconstruction of transcriptional regulatory networks in a variety of diverse microbes is one of the critical tasks

of modern Genomics and Systems Biology. It constitutes an important challenge, a prerequisite for understanding molecular mechanisms of transcriptional regulation in prokaryotes, identifying regulatory circuits, and interconnecting them with each other and with various metabolic, signaling, and other cellular pathways. A growing number of complete prokaryotic genomes allows us to extensively use comparative genomic approaches to infer *cis*-acting regulatory elements in regulatory networks of numerous groups of bacteria.

We developed a computational genomic-based approach implemented in the RegPredict web-server (regpredict.lbl.gov) facilitating fast and accurate inference and analysis of microbial regulons controlled by either DNA-binding transcription factors (TFs) or RNA regulatory elements (e.g., riboswitches). A key new concept of RegPredict is a cluster of co-regulated orthologous operons (CRON) that allows prediction of TF-binding sites (TFBSs) simultaneously in a set of taxonomically related genomes in a semi-automated way and provides a user-friendly GUI to perform comparative analysis of multiple CRONs. Major directions for genomics-based regulon inference are: (i) *regulon reconstruction* for a known regulatory motif in a set of reference genomes from a particular taxonomic group of bacteria; (ii) *ab initio prediction* of novel regulons using several scenarios for the generation of starting gene sets; (iii) *conservative propagation* of reconstructed regulons to all other microbial genomes in the same taxonomic group. The results of regulon inference performed by any implemented workflow are amenable for immediate deposition in the RegPrecise database (regprecise.lbl.gov) (Fig. 1).

We applied the integrative comparative genomics approach to infer transcriptional regulatory networks (TRNs) in various taxonomic groups of bacteria. A limited input of established regulon members in model species, e.g. *E. coli*, *B. subtilis*, *S. aureus*, was extracted from literature and publically available resources, such as RegulonDB and DBTBS databases. The obtained reference set of inferred regulons is available in the RegPrecise database and includes 420 regulons described in 94 genomes from the 9 taxonomic groups including Enterobacteria, *Bacillus*, *Shewanella*, *Streptococcus*, *Staphylococcus*, *Ralstonia*, *Desulfovibrio*, *Thermotoga*, and Cyanobacteria (Table 1). The reconstructed regulons control the key pathways involved in central metabolism, production of energy and biomass, metal homeostasis, stress response and virulence. Many of the *de novo* inferred regulons were experimentally validated in *S. oneidensis* and *T. maritima* models. The taxon-specific regulon collections will be further expanded to produce draft TRNs for other taxonomic groups of Proteobacteria and Firmicutes. To cover all sequenced genomes (e.g. strains of the same species) within the analyzed taxonomic groups (~1000 genomes), we will use automated conservative propagation procedure in the context of RegPrecise.

Overall, this project enables a wide spectrum of capabilities required by DOE Systems Biology KnowledgeBase incuding: (i) detailed TRNs for DOE-mission genomes; (ii) knowledgebase of regulatory interactions in a large set of microbial genomes; (iii) regulatory constraints for building predictive models; (iv) framework for validation and extension of gene regulatory networks reconstructed from gene expression profiling data.

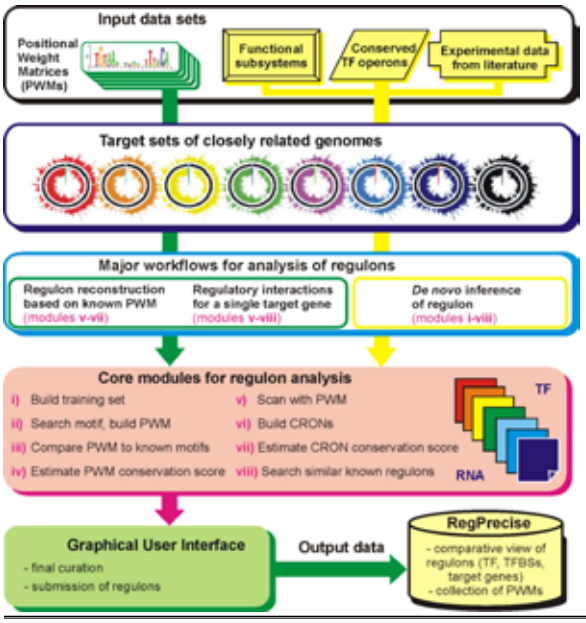



Figure 1. Regulon reconstruction workflow implemented in the RegPredict web-server and RegPrecise database. Two major components of this workflow are propagation of previously known regulons from model organisms to others, and *ab initio* prediction of novel regulons. RegPredict tool (http://regpredict.lbl.gov) applies the comparative genomic approach to the analysis of bacteria from a single taxonomic group of species. The inferred reference sets of regulons for various taxonomic groups of bacteria are collected and represented in the RegPrecise database (http://regpredict.lbl.gov).

Table 1. Taxonomic groups of bacteria targeted to genome-wide regulon reconstruction

| Taxonomic group | Analyzed genomes | Inferred TF regulons | Example model species |
|---|---|---|---|
| Bacillaceae | 11 | 105 | *Bacillus subtilis* |
| *Shewanella* | 16 | 82 | *Shewanella oneidensis* |
| Enterobacteria | 12 | 58 | *Escherichia coli* |
| *Staphylococcus* | 6 | 47 | *Staphylococcus aureus* |
| *Streptococcus* | 8 | 39 | *Streptococcus pneumoniae* |
| Thermotogales | 11 | 32 | *Thermotoga maritima* |
| Desulfovibrionales | 10 | 32 | *Desulfovibrio vulgaris* |
| *Ralstonia* | 6 | 15 | *Ralstonia eutropha* |
| Cyanobacteria | 14 | 10 | *Synechococcus sp.* PCC7002 |
| **TOTAL** | **94** | **420** | |

# 253

## Elucidation of the Transcriptional Response to Key Physiological Parameters in Bacteria: A Systems Biology Knowledgebase of Transcriptional Regulation

Stephen A. Federowicz* (sfederow@ucsd.edu), Young Seoub Park, Eric M. Knight, Donghyuk Kim, Byung-Kwan Cho, **Karsten Zengler**, and **Bernhard Ø. Palsson**

Department of Bioengineering, University of California – San Diego, La Jolla

http://systemsbiology.ucsd.edu

**Project Goals: This project aims to: (1) create a fully curated, bottom up reconstruction of the transcriptional regulatory network in bacteria, using *Escherichia coli* as a model organism, (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:**
Global transcription factors represent one of the primary means by which an organism can sense and respond to stimuli. We experimentally determined the network level mechanisms of transcriptional response in *E. coli* by integrating genome wide binding analysis of five global transcription factors with associated expression profiling and transcription start site information. ChIP-chip experiments for ArcA, Fnr, FruR, Crp, Lrp, and ArgR revealed 143, 100, 45, 247, 144, and 63 unique binding regions, respectively. Binding peaks were mapped to gene expression data generated between a wild type strain and a gene deletion strain for each of the transcription factors to discern whether a binding event resulted in gene activation or repression. This allowed us to construct a functional network of transcriptional regulation that elucidates the response to major environmental stimuli. This includes the response to oxygen availability mediated by ArcA and Fnr, shifting carbon source mediated by FruR and Crp, and primary nitrogen source mediated by Lrp and ArgR. Subsequently, we were able to elucidate core connected feed back loops through which regulatory information flows and gene expression is systematically controlled. Specifically we found that arginine and leucine act as signaling molecules by shutting down their own transport and biosynthetic pathways in response to high levels of exogenous arginine or leucine. In contrast, tryptophan and tyrosine were shown to shut down only their own biosynthesis pathways, but not their own transport systems, suggesting that they function simply as nutrients to the cell. We thus propose the existence of specific network motifs containing multiple connected feedback loops that govern the stimulatory response in *E. coli.*

Furthermore, we integrated these binding regions with experimentally determined transcription start site (TSS) data using 5'-RACEseq in order to determine regulation of alternative transcripts and individual promoter architectures. First, we identified precise transcription factor binding locations via DNA sequence motifs. We were able to clearly identify previously well-determined binding motifs for each of the transcriptions factors and then used position weight matrices to rescan the entire genome. Aligning these sequence motifs to the TSS information generated a set of over 700 experimentally derived promoter architectures. Each of these contains one or more TF binding sites and a single TSS for each transcription unit. These were then clustered to reveal conserved patterns of promoter architecture and associated transcriptional regulation. Notably, we were first able to clearly recapitulate and expand upon the well-known promoter architectures for Crp. In depth analysis of other global transcription factors revealed a wealth of novel individual and shared distance preferences. Many of the most distinct preferences occurred directly over the TSS corresponding to direct repression in the case of ArcA, Lrp, and FruR. Others occurred around the -35bp region or further upstream around -60bp and were well correlated with transcriptional activation. Taken together these results suggest the presence of modular nucleoprotein complexes which confer transcriptional activation or repression.

Overall, using promoter level information in concert with network level motifs allows for the creation of fully curated, bottom up reconstructions of global transcriptional regulatory networks. These networks represent a curated systems biology knowledgebase of the transcriptional regulatory machinery within a cell and allow for physiologically meaningful computations of environmental perturbation.

# 254

## Integrative Reconstruction of Carbohydrate Utilization Networks in Thermotogales

Dmitry Rodionov[1]* (rodionov@burnham.org), Vasiliy Portnoy,[2] Xiaoqing Li,[1] Irina Rodionova,[1] Dmitry Ravcheyev,[1] Olga Zagnitko,[3] Gordon Push,[3] **Yekaterina Tarasova,**[2] Kenneth Noll,[4] Robert Kelly,[5] Karsten Zengler,[2] **Andrei Osterman,**[1] and **Bernhard Palsson**[2]

[1]Sanford-Burnham Medical Research Institute, La Jolla, Calif.; [2]Department of Bioengineering, University of California San Diego, La Jolla, Calif.; [3]Fellowship for Interpretation of Genomes, Burr Ridge, Ill.; [4]Department of Molecular and Cell Biology, University of Connecticut, Storrs; and [5]Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh

**Project Goals: The major goals of this project are to: (1) utilize bioinformatics and computational tools to enhance, refine and fill in the knowledge gaps in the current metabolic reconstruction of *T. maritima* with a particular emphasis on hydrogen production; (2) use comparative genomics to infer the key transcription factor**

**regulons for carbohydrate metabolism in *T. maritima* and other Thermotogales.**

Bacteria of the deep-branched genus *Thermotoga* can produce hydrogen by fermenting a wide range of carbohydrates. A remarkable diversity of the Thermotogales *sugar diet* is matched by a large fraction of genes committed to carbohydrate degradation and utilization in their genomes. As a result the evolutionary plasticity of the sugar catabolic machinery in general, and due to a unique taxonomic position and lifestyle of Thermotogales, exact functions of many respective genes (and pathways) remained unclear even in the best-studied model species, *T. maritima*. To address this problem we applied an *integrative subsystems-based approach to the genomic reconstruction of metabolic and regulatory networks*. This approach established and validated in our previous studies (e.g. in the reconstruction of sugar catabolic machinery of the *Shewanella* genus) included three major levels of integration: (i) comparative analysis of 11 complete genomes from the Thermotogales order annotated by RAST server; (ii) parallel genomic reconstruction of biochemical transformations, uptake mechanisms and transcriptional regulation, and (iii) combining bioinformatic predictions with experimental testing in *T. maritima* model. Application of various bioinformatics tools implemented in the SEED (theseed.uchicago.edu) and RegPredict (regpredict.lbl.gov) web-sites allows us to substantially improve the accuracy of annotations as well as predict novel, previously uncharacterized genes and pathways. The developed detailed reconstruction integrates published and new results obtained by several research groups collaborating within the framework of the Biological Hydrogen Production Program (DOE DE-PS02-08ER08-12).

The *genomic encyclopedia of sugar utilization* in Thermotogales includes more than 300 functional roles (isofunctional protein families or *FIGfams*) spanning at least 20 distinct pathways with mosaic distribution across 11 analyzed species. The detailed results of this analysis are captured in the subsystem "*Sugar utilization in Thermotogales*" available online from the SEED web-site. The current version of the subsystem comprises >130 cytoplasmic and extracytoplasmic sugar catabolic enzymes (including ~40 glycoside hydrolases), ~90 components of carbohydrate uptake systems (mostly ABC transporters), and 18 committed transcription factors. The developed metabolic model incorporates all published experimental data and inferences about enzyme activities, substrate specificities of transporters, and differential gene expression patterns on various carbohydrates (generated mostly for *T. maritima*). Our analysis revealed substantial differences in sugar catabolic pathways between Thermotogales and other previously studied bacteria. Most common are *nonorthologous gene replacements*, when a functional role is encoded by a gene, which is not orthologous (and, often, nonhomologous) to any of the previously described genes of the same function. The repertoire of transporters and regulators involved in sugar catabolism in *Thermotoga* demonstrates the most prominent differences in comparison with other taxa. We also discovered two novel pathways for utilization of inositol and galacturonate in *T. maritima*. These pathways include previously unknown biochemical transformations driven by 4 and 6 enzymes, respectively. Both pathways as well as most of the individual recombinant purified enzymes were experimentally characterized by enzymatic assays and genetic complementation methods. A performed substrate specificity profiling of the *T. maritima sugar kinome*, 15 predicted carbohydrate kinases over a panel of >40 diagnostic sugar substrates, provided a strong experimental support of the reconstructed pathways. Finally, the predicted catabolic capabilities of *T. maritima* were assessed by monitoring growth rates, substrate consumption and gene expression on a panel of various individual and mixed mono- and disaccharides.

A transcriptional regulatory network inferred from comparative genomic analysis of Thermotogales includes 32 transcription factors and their DNA binding sites unevenly distributed across 11 studied genomes. A current collection of regulons captured in the RegPrecise database (regprecise.lbl.gov) is centered on *T. maritima* and includes 18 transcription factors that were predicted to control expression of ~185 genes involved in sugar catabolic machinery of this model organism. Remarkably, a large fraction of these genes and operons are controlled by multiple transcription factors pointing to a complexity of regulatory responses to changing environmental conditions (Fig. 1). For example, we established partial overlaps between the xylose, glucuronate, and galacturonate regulons (XylR, KdgR, and UxaR, respectively); the glucose, trehalose and inositol regulons (GluR, TreR, and IolR); and the cellobiose, mannose, and glucooligosaccharide regulons (CelR, ManR, and GloR). The experimental assessment of the reconstructed regulatory network included *in vitro* analysis of selected individual regulons and *in vivo* gene expression profiling of *T. maritima* on various carbohydrate substrates. We used the first approach based on gel-shift mobility assays to validate all predicted DNA targets and identify small molecule effectors for six regulators from the ROK family (BglR, IolR, XylR, ChiR, TreR, and ManR). We are currently expanding this effort to characterize additional transcriptional regulators, Rex, UxaR, KdgR, CelR, and RhaR. Global gene expression profiles were obtained and analyzed for the growth on 12 different carbon sources using high-density oligonucleotide tiling arrays (Nimblegen). Gene induction patterns measured for tested mono- and disaccharides (trehalose, rhamnose, xylose, etc.) showed a strong correlation and provided additional information to refine respective regulons (TreR, RhaR, XylR, etc.) reconstructed by the genomic analysis.

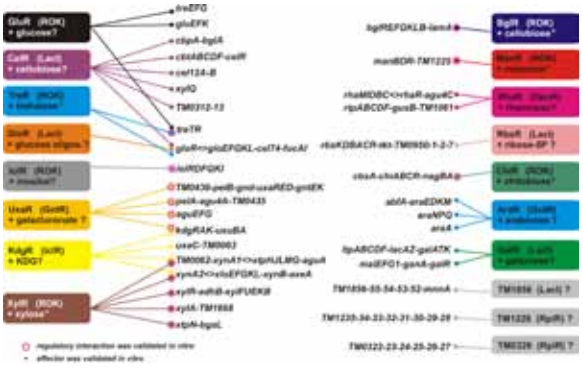


Figure 1. Reconstructed regulatory network for carbohydrate utilization genes and 18 transcription factors in *Thermotoga maritima*. Abbreviations: transcription factors (TFs) are shown in colored boxes; predicted regulatory interactions between a TF and its target operons are shown by lines; experimentally confirmed regulatory interactions are marked by magenta circles; TF protein family is denoted in parentheses; predicted small molecule effectors are listed after '+'; confirmed effectors are marked by asterisks.

*Overall*, this study yielded numerous insights into physiology, biochemistry and transcriptional regulation of the carbohydrate utilization machinery of *T. maritima* and other Thermotogales. It also provided additional validation of the established integrative approach to the genomics-driven reconstruction of metabolic and transcriptional regulatory networks, which can be applied to many other, yet unexplored, biological systems.

# 255

## Variability of Transcriptional Regulatory Network in *Desulfovibrio* Species

Alexey Kazakov[1]* (AEKazakov@lbl.gov), Inna Dubchak,[1] **Dmitry Rodionov**,[2] and **Pavel Novichkov**[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif.; and [2]Sanford-Burnham Medical Research Institute, La Jolla, Calif.

**Project Goals: Reconstruction of transcriptional regulatory network in DOE-mission genomes.**

Use of sulfate-reducing bacteria for bioremediation of heavy metal-contaminated media has economical and environmental significance. Sulfate-reducing bacteria are a matter of big concern in engineering due to problems with corrosion of metal structures. Some bacteria of *Desulfovibrionales* order are also of clinical interest because they may act as opportunistic pathogens.

In this study we carried out large-scale comparative genomics analysis of regulatory interactions in *Desulfovibrio vulgaris* and 10 related species from *Desulfovibrionales* order using the integrated RegPredict web-server for regulon reconstruction (http://regpredict.lbl.gov/).

The resulting regulatory network contains 40 regulons controlled by transcription factors (TF) from 13 TF families, more than 1,300 binding sites and 4,000 target genes involved in stress response, amino acid metabolism, metal homeostasis. The analysis of gene expression profiling data of TF mutants, which have been recently done for two transcription factors FUR (iron homeostasis) and PerR (peroxide stress response) in *Desulfovibrio desulfuricans G20*, shows nice agreement with regulons inferred by comparative genomics. The analysis of correlations between multiple microarray expression profiles provides additional information for the assessment of regulon predictions as demonstrated by analysis of MetR regulon.

The detailed analysis of three large TF families ArsR, CRP, and GntR revealed substantial variations in regulons among *Desulfovibrionales* genomes. Phylogenetic analysis of these transcription factors showed that 72 regulators from 9 sub-families are well conserved in the analyzed genomes, whereas 60 regulators are poorly conserved (present in less than 3 genomes) and showing a mosaic distribution. This pattern is most likely a result of multiple evolutionary events such as horizontal gene transfer, gene loss, and gene duplication. For regulon reconstruction in the latter group of poorly conserved TFs, the standard orthology-based approach was inefficient, and thus we utilized a modified comparative genomics approach. We found closest TF homologs outside of the analyzed group of genomes, and applied motif detection procedure for set of genes in the conserved genomic neighborhood. As a result we were able to identify binding motifs for 40 poorly conserved TFs and for all well-conserved TF sub-families. Within each TF family we observed fair similarity between TF binding motifs. This observation can be utilized for development of automatic approach for large-scale inference of poorly conserved regulons.

An overall reference collection of 40 *Desulfovibrionales* regulogs can be accessed through RegPrecise database (http://regprecise.lbl.gov/).

# 256

## Computational Modeling of Fluctuations in Energy and Metabolic Pathways of Microbial Organisms

Elijah Roberts*[1] (erobert3@illinois.edu), Piyush Labhsetwar,[2] **Nathan D. Price**,[2,3] **Carl R. Woese**,[4] and **Zan Luthey-Schulten**[1,2] (zan@illinois.edu)

[1]Department of Chemistry, [2]Center for Biophysics and Computational Biology, [3]Department of Chemical and Biomolecular Engineering, [4]Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL

http://www.scs.illinois.edu/schulten

**Project Goals: In this project, we propose to develop a computational methodology for modeling individual microbes that connects stochastic reaction-diffusion dynamics of selected biochemical pathways under *in vivo* conditions with genome-scale modeling of metabolic and regulatory networks, techniques from systems biology. Using such a combined model, we propose that the dynamic behavior of a cell's entire biochemical network can be approximated in a way that will enable discovery of new phenomena regarding how cells respond to fluctuating environmental conditions.**

Stochastic expression of genes produces heterogeneity in clonal populations of microorganisms under identical conditions. We analyze and compare the behavior of the inducible *lac* genetic switch using well-stirred and spatially resolved simulations for *Escherichia coli* cells modeled under fast and slow-growth conditions. Our new kinetic model describing the switching of the *lac* operon from one phenotype to the other incorporates parameters obtained from recently published *in vivo* single-molecule fluorescence experiments along with *in vitro* rate constants. For the well-stirred system, investigation of the intrinsic noise in the circuit as a function of the inducer concentration and in the presence/absence of the feedback mechanism reveals that the noise peaks near the switching threshold. Applying maximum likelihood estimation, we show that the analytic two-state model of gene expression can be used to extract stochastic rates from the simulation data. The simulations also provide mRNA-protein probability landscapes, which demonstrate that switching is the result of crossing both mRNA and protein thresholds. Using cryoelectron tomography of an *E. coli* cell and data from proteomics studies, we construct spatial *in vivo* models of cells and quantify the noise contributions and effects on repressor rebinding due to cell structure and crowding in the cytoplasm. Compared to systems without spatial heterogeneity, the model for the fast-growth cells predicts a slight decrease in the overall noise and an increase in the repressor's rebinding rate due to anomalous subdiffusion. The tomograms for *E. coli* grown under slow-growth conditions identify the positions of the ribosomes and the condensed nucleoid. The smaller slow-growth cells have increased mRNA localization and a larger internal inducer concentration, leading to a significant decrease in the lifetime of the repressor-operator complex and an increase in the frequency of transcriptional bursts.

Funding: This work was supported by U.S. Department of Energy (Office of Biological and Environmental Research) under award DE-FG02-10ER6510.

# 257

## Microbial Data Integration and Metagenomic Workflows and the Development of the Biological Resource Network

Brian C. Thomas[1]* (bcthomas@berkeley.edu), Chongle Pan,[2] Itai Sharon,[1] Vincent Denef,[1] Robert Hettich,[2] Trent Northen,[3] Ben Bowen,[3] and **Jill Banfield**[1]

[1]University of California, Berkeley; [2]Oak Ridge National Laboratory; and [3]Lawrence Berkeley National Laboratory

**Project Goals: Our goal is to construct a network of integrated metagenomic, proteomic, metabolomic, and transcriptomic data from microbial communities. Additionally, we will develop methods for analyzing, visualizing, and understanding the complex and highly redundant nature of these systems. Our research will generate a scalable microbial community knowledge base that addresses the challenges of complex systems.**

Microbial systems biology relies upon the coordination of multiple data types, such as genomic, proteomic, metabolomic, and transcriptomic data. Commonly, metagenomic workflows begin with just sequence information, but quickly expand to include a variety of data sources, including metadata such as geochemical measurements, time of sample collection etc. Integration of these other data streams is challenging, yet their dynamic interconnection is a critical goal in bioinformatics and will help provide a functional understanding of the microbial systems. Effective data integration and handling must also address the reality that data (especially gene calls, functional annotations, assignment of genes to organisms within a community, etc.) can exist in many states – live (updated), stale (out of date), dead (wrong), lost, etc. The goal of our DOE Kbase project is to make sense of the breadth and diversity of biological data from a single system by providing a knowledge base and workflows to facilitate integration of 'live' data and promote data access.

Towards this goal, we present an update on our development of a microbial data integration system, called the **B**iological **R**esource **N**etwork (BRN). BRN is a network of data resources—spanning genomic, proteomic, metabolomic and transcriptomic data types—interlinked via an underlying software system based on representational state transfer (REST, or "RESTful") services. RESTful software architectures simplify component implementation, reduce the complexity of component semantics, improve the effectiveness of performance tuning, and increase the scalability of components. BRN components are separated by data type:

our metagenomic service is separate from our proteomic service or metabolomic service etc. This modular architecture allows us to scale individual components as data content or complexity demands change, and allows key design choices for a resource to be made by component experts. This feature is well suited for cloud-based computing, where new nodes can be allocated to BRN components as demand increases (for example, during a data update stage or during a normalization calculation step). Integration of the various BRN components is achieved with a unified application-programming interface (API). Requests can be made to the individual BRN components if a specific type of data is desired, however, using the API any service in the BRN can make proxy requests to other BRN-aware services. For example, the metagenomics resource is informed of the proteomics resource: It can therefore respond to requests such as "retrieve proteins from metagenomic contig 'Lepto2_contig42' with uniquely expressed proteins found in sample 'AMD_FloatingBiofilm_A'". This simple query represents a joining of two diverse data sets originating from the same sample—the BRN imposes no restrictions on the individual resources with regard to data integrity or format. This is the opposite of what is commonly found in federated data repositories, where data from diverse sources are combined into a single large entity, separated from the original, "live" data. This splintering of data from its maintainers immediately renders it, at the very least, "stale." Federated databases do not benefit from updates made to the original data source in real time. BRN components can be individually scaled and are not required to be served from the same location: BRN web services are connected via the Internet, allowing resources located anywhere in the world to be linked into the Network. Currently the BRN contains services for metagenomics and proteomics resources (hosted at U.C. Berkeley) and initial progress has been made incorporating a metabolomics resource (hosted at LBNL).

GeneGrabber is a database, web application and workflow for processing, analyzing, visualizing and serving metagenomic and linked data. As a BRN component, data are easily accessible using simple REST requests via HTTP (either from a web browser or programmatically using the BRN API). GeneGrabber has workflows that assist in data processing, import, and annotation updates. GeneGrabber is a "gene-centric" view of metagenomic data. As mentioned above, GeneGrabber can communicate with any other resource that has a RESTful interface (other BRN services or non-BRN resources, such as UniRef etc).

A critical component of any resource is access and incorporation of new data. Metagenomics begins with tens of millions of sequencing reads that must be assembled, refined, searched, annotated, and assigned to organisms. Workflows greatly help this endeavor. We have developed a metagenomics workflow for handling next-generation sequencing data. The approach has been refined during assembly of genomes from short read Illumina data for several strains of *Sulfobacillus*, all of them with little or no similarities to previously sequenced *Sulfobacillus* strains. New tools and strategies have been developed, including a novel iterative assembly strategy that optimizes assembly of genomes of different coverage depths from the same sample. Most importantly, we have established a scaffolding method that automates the manual assembly curation step—this new method has performed well in preliminary analyses. We have also refined an unsupervised clustering strategy for the purpose of data binning that is based on emergent self-organizing maps (ESOM). The approach has been extended to incorporate data types other than sequence signatures (e.g. temporal distribution patterns), enabling much improved genome classification.

Additionally, we have developed a workflow to process proteomics data using a large computer cluster at ORNL. We use the Sequest and DTASelect algorithms to process standard shotgun proteomics data. A new open-source algorithm, Sipros, was developed to process proteomic stable isotope probing data (Pan et al. In press). The results are integrated with sample metadata and integrated into the BRN using the ProteomeDB service. ProteomeDB enables analyzing proteomic results using information from other resources in the BRN. For example, protein co-expression patterns can be correlated with operon information retrieved from GeneGrabber

Initial progress towards our DOE Kbase project goal can be seen in GeneGrabber and ProteomeDB and their underlying RESTful API and we will provide demonstrations of the underlying data access. Future work will focus on expanding the BRN to include metabolomics (both GCMS and MS/MS), transcriptomics data resources and continued development of workflows facilitating data integration.

# 258

## Insights from the Reconstruction of 3000 Metabolic Models

**Christopher Henry**,[1,2]* **Ross Overbeek**,[3] Matt DeJongh,[4] Aaron Best,[4] Fangfang Xia,[1,2] Scott Devoid,[2] and **Rick Stevens**[1,2]

[1]University of Chicago, Ill.; [2]Argonne National Laboratory, Ill.; [3]Fellowship for Interpretation of Genomes, Ill.; and [4]Hope College, Mich.

**Project Goals: This project aims to develop software and data infrastructure to support programmatic access to the SEED and Model SEED resources for the annotation of genomes, integration of omics data, and reconstruction of genome-scale metabolic models. Within this overarching objective, are four specific aims: (i) enhancing the computational infrastructure behind the SEED framework to improve extensibility, accessibility, and scalability; (ii) integrate extensions into the SEED data-model and software to accommodate new data types including regulatory networks, genome-scale metabolic models, structured assertions, eukaryotic genome data, and growth phenotype data; (iii) develop an application programming interface (API) to provide remote access to the SEED database and tools; (iv) apply SEED infrastructure, data,**

**and APIs to annotate genomes and construct genome-scale metabolic models for organisms with applications to bioenergy, carbon cycle, and bioremediation.**

Genome-scale metabolic models have emerged as a valuable resource for generating predictions of global organism behavior based on the sequence of nucleotides in the genome. These models can accurately predict essential genes, organism phenotypes, organism response to mutation, and metabolic engineering strategies. Recently we developed the Model SEED resource (http://seed-viewer.theseed.org/models/) for the high-throughput reconstruction of new genome-scale metabolic models for microbial genomes. We demonstrated the ability of this resource to produce 130 new genome-scale models that are comparable in scale to existing published models. We also validated and optimized these models against available Biolog and gene essentiality data. Now we have applied the Model SEED to producing draft metabolic models for over 3000 microbial genomes, representing nearly all complete microbial genomes currently available in GenBank. New algorithms were developed for the gap-filling of these models to enable the activation of every possible reaction in the models; new algorithms were applied for the generation of biomass reactions based on completeness of annotated pathways for biomass precursors; and finally, new algorithms were applied for using the SEED tools to identify gene candidates that may be associated with the gap-filled reactions. This work reveals insights into the diversity of microbial genomes, the completeness of our knowledge of these genomes, and the areas of our knowledge where more gaps presently exist.

# 259

## Toward the Genomic Organization of Superorganisms: Trends in the Functional Content of Metagenomics Samples

Koon-Kiu Yan[1,2]* (koon-kiu.yan@yale.edu), **Sergei Maslov**,[3] and **Mark Gerstein**[1,2,4]

[1]Program in Computational Biology and Bioinformatics, [2]Department of Molecular Biophysics and Biochemistry, and [4]Department of Computer Science, Yale University, New Haven, Conn.; and [3]Department of Condensed Matter Physics and Materials Science, Brookhaven National Laboratory, Upton, N.Y.

**Project Goals: To better understand how natural habitats shape the organization of bacterial communities from a system-wide analysis of the trends between the average genome size and the proportion of different functionally categorized genes in metagenomic samples.**

In prokaryotes the proportion of genes attributed to particular cellular processes varies with the size of the genome. More specifically, it has been reported that in many high level functional categories the number of genes in each category scales $N_c$ as a power-law of the genome size ($N$), i.e. $N_c \sim N^{\alpha}$. These scaling laws, shaped by the underlying evolutionary processes, shed light on the organization of prokaryotic genomes. Nevertheless, in natural environments, bacterial species do not live in an isolated fashion but in forms of bacterial communities whose organizations are shaped by the interplay between species, for instance bacterial symbiosis. To a certain extent, a community of many bacterial species behaves as a superorganism, where individual species work collectively to survive. Toward this end, it is instructive to examine the functional content of the "genomes" of these superorganisms.

In this work, we explore the idea of superorganisms using metagenomic samples collected in different environmental habitats. We download metagenomic samples from sources such as the CAMERA database, and map the reads to curated functional categories such as the Clusters of Orthologous Groups (COG), and the Kyoto Encyclopedia of Genes and Genomes (KEGG). The ratio between, for instance, the number of reads corresponding to transcription factors and the total number of reads, is related to the scaling relationships described and we could infer parameters such as the average genome size of the sample. The inferred average genome size sheds light on whether large genomes or small genomes are favored in the corresponding habitat. We can further estimate, for each sample, the average genome size of species and the proportion of reads mapped to different functional categories and look for systematic trends similar to scaling laws demonstrated among individual genomes. Trends in superorganisms will be compared with the corresponding trends among individual genomes.

# 260

## Numerical Optimization Algorithms and Software for Systems Biology: An Integrated Model of Macromolecular Synthesis and Metabolism of *Escherichia coli*

**Ines Thiele**[1]* (ines.thiele@gmail.com), R.M.T. Fleming,[1] Stefan Gretar Thorleifsson,[1] A. Bordbar,[2] R. Que,[2] and B.O. Palsson[2]

[1]Center for Systems Biology, University of Iceland, Reykjavik, Iceland; and [2]Bioengineering Dept. University of California, San Diego, La Jolla

**Project Goals: This project aims to reconstruct genome-scale models of metabolism and macromolecular synthesis and to develop algorithms capable of solving the resulting large, stiff and ill-scaled matrices. We aim to combine state of the art reconstruction and constraint-based modeling and analysis tools with high-end linear optimization solvers and convex flux balance analysis. The incorporation of thermodynamic information in addition to environmental constraints will allow an accurate assessment of feasible steady states. While we will prototype the reconstruction and algorithm developments with *Escherichia coli*, we will employ the resulting networks to**

**determine thermodynamically favorable pathways for hydrogen production by *Thermotoga maritima*.**

Systems biology aims to understand the mechanisms of the genotype-phenotype relationship by investigating interactions between cellular components. We constructed a high-resolution, gene-sequence dependent network of an *E. coli* cell accounting for major cellular processes, e.g., metabolism, transcription and translation using a bottom-up approach relying on genomic and biochemical information. Using a constraint-based modeling approach (COBRA), we found that changes in codon usage altered the fitness of the in silico *E. coli* in various environmental conditions. This result agrees with experimental observations that not only gene content but also gene sequence determine the environment an organism can occupy. A gene's nucleotide composition is governed by the demand-supply principle of codon usage and tRNA abundance. For the first time, we were able to model the genotype-environment-phenotype relationship while previous efforts suggested such relationship merely based on statistical analysis of codon bias and environmental niche preference.

# 261

## *Thermotoga maritima* Systems Biology Knowledgebase: A Computational Platform for Multi-Subsystem Reconstruction and Multi-Scale Stoichiometric Modeling

Joshua A. Lerman[1]* (jalerman@ucsd.edu), Daniel R. Hyduke,[1] Haythem Latif,[1] Vasiliy A. Portnoy,[1] Alexandra C. Schrimpe-Rutledge,[2] Joshua N. Adkins,[2] Richard D. Smith,[2] Ines Thiele,[3] **Michael A. Saunders,**[4] Karsten Zengler,[1] and Bernhard Ø. Palsson[1]

[1]Department of Bioengineering, University of California – San Diego, La Jolla; [2]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash., [3]Center for Systems Biology, University of Iceland, Reykjavík, Iceland., and [4]Department of Management Science and Engineering, Stanford University, Stanford, Calif.

http://systemsbiology.ucsd.edu

**Project Goals: This project aims to: (1) develop a genome-scale model of macromolecular synthesis for *Thermotoga maritima*, (2) fully integrate it with the metabolic model, (3) guide the development of algorithms capable of finding optimal steady-state solutions to the combined model despite its large, sparse, and ill-scaled constraint matrix, (4) achieve phenotypic predictions with respect to gene expression, (5) classify the failure modes of the model to prioritize the reconstruction of regulatory circuits, and (6) enable the direct mapping of omics data to aid in the determination of the governing constraints on biohydrogen production from organic waste products.**

**Project Description:**
Over the past decade, the process of reconstructing metabolic networks at the genome scale has become prevalent in molecular systems biology. There is growing interest in using these genome-scale models of metabolism as contextual frameworks for the analysis of the vast amount of omics data currently generated. Transcriptome and proteome data have been integrated with metabolic network models and have provided insight into regulation, the frugality of the genomic program under growth selection pressure, host-pathogen interactions, and drug-responses. However, the omics integration has so far only been accomplished in a Boolean or indirect fashion. This is because the metabolic network reconstructions in wide use employ a Boolean mapping between genetic loci and enzymatic activity termed the gene-protein-reaction (GPR) relationship.

We have previously shown that it is possible to construct a genome-scale model of RNA and protein expression based on a set of basic biochemical reactions. This process was first completed in *Escherichia coli*, and the resulting model was called the 'E'-matrix, which stands for the reconstruction of gene Expression.

Here, we completed a model of macromolecular synthesis for *T. maritima*, which has a relatively small genome of 1.8 Mb. This reconstruction relied heavily on an accurate transcription unit architecture (see abstract by Latif et al.), which detailed the machinery and regulatory conditions required for RNA synthesis. The metabolic network and the macromolecular synthesis network were subsequently merged into a single multi-scale model (termed the 'ME' matrix, for Metabolism and Expression). This ME matrix is shown to be computable and has unique capabilities compared to standard metabolic reconstructions. We found that the incorporation of the biochemical reactions underlying the expression of gene products within a metabolic network reconstruction allowed for the removal of the artificial Boolean GPRs and facilitated the simulation of variable enzyme concentrations, opening a range of new applications of flux balance analysis. For example, the explicit representation of transcription and translation provided an opportunity to directly employ quantitative transcriptomic and proteomic measurements as model constraints.

Ultimately, after adding mathematically derived constraints to couple certain reactions in this model, we achieved a model fully linking the functions of over 650 genes in *T. maritima*, representing around 35% of the entire genome. Note that the current metabolic reconstruction for *T. maritima* covers only 25% of the total genome. This model represents one of the most comprehensive models of any organism to date. Our formulation leads to a reduced dependence on artificial objective functions, such as the biomass objective function, which do not have a mechanistic biochemical basis. For example, nucleotides and amino acids are no longer drawn out of the cell in bulk. Instead, individual RNA and protein synthesis fluxes are decision variables in the optimization problem and the *in silico* cell must decide how it should invest its building blocks given a finite capacity to synthesize them.

We show promise of long-term applications of this type of model, including the prediction of the governing constraints on biohydrogen production from organic waste products, thermostable protein engineering, interpretation of adaptive evolution, and minimal genome design. Incorporating expression of macromolecules into a genome-scale model of metabolism resulted in a quantitative model for transcription and translational processes, but it did not approximate the regulatory and signaling mechanisms that control expression. Analysis of the failure modes of our model will help us prioritize regulons to add to the model to increase predictive power. When no such information is available, the model will lead to discovery in terms of the suggested potential for novel regulatory circuits. Interestingly, this may be immediately tractable in *T. maritima* as its genome supports relatively few transcriptional regulatory states, with only a limited number of transcription factors (see abstract by D. Rodionov et al.), some of which have been experimentally characterized.

Grant Information: Numerical Optimization Algorithms and Software for Systems Biology (DOE Award ER25917 / DE-PS02-08ER08-01)

# 262

## Deciphering the Growth Dynamics of *Shewanella oneidensis* by Integrating Metabolite and Gene Expression Profiles with Stoichiometric Modeling

Ed Reznik,[1]* Sara Baldwin,[1]* Qasim Beg,[2] and **Daniel Segrè**[1,2,3] (dsegre@bu.edu)

[1]Program in Bioinformatics, [2]Department of Biomedical Engineering, [3]Department of Biology, Boston University, Boston, Mass.

http://prelude.bu.edu

**Project Goals: This project is aimed at combining stoichiometric genome-scale modeling of metabolism in *Shewanella oneidensins* MR-1 with gene expression data and metabolic profiles to understand how metabolic regulation affects growth and byproduct secretion.**

A major challenge in systems biology is the integration of multiple types of data in order to understand the dynamics of complex biological processes. To this end, we present a flexible method for combining time-series measurements of growth rate, metabolite concentrations in the supernatant, and gene expression data with genome-scale metabolic modeling. We apply our approach to a batch growth experiment in which the bacterium *Shewanella oneidensis* MR-1 was observed during the transition from exponential to stationary phase, in minimal lactate medium.

In our algorithm, we implement a modified, time-dependent version of the GIMME approach, previously developed for mapping gene expression data onto flux balance models (1). GIMME searches for a set of fluxes that minimizes the inconsistency with gene expression data, based on a scoring function that considers "on" any gene with expression above a given threshold.

In the original version of GIMME, a universal threshold value was used to determine reaction activity across all genes. In our new method we use a large compendium of microarrays to determine a unique threshold for each gene, and we demonstrate that, by doing so, we improve the predictive power of the algorithm. In addition, taking advantage of our time series data, we link together successive optimizations using a dynamic flux balance analysis framework (2). This enables us to investigate how the metabolism of *S. oneidensis* adjusts over a prolonged period of time as nutrients become depleted in the environment.

Initial versions of our algorithm used biomass data and expression data to determine the distribution of fluxes over time. The flux predictions were compared with experimental measurements of metabolite concentrations in the media to assess the predictive capabilities our algorithm. We next wanted to assess which internal fluxes would need to change in order to optimally align our external flux predictions with experimental measurements. Towards this goal, we implemented a quadratic programming minimization of a global inconsistency score which includes deviations from both expression data and nutrient concentrations in the supernatant. This procedure allowed us to make specific biological hypotheses regarding which metabolic pathways would need to carry more flux in order to produce the observed metabolite profiles.

Our algorithm produces flux predictions that are qualitatively better than those obtained by a standard dynamic flux balance algorithm. In particular, we capture previously uncharacterized time-dependent profiles for pyruvate and acetate utilization and excretion. Based on our current results, we propose that this method can be extended to inform future efforts in metabolic modeling, with potential applications in metabolic engineering and microbial ecology.

### References

1. Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. *PLoS Computational Biology*. (2008) 4(5): e1000082.
2. Mahadevan R, Edwards J, Doyle FJ. Dynamic Flux Balance Analysis of Diauxic Growth in *Escherichia coli*. *Biophysical Journal*. (2002) 83(3) pp.1331-1340: doi:10.1016/S0006-3495(02)73903-9.

# 263

## Building and Testing an Open Source Platform for Spatio-Temporal Stoichiometric Modeling of Metabolism in Microbial Ecosystems

William J. Riehl[1]* (briehl@bu.edu), Niels Klitkord,[1] William Harcombe,[2] Christopher J. Marx,[2] Nathaniel C. Cady,[3] and **Daniel Segrè**[1,4] (dsegre@bu.edu)

[1]Graduate Program in Bioinformatics, Boston University, Boston, Mass.; [2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Mass; [3]College of Nanoscale Science and Engineering, University at Albany, Albany, N.Y; and [4]Department of Biology and Department of Biomedical Engineering, Boston University, Boston, Mass.

http://prelude.bu.edu

**Project Goals: The goal of this project is to build a framework for spatio-temporal flux balance modeling of multiple interacting microbial species. The platform will be used to study syntrophic interactions in microbial ecosystems. Predictions will be compared to experimental measurements performed using cell microprinting technology.**

Genome-scale models of microbial metabolism are becoming increasingly important in addressing questions related to bioenergy production, bioremediation, and carbon and nitrogen cycling in the biosphere. As automated gene annotation pipelines, network gap-filling algorithms, and high throughput experimental methods improve, it will become gradually feasible to model virtually any sequenced microbe using this approach. Yet, some of the most fundamental properties of natural microbial ecosystems crucially depend on aspects that are beyond the stoichiometries of individual biochemical species. These include contact- or metabolite-mediated interactions between different microbes, dynamical changes of the environment, spatial structure of the underlying geography and evolutionary competition between distinct subpopulations.

As part of the DOE Systems Biology Knowledgebase we are developing COMETS (**C**omputation **O**f **M**icrobial **E**cosystems in **T**ime and **S**pace), an open source, broadly applicable and user-friendly platform for performing spatially distributed time-dependent flux balance simulations of microbial ecosystems. By taking advantage of the computational efficiency of flux balance model calculations, we implement a spatially structured lattice of interacting metabolic subsystems. These subsystems represent a level of detail that is intermediate between a fine-grained single-cell modeling approach, and a global mean-field approximation. The COMETS simulation engine combines a modified version of dynamic flux balance analysis (dFBA[1]) with a finite differences approximation of diffusion dynamics, to simultaneously track the spatio-temporal fate of multiple environmental molecules and microbial species. The COMETS platform has the capacity to bridge multiple spatial and temporal scales, making it possible to observe long term dynamics of microbial populations growing in a given environmental setting, based on constant updates of local nutrient availabilities and exchanges, and ultimately determined by the metabolic activity of individual microbial species. Thus, it can be used as a platform for modeling the growth of a single bacterial species in a Petri dish, biofilm formation on complex substrate morphologies, seasonality of microbial communities in a specific geographical setting, or the growth and diffusion of a microbe that has been genetically engineered toward bioremediation in a contaminated body of water.

We present a first fully working first version of our platform, which uses the open-source GNU Linear Programming Kit (GLPK) for performing the dFBA calculations, and a Java-based language (Processing) for coordinating simulations and rapid visualization. COMETS can run on individual CPUs, as well as on a dedicated high performance 48-core machine. We have started applying COMETS to the analysis of several different examples, including the growth of a single species in a 2-dimensional environment and syntrophic growth of natural and engineered small microbial consortia. In ongoing work, computer simulations will be integrated with experiments, allowing us to (i) calibrate the simulation parameters towards faithful representation of microbial growth patterns, and (ii) perform pilot studies on microbial ecosystem dynamics. Specifically, we are currently employing quill-pen microprinting technology to print patterns of cells onto solid substrates in a variety of patterns and species combinations. The ability to print such combinations of cells on a surface allows us to measure several quantities used or predicted by the model, including nutrient uptake rate, cell growth rate, and metabolic byproduct diffusion in a microscale environment. These initial microscale models and experiments will be gradually scaled up to larger environments, and more complex microbial ecosystem.

### Reference

1. Mahadevan R, Edwards JS, and Doyle FJ. *Dynamic flux balance analysis of diauxic growth in Escherichia coli.* Biophys J. 2002, Sep; 83(3):1331-40.

# 264

## Modularity in Biological Systems Affects the Phenotypic Outcome of Multiple Genetic Perturbations

Hsuan-Chao Chiu[1*] (hcchiu@bu.edu), Christopher J. Marx,[2] and **Daniel Segrè**[1,3] (dsegre@bu.edu)

[1]Graduate Program in Bioinformatics, Boston University, Boston, Mass.; [2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Mass.; and [3]Departments of Biology and Biomedical Engineering, Boston University, Boston, Mass.

http://prelude.bu.edu

**Project Goals: The goal of the project is to understand how multiple genetic perturbations jointly affect microbial phenotypes (such as fitness or metabolic fluxes).**

An important application of systems biology is to provide a predictive understanding of complex biological network to help generate modified microbial strains useful for metabolic engineering applications. Both evolutionary and rational design strategies towards increasing productivity of modified strains often face the challenge of understanding how individual or combined multiple genetic perturbations affect phenotypes, such as growth rate, or the rate of byproduct secretion.

The degree of interdependency between two alleles in determining phenotypes, referred to as epistasis, is the subject of active experimental and computational research. For example, epistasis has been recently shown to be useful for identifying genes that belong to mutually dependent pathways or functional units, providing valuable information both about specific processes and about the global organization of cellular functions.

Here we propose a new theoretical framework for studying epistasis. Specifically, we study how the modular organization of a biological system affects the magnitude or sign of epistasis. Based on the assumption that different modules in a biological system can be associated with measurable phenotypes, we ask whether it is possible to infer the degree of epistasis relative to a given phenotype (e.g. fitness) based on how this phenotype depends on other phenotypes. In particular, we start by analyzing an explicit example of fitness function, in which fitness is expressed as the difference between a benefit and a cost component. Following a null assumption that mutations affect each component in a multiplicative manner, our model analytically predicts a prevalent negative epistasis distribution between pairs of mutations that are either both beneficial or both deleterious. Our results suggest that antagonistic interaction may prevail among beneficial mutations, providing diminishing returns payoff on multiple perturbations, and potentially slowing down microbial adaptation. We next provide a general expression for epistasis relative to a phenotype that has an arbitrary functional dependence on other phenotypes, under the assumption of small perturbations. This expression successfully reflects the lack-of-epistasis intuition for perturbations that act on independent modules, and provides a general "epistasis propagation law" describing how epistasis on a given high level trait can be quantified based on epistasis on lower-level traits.

Our results suggest a possible connection between the architecture of genetic interaction networks and the modular organization of biological systems, and provide general insight on the inherent benefits and limitations of multi-perturbation microbial strain design strategies.

Student Oral Presentation–Tuesday

# 265

## Algorithms for Synthetic Ecology: Microbial Cross-Feeding Induced by Environmental Transformations

Niels Klitgord[1*] (klitgord@gmail.com), and **Daniel Segrè**[1,2]

[1]Program in Bioinformatics, [2]Department of Biology and Department of Biomedical Engineering, Boston University, Boston, Mass.

http://prelude.bu.edu

**Project Goals: We use stoichiometric genome-scale models of metabolism to identify environmental conditions that induce cross-feeding interactions between different microbes.**

Interactions between microbial species can be mediated by the exchange of small molecules, secreted by one species and metabolized by another. Both one-way (commensal) and two-way (mutualistic) interactions may contribute to complex networks of interdependencies. Understanding these interactions constitutes an open challenge in microbial ecology, with important applications in metabolic engineering and environmental sustainability. In parallel to natural communities, it is possible to explore interactions in artificial microbial ecosystems, e.g. pairs of genetically engineered mutualistic strains. Here we computationally generate artificial microbial ecosystems without re-engineering the microbes themselves, but rather by predicting their growth on appropriately designed media. We use genome-scale stoichiometric models of metabolism to identify media that can sustain growth for a pair of species, but fail to do so for one or both individual species, thereby inducing putative symbiotic interactions. We first tested our approach on two previously studied mutualistic pairs, and on a pair of highly curated model organisms, showing that our algorithms successfully recapitulate known interactions, robustly predict new ones, and provide novel insight on exchanged molecules. We then applied our method to all possible pairs of seven microbial species, and found that it is always possible to identify putative media that induce commensalism or mutualism. Our analysis also suggests that symbiotic interactions may arise more readily through environmental fluc-

tuations than genetic modifications. We envision that our approach will help generate microbe-microbe interaction maps useful for understanding microbial consortia dynamics and evolution, and for exploring the full potential of natural metabolic pathways for metabolic engineering applications.

For more information: http://www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.1001002

# 266

## Integration of Flux-Balance Analysis and Pathway Databases

**Peter D. Karp*** (pkarp@ai.sri.com), and **Mario Latendresse**

SRI International

We describe new computational techniques for generating metabolic flux models from pathway databases. The Pathway Tools[1] software is a software package for creating, updating, visualizing, and analyzing Pathway/Genome Databases (PGDBs) for organisms with sequenced genomes. We have recently developed software for generating a linear programming model of a metabolic reaction network that is stored within a PGDB. Pathway Tools automatically invokes the SCIP solver on that model. The resulting optimized fluxes are then displayed on a metabolic pathway map for the PGDB by Pathway Tools, to accelerate a user's understanding of the predicted fluxes.

Benefits of this approach are that the metabolic flux model is closely coupled with an integrated genomic/metabolic knowledge base, and with other computational tools for manipulating that knowledge base. For example, users can visualize reactions, metabolites, pathways, and genome information using the rich visualization capabilities of Pathway Tools. Users can update metabolic reactions, substrates, and pathway definitions using the interactive editors within Pathway Tools, and those updates are reflected in the flux-balance model that is generated from the PGDB. In addition, metabolic model debugging tools within Pathway Tools can be applied to the metabolic flux model. Example debugging tools include tools for element balancing of metabolic reactions, and for detecting dead-end metabolites in the metabolic network.

In addition we have developed novel methods for completing a metabolic model. We have extended the gap-filling work of Maranas and colleagues to yield a multiple gap-filling approach. Using a meta-optimization procedure that is also automatically generated from a PGDB, our software will extend an incomplete metabolic model by postulating reversals of unidirectional reactions in the metabolic model, and by postulating additions of new reactions to the metabolic model from the MetaCyc database. These two approaches extend metabolic models to produce biomass compounds that they were previously unable to synthesize. Additionally, our software will gap-fill the nutrient compounds, that is, adding additional nutrient compounds that will produce biomass compounds that could not be produced. Finally, the software will identify which biomass compounds can still not be produced even after the preceding types of gap filling, thus further focusing the user's model debugging efforts. Taken together, these techniques can radically shorten the time required to develop FBA models from months to days.

**Reference**

1. Karp, P.D., Paley, S.M., Krummenacker, M., Latendresse, M., Dale, J.M., Lee, T., Kaipa, P., Gilham, F., Spaulding, A., Popescu, L., Altman, T., Paulsen, I., Keseler, I.M., and Caspi, R. (2010) "Pathway Tools version 13.0: Integrated Software for Pathway/Genome Informatics and Systems Biology," *Briefings in Bioinformatics* 11:40-79.

# 267

## Curation and Computational Design of Bioenergy-Related Metabolic Pathways

**Peter D. Karp*** (pkarp@ai.sri.com), **Mario Latendresse**, and **Ron Caspi**

SRI International

**Project Goals: Goal 1: Enhance the MetaCyc metabolic pathway database to contain bioenergy-related enzymes and pathways. Goal 2: Develop computational tools for engineering metabolic pathways that satisfy specified design goals, in particular for bioenergy-related pathways.**

Pathway Tools[1] is a systems-biology software package written by SRI International (SRI) that produces Pathway/Genome Databases (PGDBs) for organisms with sequenced genomes. Pathway Tools also provides a wide range of capabilities for analyzing the predicted metabolic networks, and analyzing user-generated omics data. More than 1,500 academic, industrial, and government groups use Pathway Tools. An integral part of the Pathway Tools software is MetaCyc[2], a large, multiorganism database of metabolic pathways and enzymes that is manually curated by SRI and its collaborators. Our project has two goals:

1. Enhance MetaCyc to contain bioenergy-related metabolic enzymes and pathways.
2. Develop computational tools for engineering metabolic pathways that satisfy specified design goals, in particular for bioenergy-related pathways.

We are significantly expanding the coverage of bioenergy-related metabolic information in MetaCyc. Version 14.6 of MetaCyc contains 1,600 pathways and 9,000 bioreactions from 2,100 organisms. The information in MetaCyc has been curated from more than 26,000 publications. We are curating additional bioenergy-related metabolic pathways into MetaCyc from the biomedical literature. By adding these additional pathways to MetaCyc, the pathway prediction component of the Pathway Tools software will be able to recognize these pathways in newly sequenced genomes. We plan to use Pathway Tools to generate organism-specific

Pathway/Genome Databases (PGDBs) for all energy-relevant organisms sequenced at JGI. SRI will make these databases freely available to the public via its BioCyc website

Our second goal is to develop an efficient computational tool for the engineering of metabolic pathways. The tool will satisfy design goals specified by the user and will offer two operational modes: a fast, approximate mode that will quickly find a near-optimal pathway and an exact mode that will find the optimal pathway. Design goals will include starting and ending compounds for the pathway, and will enable the specification of constraints such as preferred or disallowed intermediates. The tool will utilize the large collection of enzymes in the MetaCyc database as a reference, and its suggestions will be validated with the bioenergy-related pathways we are curating into MetaCyc. The pathways generated by this tool will be ranked and clustered according to several optimality criteria.

**References**


1. Karp, P.D., Paley, S.M., Krummenacker, M., Latendresse, M., Dale, J.M., Lee, T., Kaipa, P., Gilham, F., Spaulding, A., Popescu, L., Altman, T., Paulsen, I., Keseler, I.M., and Caspi, R. (2010) "Pathway Tools version 13.0: Integrated Software for Pathway/Genome Informatics and Systems Biology," *Briefings in Bioinformatics* 11:40-79.
2. Caspi, R., Altman, T., Dale, J.M., Dreher, K., Fulcher, C.A., Gilham, F., Kaipa, P., Karthikeyan, A.S., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Paley, S., Popescu, L., Pujar, A., Shearer, A.G., Zhang, P., and Karp, P.D., (2010) "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases," *Nuc. Acids Res*. 36:D623-31.


# 268

## Gene Ontology for Microbial Processes Useful for Bioenergy Production


Trudy Torto-Alalibo[1]* (trudy@vbi.vt.edu), Endang Purwantini,[1] Biswarup Mukhopadhyay,[1] T.M. Murali,[2] Brett M. Tyler,[1] and **Joao C. Setubal**[1]

[1]Virginia Bioinformatics Institute, Blacksburg; and [2]Virginia Polytechnic Institute and State University, Blacksburg


**Project Goals: In collaboration with the community of microbiologists interested in energy-related processes and with the Gene Ontology (GO) consortium, develop a comprehensive set of Gene Ontology terms that describe biological processes relevant to energy-related functions. Annotate microbial genomes relevant to bioenergy-production with appropriate GO terms Develop a database and web interface for storing and displaying manual annotations.**

The MENGO project (http://mengo.vbi.vt.edu/) is a community-oriented multi-institutional collaborative effort that aims to develop new Gene Ontology (GO) terms to describe microbial processes of interest to bioenergy. Such terms will aid in the comprehensive annotation of gene products from diverse energy-related microbial genomes. The Gene Ontology consortium was formed in 1998 to create universal descriptors, which can be used to describe functionally similar gene products and their attributes across all organisms. MENGO, an interest group of the GO consortium, solicits help from the bioenergy community in developing GO terms relevant for bioenergy-production related processes such as biomass deconstruction, solventogenic fermentation, $H_2$ production, methanogenesis, and synthesis of hydrocarbons. Currently, based on community input, five potential working groups have been formed. These working groups will work on GO term development in focal areas that include biomass deconstruction, central metabolism and protein structure-function relationships. A few general concepts associated with the selected focal areas have been developed. These and related GO terms will be presented. The MENGO interest group will host a workshop right after the DOE contractor-grantee meeting from April 13-14 at the same venue. This workshop will introduce participants to the Gene Ontology. Additionally, we will have an open forum to seek input from participants on general concepts relevant to bioenergy-production related processes, which will subsequently inform MENGO term development. Funding for the MENGO project is provided by the Department of Energy as part of the Systems Biology Knowledgebase program.

# Communication and Ethical, Legal, and Societal Issues

## 269

### "Soft Law" Approaches to Nanotechnology Oversight

**Gary Marchant**[1]* (gary.marchant@asu.edu), Kenneth Abbott,[1] Douglas Sylvester,[1] and Elizabeth Corley[2]

[1]Center for Law, Science and Innovation, Sandra Day O'Connor College of Law, Arizona State University; and [2]School of Public Affairs, Arizona State University

**Project Goals: We seek to identify, study, evaluate and recommend innovative soft law governance approaches that may apply to nanotechnology and could be scaled up to a transnational level. Specifically, we seek to: 1. Identify and classify the wide variety of public, private and collaborative soft law governance initiatives that have been implemented or proposed for nanotechnology, as well as similar models for other emerging technologies and may be adaptable to nanotechnology. 2. Evaluate the outcomes, design choices, strengths and weaknesses of the mechanisms studied. Based on the results of our analysis and an expert workshop, we will select four of the most promising oversight models for further, in-depth analysis. 3. Using a series of evaluative questions, a survey of relevant stakeholders and a second expert workshop, conduct an in-depth analysis of the four most promising soft law governance models for nanotechnology.**

Regulatory agencies are in the early stages of developing regulatory frameworks for nanotechnology, but it will be many years before comprehensive federal regulatory oversight is in place for nanotechnology. In the mean time, a number of innovative, voluntary or partnership-based programs, which we refer to as "soft law" initiatives because of their non-binding nature, have emerged to try to fill the oversight gap for nanotechnology. In this presentation, we analyze and compare six existing soft law initiatives for nanotechnology: (1) the Dupont-EDF NanoRisk Framework; (2) the Responsible NanoCode; (3) The NanoSafety Consortium for Carbon (NCC); (4) the EU Code of Conduct for Responsible Nanosciences and Nanotechnologies; (5) the Nanomaterial Product Stewardship Program; and (6) CENARIOS nanotechnology certification program. We assess the relative strengths and weaknesses of these various soft law programs, identify different positions on the role of these programs vis-à-vis traditional regulatory programs, and propose a typology for these soft law programs based on the participation and parameters of the programs.

## 270

### Biological and Environmental Research Information System: A Multifaceted Approach to DOE Systems Research Communication

Anita J. Alton, Shirley H. Andrews, Jennifer L. Bownas, Christopher E. Caldwell, Kris S. Christen, Holly L. Haun, 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.

**Project Goals: Provide programmatic information via printed and online materials to help build the critical multidisciplinary community needed to advance systems research for DOE energy and environmental missions. The Biological and Environmental Research Information System group works with program managers and the scientific community to help develop and communicate key scientific and technical concepts for research community and public discourse. Ideas are welcome to extend and improve communications and program integration and thus represent BER's research more comprehensively.**

Concerted communication is key to progress in cutting-edge science and public accountability. Our goals focus on three objectives: (1) facilitate science planning, research, and communication; (2) inform a broader audience about DOE research projects, progress, and significance to science and society; and (3) respond to outreach and information exchange needs of related DOE projects.

For the past 22 years, our group, the Biological and Environmental Research Information System, has focused on presenting all facets of genomics research for the Department of Energy's (DOE) Office of Science. The materials we produce have helped ensure that scientists can participate in and reap the bounty of the genome revolution, that new generations of students can be trained in genomics and systems biology, and that the public can make informed decisions regarding genetics issues.

In 2009, our scope was extended to include all 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. These programs explore 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. We work with DOE staff and the research community to produce and disseminate information in various formats: technical reports, roadmaps, websites, brochures, databases, technical compilations, presentations, exhibits for scientific meetings, text, graphics, and posters. We staff the BER and Genomic Science exhibit at more than 10 scientific meetings each year. We also work with DOE grantees and members of the extended DOE BER community, especially with the outreach efforts of the Bioenergy Research Centers, the Joint Genome Institute, the Environmental Molecular Sciences Laboratory, and the Atmospheric Radiation Measurement Climate Research Facility to help increase their reach and impact.

**For BER's Biological Systems Science Division (BSSD), our recent Genomic Science program accomplishments include**

**Research plans and reports produced with the scientific community:**

- *Systems Biology Knowledgebase Implementation Plan* (September 2010), plus a series of individual workshop reports (January –June 2010)

**BER BSSD booklets and brochures include:**

- *Advanced Technologies for Biology: Overview of Structural Biology Infrastructure* (March 2011)
- *Overview of Projects Underpinning Knowledgebase Development* (revision, February 2011)
- *Biological Systems Research on the Role of Microbial Communities in Carbon Cycling: Summary of Projects of Awarded in Summer 2010* (October 2010)
- *Bioenergy Research Centers: An Overview of the Science* (revision, August 2010)
- *Plant Feedstock Genomics for Bioenergy: Joint Awards* (July 2010)
- *BSSD Overview* (May 2010)
- *Bioenergy Research Centers: Education and Outreach* (March 2010)
- *Genomic Science Program Overview* (February 2010)

Other recently produced BER BSSD materials include the *Genomics for Energy and Environmental Science* placemat (January 2011), abstracts book for the *DOE Genomic Science Awardee Workshop VIII and Knowledgebase Workshop* (February 2010), and an exhibit created for the DOE Office of Science Genomic Science Program (October 2009). BER BSSD works in progress include this abstracts book.

We also continuously update and enhance numerous websites including the Genomic Science website (*genomicscience.energy.gov*) and public image gallery (*genomics.energy.gov/gallery/*). A major redesign of the Genomic Science website was completed in November 2010. The updated site streamlines content and design, while improving navigation and increasing functionality and accessibility. New sections gives greater access to information about DOE user facilities and the Genomic Science Knowledgebase.

**BER-wide projects completed include the creation and dissemination of:**

- Searchable public BER Research Highlights database (*public.ornl.gov/hgmis/bernews/*)
- *DOE BER Overview* (revisions: March 2011, August 2010)
- *BER Grand Challenges: A Long-Term Vision* (December 2010)
- DOE BER poster (August 2010)
- DOE BER exhibit (July 2010)

**For BER's Climate and Environmental Sciences Division (CESD), our recent accomplishments include**

**Research plans and reports produced with the scientific community:**

- *Climate Research Roadmap* (September 2010)
- *Complex Systems Science for Subsurface Fate and Transport* (March 2010)

**BER CESD placemat and brochures include:**

- *Energy-Climate Nexus* placemat (January 2011)
- *Terrestrial Ecosystem Science Overview (May 2010)*
- *CESD Overview* (March 2010)
- *DOE Environmental Molecular Sciences Laboratory (EMSL) Overview* (March 2010)
- *DOE Atmospheric Radiation Measurement (ARM) Climate Research Facility Overview* (February 2010)
- *Climate and Earth System Modeling Overview* (in progress)

Office of Biological and Environmental Research (BER) http://science.doe.gov/ober/ober_top.html

BER documents http://www.berscience.org/

The Biological and Environmental Research Information System is supported by the U.S. Department of Energy Office of Biological and Environmental Research in the DOE Office of Science.

# 270A‡

submitted post-press

## Societal Implications of Nanoscale and Bioenergy Science and Technology Undertaken at U.S. Department of Energy Research Centers

**Co-PIs: Amy K. Wolfe, David J. Bjornstad*** (bjornstaddj@ornl.gov); Co-authors: W. Christopher Lenhardt, Barry Shumpert, and Mitch Doktycz

Oak Ridge National Laboratory

**Project Goals: The SFA seeks to determine the degree to which its findings can be applied to similar choices for**

**other technologies, uses, and management approaches. In meeting these goals, the SFA will establish a unique ELSI resource for SC, and for the broader suite of communities interested in societal implications associated with emerging S&T.**

This poster presents an overview of the Ethical, Legal, and Social Issues (ELSI) Scientific Focus Area (SFA) at Oak Ridge National Laboratory. This SFA's two long-term goals are to:

- Create a new base of knowledge that identifies the kinds of societal issues that arise at key junctures over time, as S&T moves from research and development (R&D) into use, and
- Analyze the potential implications of alternative choices upstream (for R&D) and downstream (for use).

The SFA seeks to determine the degree to which its findings can be applied to similar choices for other technologies, uses, and management approaches. In meeting these goals, the SFA will establish a unique ELSI resource for SC, and for the broader suite of communities interested in societal implications associated with emerging S&T.

In this SFA, "ELSI issues" refers to the set of choices, impacts, and implications that determine the manner in which S&T and their products are developed and integrated into society to achieve SC goals. "Societal considerations" are those issues brought to bear when making choices. "Junctures" are key decision points along the continuum of R&D through initial use. SFA research will identify how parties involved in making choices at these junctures tend to weigh multiple, potentially competing and conflicting considerations, in a manner that links to specific technologies, applications, and institutional processes. Thus, SFA research is structured to identify and analyze the:

- implications of ***S&T process and product-related*** societal considerations associated with specific choices that arise in moving from research to the eventual use of DOE S&T;
- implications of specific ***institutional and organizational management*** choices associated with moving from research to the eventual use of DOE S&T, and societal issues associated with those choices;
- extent to which our findings are ***generalizable*** across technologies, contexts, applications, and R&D lifecycle phases.

ORNL's ELSI SFA is designed to be gap-filling. ELSI-related studies tend not to have been organized to answer cross-cutting questions, although DOE would benefit from a stronger foundation upon which to anticipate issues that arise over specific technologies. Moreover, ELSI and related studies generally have not focused on SC-funded S&T, SC's mission, or on the extent to which their findings are applicable to DOE. These gaps increase SC's challenges in making sets of choices that can affect its ability to achieve its mission. This SFA seeks to fill these gaps by combining an explicit focus on SC S&T with targeted research aims and questions.

Focusing on SC's Nanoscale Science Research Center (NSRC) and Bioenergy Research Center (BRC) located at ORNL, the SFA anchors its investigations to specified technologies, potential applications, and institutional practices. Alternative methods are used to identify key choices at junctures along the pathway from laboratory to use, parties who make those choices, the considerations that influence their decisions, and the implications of alternative choices upstream for R&D and downstream for early and mature use, disposal, and decommissioning. All analyses address a principal, core objective: to ***identify and analyze issues of societal concern***, as they may be reflected in key junctures along the R&D-through-initial use trajectory.

To impose structure and bounds on our inquiries, we ask how three sets of attributes influence key issues (e.g., energy-environment-society tradeoffs) that are likely to be at the heart of tough choices concerning nanoscale and bioenergy S&T in different contexts and at different points along R&D-to-use lifecycle phases. These three sets of attributes are: (a) technology products and processes; (b) their potential applications; and (c) institutional and organizational management of research centers—here focusing on intellectual property and information management. More specifically, we ask:

- Holding constant a single category of technology, how do societal issues and tradeoffs vary when that technology is used in different applications?
- Holding constant a single application—here defined in terms of a specific societal goal, how do societal issues and tradeoffs vary when that goal is pursued with different technological options?
- Holding constant guidelines for intellectual property and information management, how do research centers' implementation practices influence the nature of their S&T R&D and how does information flow from inside to outside of the research centers?

By disaggregating and delving into different layers of attributes that influence societal considerations associated with key decision points, this Science Plan will develop new understandings about the societal implications of emerging technologies.

# 270B‡

submitted post-press

## Nanoscale Science Ethical, Legal, and Social Issues (ELSI) Analyses: Issues, Nano-Attributes, and Potential Applications

**Co-PIs: Amy K. Wolfe**[1]* (wolfeak@ornl.gov), **David J. Bjornstad**,[1]* Co-authors: W. Christopher Lenhardt,[1] Barry Shumpert,[1] Mitch Doktycz,[1] and Stephanie Wang[2]

[1]Oak Ridge National Laboratory and [2]Grinnell College

**Project Goals: One avenue for inquiry within the Ethical, Legal, and Social Issues (ELSI) Scientific Focus Area (SFA) at Oak Ridge National Laboratory seeks to identify the kinds of societal issues that arise at key junctures over time, as S&T moves from research and development (R&D) into use. More specifically, the SFA focuses on the set of choices, impacts, and implications that determine the manner in which S&T and their products are developed and integrated into society to achieve DOE Office of Science (SC) goals.**

This poster describes initial analyses that begin to disaggregate factors influencing decision points and their associated societal issues along the pathway from R&D to use within the realm of DOE Nanoscale Science Research Centers (NSRCs). As one starting point, we look to existing ELSI scholarship to help us categorize key *choices*, the *issues* parties typically involved in making these choices raise (or do not raise), and how involved parties *weigh or trade-off* among the multiple salient issues in determining how to proceed.

A critical aspect of this ELSI SFA is it's strong linkage to DOE-SC S&T. History repeatedly has shown that societal responses to the same technology vary, even in seemingly similar contexts. The overall ELSI SFA is structured to help sort out the extent to which S&T-related attributes versus application-related attributes (e.g., energy versus environmental applications) influence choices, issues, and tradeoffs. Therefore, we also are examining nano-related ELSI literature with regard to its linkage with particular categories of:

1. nanoscale science and technology—the extent to which the kind of nanomaterial or process matters; and
2. sphere of application—the extent to which different uses of nanomaterials or processes matter, where we focus on energy versus environmental applications.

Our examination of approximately 85 nano-related ELSI publications indicates that the predominant issues on which ELSI scholars have focused are the role of ELSI, perceptions, governance, and equity. We found that the vast majority of publications do not specify kinds of nanomaterials or processes, so that it is unclear whether these publications intended to be general across all nanomaterials and processes. Likewise, most articles refer in general to the potential applications of nanomaterials and processes rather than to specific spheres of application. Despite these gaps, the literature does help to identify a variety of considerations that may be important to understanding choices, issues, and tradeoffs along the pathway from R&D to use. This poster presents our interim findings from this literature-based analysis and implications for our goal of disaggregating the evolving societal considerations associated with DOE's emerging S&T.

# Appendix 1: Participants

As of March 29, 2011

John Aach
Harvard Medical School, Department of Genetics
aach@receptor.med.harvard.edu

Mike Adams
University of Georgia
adams@bmb.uga.edu

Paul Adams
Lawrence Berkeley National Laboratory
PDAdams@lbl.gov

Santiago Akle
Stanford University
saunders@stanford.edu

Muktak Aklujkar
University of Massachusetts, Amherst
muktak@microbio.umass.edu

Magdy Alabady
University of Illinois
msalabad@uiuc.edu

Steven Allison
University of California, Irvine
allisons@uci.edu

Gorur Amita
Lawrence Berkeley National Laboratory
Agorur@lbl.gov

Gordon Anderson
Biological Sciences and Mass Spectrometry
gordo@pnl.gov

Adam Arkin
Lawrence Berkeley National Laboratory
aparkin@lbl.gov

Terry Astrid
Lawrence Berkeley National Laboratory
AYTerry@lbl.gov

Manfred Auer
DOE Joint BioEnergy Institute (JBEI)
Lawrence Berkeley National Laboratory
mauer@lbl.gov

Sara Baldwin
Boston University
baldwin2@bu.edu

Nitin Baliga
Institute for Systems Biology
nbaliga@systemsbiology.org

Anindita Bandyopadhyay
Washington University in Saint Louis
anindita@wustl.edu

Goutami Banerjee
DOE Great Lakes Bioenergy Research Center (GLBRC)
gbanerjee11@gmail.com

Jillian Banfield
University of California, Berkeley
jbanfield@berkeley.edu

Richard Baran
Lawrence Berkeley National Laboratory
RBaran@lbl.gov

Jessica Barb
University of Georgia
jgbarb@gmail.com

Christopher Bare
Institute for Systems Biology
nbaliga@systemsbiology.org

Paul Bayer
U.S. Department of Energy
paul.bayer@science.doe.gov

Larry Beach
US Agency for International Development
lbeach@usaid.gov

Qasim Beg
Boston University
qasimbeg@bu.edu

Alex Beliaev
Pacific Northwest National Laboratory
jim.fredrickson@pnl.gov

Harry Beller
DOE Joint BioEnergy Institute (JBEI)
Lawrence Berkeley National Laboratory
HRBeller@lbl.gov

Judith Bender
Brown University
judith_bender@brown.edu

Jeff Bennetzen
University of Georgia
maize@uga.edu

William Benton
DOE Great Lakes Bioenergy Research Center (GLBRC)
dbenton@glbrc.wisc.edu

Rajiv Bharadwaj
DOE Joint BioEnergy Institute (JBEI)
AKSingh@lbl.gov

Mark Biggin
Lawrence Berkeley National Laboratory
MDBiggin@lbl.gov

Ajaya Biswal
University of Georgia
biswalajaya@ccrc.uga.edu

David Bjornstad
Oak Ridge National Laboratory
bjornstaddj@ornl.gov

Chris Blackwood
Kent State University
cblackwo@kent.edu

Paolo Boccazzi
Pharyx, Inc.
boccazzi@pharyx.com

Harvey Bolton
Pacific Northwest National Laboratory
harvey.bolton@pnl.gov

Yannick Bomble
DOE BioEnergy Science Center
gilnap@ornl.gov

Gregory Bonito
Duke University
fungi@duke.edu

Adam Book
DOE Great Lakes Bioenergy Research Center (GLBRC)
ajbook@wisc.edu

Franklin Boteler
USDA-NIFA
fboteler@nifa.usda.gov

Nicholas Bouskill
Lawrence Berkeley National Laboratory
njbouskill@lbl.gov

Ben Bowen
Lawrence Berkeley National Laboratory
bpbowen@lbl.gov

Jennifer Bownas
Oak Ridge National Laboratory
bownasjl@ornl.gov

James Brainard
National Renewable Energy Laboratory
james.brainard@nrel.gov

Eoin Brodie
Lawrence Berkeley National Laboratory
elbrodie@lbl.gov

Patrick Brown
University of Illinois
pjb34@illinois.edu

Steven Brown
DOE BioEnergy Science Center
gilnap@ornl.gov

Edward C Brummer
The Samuel Noble Roberts Foundation
ecbrummer@noble.org

Amy Brunner
Virginia Tech
abrunner@vt.edu

Donald Bryant
Pennsylvania State University
Montana State University
dab14@psu.edu

Marcia Buanafina
Pennsylvania State University
mmb26@psu.edu

Edward Buckler
USDA-ARS, Cornell University
esb33@cornell.edu

Daniel Buckley
Cornell University
dbuckley@cornell.edu

Robin Buell
Michigan State University
buell@msu.edu

Victor Busov
Michigan Tech
vbusov@mtu.edu

Gareth Butland
Lawrence Berkeley National Laboratory
GPButland@lbl.gov

Stephen Callister
MBARI/UCSC
azworden@mbari.org

Allison Campbell
Pacific Northwest National Laboratory
allison.campbell@pnl.gov

William Cannon
Pacific Northwest National Laboratory
william.cannon@pnl.gov

Kevin Carpenter
Lawrence Livermore National Laboratory
carpenter37@llnl.gov

Nicholas Carpita
Purdue University
carpita@purdue.edu

Julie Carruthers
DOE Office of Science
julie.carruthers@science.doe.gov

Michael Casler
USDA-ARS
mdcasler@wisc.edu

Cynthia Cass
Illinois State University
ccass@ilstu.edu

Javier Arturo Ceja Navarro
Lawrence Berkeley National Laboratory
jcnavarro@lbl.gov

Romy Chakraborty
Lawrence Berkeley National Laboratory
RChakraborty@lbl.gov

John-Marc Chandonia
Lawrence Berkeley National Laboratory
jmchandonia@lbl.gov

Chieh Chang
DOE Joint BioEnergy Institute (JBEI)
AKSingh@lbl.gov

Liang Chen
Lawrence Berkeley National Laboratory
lchen2@lbl.gov

Xinlu Chen
University of Tennessee
xchen24@utk.edu

Gang Cheng
DOE Joint BioEnergy Institute (JBEI)
Sandia National Laboratories
mskent@lbl.gov

Swapnil Chhabra
Lawrence Berkeley National Laboratory
srchhabra@lbl.gov

Kevin Childs
Michigan State University
kchilds@plantbiology.msu.edu

Hsuan-CHao Chiu
Boston University
dsegre@bu.edu

Dylan Chivian
Lawrence Berkeley National Laboratory
DCChivian@lbl.gov

Karuna Chourey
Oak Ridge National Laboratory
choureyk@ornl.gov

Kris Christen
Oak Ridge National Laboratory
christenks@ornl.gov

Shishir Chundawat
DOE Great Lakes Bioenergy Research Center (GLBRC)
chundawa@msu.edu

George Church
Harvard Medical School
ymasar@genetics.med.harvard.edu

Robert Clubb
UCLA-DOE Institute of Genomics and Proteomics
rclubb@mbi.ucla.edu

Dean Cole
U.S. Department of Energy
dean.cole@science.doe.gov

James Cole
Michigan State University
colej@msu.edu

Frank Collart
Argonne National Laboratory
fcollart@anl.gov

Denise Costich
USDA-ARS, Cornell University
dc58@cornell.edu

Stephen Cramer
University of California, Davis
spcramer@ucdavis.edu

Michael Crowley
DOE BioEnergy Science Center
gilnap@ornl.gov

Leland Cseke
University of Alabama in Huntsville
csekel@uah.edu

Alan Darvill
DOE BioEnergy Science Center
adarvill@ccrc.uga.edu

Maude David
Lawrence Berkeley National Laboratory
MMDavid@lbl.gov

Mark Davis
National Renewable Energy Laboratory
Mark.Davis@nrel.gov

Brian Davison
Oak Ridge National Laboratory
davisonbh@ornl.gov

Natalia de Leon
Great Lakes Bioenergy Research Center
University of Wisconsin
ndeleongatti@wisc.edu

Paramvir Dehal
Lawrence Berkeley National Laboratory
pdadams@lbl.gov

Jaclyn DeMartini
University of California, Riverside
jdema001@ucr.edu

Kai Deng
DOE Joint BioEnergy Institute (JBEI)
kdeng@lbl.gov

Ye Deng
University of Oklahoma
dengye@ou.edu

Adam Deutschbauer
Lawrence Berkeley National Laboratory
AMDeutschbauer@lbl.gov

Patrik D'haeseleer
DOE Joint BioEnergy Institute (JBEI)
Lawrence Livermore National Laboratory
patrikd@gmail.com

Shi-You Ding
National Renewable Energy Laboratory
Shi.you.Ding@nrel.gov

Rick Dixon
Samuel Roberts Noble Foundation
radixon@noble.org

Mitchel Doktycz
Oak Ridge National Laboratory
gaudreaucl@ornl.gov

Timothy Donohue
DOE Great Lakes Bioenergy Research Center (GLBRC)
tdonohue@bact.wisc.edu

Dan Drell
U.S. Department of Energy
daniel.drell@science.doe.gov

Inna Dubchak
Lawrence Berkeley National Laboratory
ildubchak@lbl.gov

David Duncan
DOE Great Lakes Bioenergy Research Center (GLBRC)
University of Wisconsin-Madison
dsduncan@wisc.edu

John Dunn
Brookhaven National Laboratory
jdunn@bnl.gov

Ismail Dweikat
University of Nebraska
idweikat2@unl.edu

Joseph Ecker
The Salk Institute
ecker@salk.edu

Mike Edgerton
Monsanto Company
mmedge@monsanto.com

Dwayne Elias
Oak Ridge National Laboratory
eliasda@ornl.gov

Barbara Evans
Oak Ridge National Laboratory
evansb@ornl.gov

Gang Fang
Yale University
gang.fang@yale.edu

Stephen Federowicz
University of California, San Diego
sfederow@ucsd.edu

John Ferrell
EERE/Biomass
john.ferrell@ee.doe.gov

Matthew Fields
Montana State University
matthew.fields@erc.montana.edu

Mary Firestone
University of California, Berkeley
mkfstone@berkeley.edu

Zoë Fisher
Los Alamos National Laboratory
zfisher@lanl.gov

Ronan Fleming
University of Iceland
ronan.mt.fleming@gmail.com

Karen Foley
Institute for Systems Biology
kfoley@systemsbiology.org

Ellen Ford
JBEI
eford@lbl.gov

Virginia Fowler
Oak Ridge National Laboratory
fowlervs@ornl.gov

Jim Fredrickson
Pacific Northwest National Laboratory
jim.fredrickson@pnl.gov

Andrew Frock
North Carolina State University
adfrock@ncsu.edu

Andrei Gabrielian
NIAID NIH
gabr@niaid.nih.gov

Xiadi Gao
University of California, Riverside
xiadi.gao@email.ucr.edu

Hector Garcia Martin
Lawrence Berkeley National Laboratory
hectorg@gmail.com

George Garrity
Michgan State University and
NamesforLife, LLC
garrity@msu.edu

Siuzdak Gary
Lawrence Berkeley National Laboratory
siuzdak@gmail.com

Simon George
Lawrence Berkeley National Laboratory
University of California, Davis
SJGeorge@lbl.gov

Maria Ghirardi
National Renewable Energy Laboratory
maria.ghirardi@nrel.gov

Richard Giannone
Oak Ridge National Laboratory
giannonerj@ornl.gov

Paul Gilna
Oak Ridge National Laboratory
gilnap@ornl.gov

Carol Giometti
Argonne National Laboratory
csgiometti@anl.gov

Dennis Glass
University of Tennessee
Oak Ridge National Laboratory
dennisglass84@gmail.com

Jose Gonzalez
South Dakota State University
jose.gonzalez@sdstate.edu

Andrey Gorin
DOE BioEnergy Science Center
gilnap@ornl.gov

Joe Graber
U.S. Department of Energy
joseph.graber@science.doe.gov

Pamela Green
University of Delaware
green@dbi.udel.edu

Susan Gregurick
U.S. Department of Energy
susan.gregurick@science.doe.gov

Nicholas Guido
Harvard Medical School
guido@genetics.med.harvard.edu

Robert Gunsalus
University of California, Los Angeles
robg@microbio.ucla.edu

Adam Guss
Oak Ridge National Laboratory
gussam@ornl.gov

Murray Hackett
University of Washington
mhackett@u.washington.edu

Masood Hadi
DOE Joint BioEnergy Institute (JBEI)
Sandia National Laboratories
mzhadi@sandia.gov

Bong-Gyoon Han
Lawrence Berkeley National Laboratory
bghan2@gmail.com

Zhao Hao
Lawrence Berkeley National Laboratory
zhao@lbl.gov

William Harcombe
Harvard University
wharcombe@oeb.harvard.edu

Scott Harding
University of Georgia
sharding@uga.edu

Maria Harrison
Boyce Thompson Institute
mjh78@cornell.edu

Caroline Harwood
University of Washington
csh5@uw.edu

Holly Haun
Oak Ridge National Laboratory
han@ornl.gov

Loren Hauser
Oak Ridge National Laboratory
hauserlj@ornl.gov

Samuel Hazen
University of Massachusetts, Amherst
hazen@bio.umass.edu

Terry Hazen
DOE Joint BioEnergy Institute (JBEI)
tchazen@lbl.gov

Zhili He
University of Oklahoma, Institute for Environmental Genomics
zhili.he@ou.edu

Joshua Heazlewood
Lawrence Berkeley National Laboratory
JLHeazlewood@lbl.gov

Eric Hegg
Michigan State University
EricHegg@msu.edu

Richard Heins
DOE Joint BioEnergy Institute (JBEI)
raheins@lbl.gov

Christopher Henry
Argonne National Laboratory
chrisshenry@gmail.com

Robert Hettich
Oak Ridge National Laboratory
hettichrl@ornl.gov

Nathan Hillson
DOE Joint BioEnergy Institute (JBEI)
Lawrence Berkeley National Laboratory
njhillson@lbl.gov

Yasawantha Hiripitiyage
Kent State University
devindahiri@msn.com

Roland Hirsch
Biological Systems Sciences Division, Office of Biological & Environmental Research, Department of Energy
roland.hirsch@science.doe.gov

Hoi-Ying Holman
Lawrence Berkeley National Laboratory
hyholman@lbl.gov

John Houghton
U.S. Department of Energy
john.houghton@science.doe.gov

Shifang Hsu
Kent State University
shsu2@kent.edu

Matthew Hudson
University of Illinois
mhudson@illinois.edu

Greg Hura
Lawrence Berkeley National Laboratory
glhura@gmail.com

Greg Hurst
Oak Ridge National Laboratory
hurstgb@ornl.gov

Saheed Imam
University of Wisconsin-Madison
srimam@wisc.edu

William Inskeep
Montana State University
binskeep@montana.edu

Christer Jansson
Lawrence Berkeley National Laboratory
CGJansson@lbl.gov

Bing Jap
Lawrence Berkeley National Laboratory
bkjap@lbl.gov

Yongqin Jiao
Lawrence Livermore National Laboratory
jiao1@llnl.gov

Zhou Jizhong
University of Oklahoma, Institute for Environmental Genomics
jzhou@ou.edu

Andrzej Joachimiak
Argonne National Laboratory
andrzejj@anl.gov

Ari Jumpponen
Kansas State University
ari@ksu.edu

Nicholas Justice
University of California, Berkeley
njustice@berkeley.edu

Shawn Kaeppler
University of Wisconsin
smkaeppl@wisc.edu

Justin Kaffenberger
University of Minnesota
kaffe002@umn.edu

Edward Kaleikau
USDA National Institute of Food and Agriculture
ekaleikau@nifa.usda.gov

Udaya Kalluri
Oak Ridge National Laboratory
kalluriudayc@ornl.gov

Marina Kalyuzhnaya
University of washington
mkalyuzh@uw.edu

Ulas Karaoz
Lawrence Berkeley National Laboratory
UKaraoz@lbl.gov

Peter Karp
SRI International
pkarp@ai.sri.com

Tatiana Karpinets
Oak Ridge National Laboratory
k2n@ornl.gov

Abhijit Karve
Oak Ridge National Laboratory
karveaa@ornl.gov

Irina Kataeva
University of Georgia
kataeva@uga.edu

Arthur Katz
U.S. Department of Energy
arthur.katz@science.doe.gov

Alexey Kazakov
Lawrence Berkeley National Laboratory
aekazakov@lbl.gov

Jay Keasling
DOE Joint BioEnergy Institute (JBEI)
Lawrence Berkeley National Laboratory
jdkeasling@lbl.gov

David Keating
DOE Great Lakes Bioenergy Research Center (GLBRC)
dkeating@glbrc.wisc.edu

Kenneth Keegstra
DOE Great Lakes Bioenergy Research Center (GLBRC)
Michigan State University
keegstra@msu.edu

Kimberly Keller
University of Missouri
kellerkl@missouri.edu

Martin Keller
Oak Ridge National Laboratory, Energy and Environmental Sciences Directorate
kellerm@ornl.gov

Robert Kelly
North Carolina State University
rmkelly@eos.ncsu.edu

Zvi Kelman
University of Maryland
zkelman@umd.edu

Changsoo Kim
University of Georgia
changkim@uga.edu

Matias Kirst
University of Florida
mkirst@ufl.edu

Niels Klitgord
Boston University, Segre Lab
niels@bu.edu

Guruprasad Kora
Oak Ridge National Laboratory
koragh@ornl.gov

Helen Kreuzer
Pacific Northwest National Laboratory
helen.kreuzer@pnl.gov

Akhil Kumar
Pennsylvania State University
azk172@psu.edu

Cheryl Kuske
Los Alamos National Laboratory
kuske@lanl.gov

Shing Kwok
USDA-NIFA
skwok@nifa.usda.gov

Jessy Labbé
Oak Ridge National Laboratory
labbejj@ornl.gov

Catherine Lacayo
Lawrence Livermore National Laboratory
lacayo2@llnl.gov

Philip Laible
Argonne National Laboratory
laible@anl.gov

William Lancaster
University of Georgia, Athens
alancaster@bmb.uga.edu

Paul Langan
Los Alamos National Laboratory
langan_paul@lanl.gov

Carolyn Larabell
Lawrence Berkeley National Laboratory
larabell@lbl.gov

Peter Larsen
Argonne National Laboratory
plarsen@anl.gov

Haythem Latif
University of California, San Diego
halatif@ucsd.edu

Daniel Lauffer
DOE Great Lakes Bioenergy Research Center (GLBRC)
dlauffer@glbrc.wisc.edu

Mark Le Gros
Lawrence Berkeley National Laboratory
MALEGROS@LBL.GOV

Jan Leach
Colorado State University
Jan.Leach@colostate.edu

Harry Lee
Pharyx, Inc.
harrylee@pharyx.com

Taek Soon Lee
Lawrence Berkeley National Laboratory
tslee@lbl.gov

Laura Leff
Kent State University
lleff@kent.edu

John Leigh
University of Washington
leighj@u.washington.edu

Rebecca Lennen
University of Wisconsin-Madison
lennen@wisc.edu

Joshua Lerman
University of California, San Diego
jlerman44@gmail.com

Dan Li
UCLA-DOE Institute of Genomics and Proteomics
danli@mbi.ucla.edu

Ting Li
ORISE
lit1@ornl.gov

Jim Liao
University of California, Los Angeles
liao@seas.ucla.edu

Ann Lichens-Park
USDA National Institute of Food and Agriculture
apark@nifa.usda.gov

Liang-Shiou Lin
USDA National Institute of Food and Agriculture
llin@nifa.usda.gov

Benjamin Lindner
Oak Ridge National Laboratory
ben@benlabs.net

Mary Lipton
Pacific Northwest National Laboratory
mary.lipton@pnl.gov

Andrew Loder
North Carolina State University
ajloder@ncsu.edu

Sharon Long
Stanford University
sharon.long@stanford.edu

Dominique Loque
DOE Joint BioEnergy Institute (JBEI)
Lawrence Berkeley National Laboratory
dloque@lbl.gov

Derek Lovley
University of Massachusetts
dlovley@microbio.umass.edu

Scott Lowman
Institute for Sustainable and Renewable Resources
jamesl71@vt.edu

Meredith Luschen
DOE Great Lakes Bioenergy Research Center (GLBRC)
mluschen@glbrc.wisc.edu

Suzanne Ma
DOE Joint BioEnergy Institute (JBEI)
srchhabra@lbl.gov

Jon Magnuson
Pacific Northwest National Laboratory
scott.baker@pnl.gov

Yuzuki Manabe
DOE Joint BioEnergy Institute (JBEI)
ymanabe@lbl.gov

Pin-Ching Maness
National Renewable Energy Laboratory
pinching.maness@nrel.gov

David Mann
DOE BioEnergy Science Center
gilnap@ornl.gov

Reinhold Mann
Brookhaven National Laboratory
mannrc@bnl.gov

Betty Mansfield
Oak Ridge National Laboratory
mansfieldbk@ornl.gov

Costas Maranas
Pennsylvania State University
costas@psu.edu

Gary Marchant
Arizona State University
gary.marchant@asu.edu

Victor Markowitz
DOE Joint Genome Institute
vmmarkowitz@lbl.gov

Sheryl Martin
Oak Ridge National Laboratory
martinsa@ornl.gov

Adam Martiny
University of California, Irvine
amartiny@uci.edu

Christopher Marx
Harvard University
cmarx@oeb.harvard.edu

Jon Massman
University of Minnesota
massm016@umn.edu

Fields Matthew
Montana State University
matthew.fields@erc.montana.edu

Xavier Mayali
Lawrence Livermore National Laboratory
pettridge2@llnl.gov

Lee Ann McCue
Pacific Northwest National Laboratory
leeann.mccue@pnl.gov

James McKinlay
University of Washington
mckinla1@uw.edu

Gail McLean
U.S. Department of Energy
Gail.McLean@science.doe.gov

David Mead
Lucigen Corp
dmead@lucigen.com

John Meeks
University of California, Davis
jcmeeks@ucdavis.edu

Chuansheng Mei
Institute for Advanced Learning and Research
chuansheng.mei@ialr.org

Bill Metcalf
University of Illinois
metcalf@life.illinois.edu

Noelle Metting
U.S. Department of Energy
noelle.metting@science.doe.gov

Thomas Metz
Pacific Northwest National Laboratory
rds@pnl.gov

Folker Meyer
Argonne National Laboratory
folker@mcs.anl.gov

Marissa Mills
Biological and Environmental Research Management Information System
millsmd@ornl.gov

Aleksandar Milosavlejvic
Baylor College of Medicine
AMILOSAV@BCM.EDU

Julie Mitchell
University of Wisconsin-Madison
jcmitchell@wisc.edu

Todd Mockler
Oregon State University
tmockler@cgrb.oregonstate.edu

Debra Mohnen
University of Georgia - Complex Carbohydrate Research Center
dmohnen@ccrc.uga.edu

Matthew Monroe
Pacific Northwest National Laboratory
matthew.monroe@pnl.gov

Stephen Moose
University of Illinois
smoose@illinois.edu

James Moran
Pacific Northwest National Laboratory
James.Moran@pnl.gov

Joanne Morello
U.S. Department of Energy
joanne.morello@ee.doe.gov

Jennifer Morrell-Falvey
Oak Ridge National Laboratory
morrelljl1@ornl.gov

Bill Moses
Lawrence Berkeley National Laboratory
wwmoses@lbl.gov

Wellington Muchero
DOE BioEnergy Science Center
gilnap@ornl.gov

Arijit Mukherjee
Great Lakes Bioenergy Research Center
University of Wisconsin, Madison
amukherjee@wisc.edu

Aindrila Mukhopadhyay
Lawrence Berkeley National Laboratory
amukhopadhyay@lbl.gov

Biswarup Mukhopadhyay
Virginia Tech
biswarup@vt.edu

John Mullet
Texas A&M University
jmullet@tamu.edu

David Myrold
Oregon State University
david.myrold@oregonstate.edu

Harish Nagarajan
University of California, San Diego
nh@ucsd.edu

Ali Navid
Lawrence Livermore National Laboratory
navid1@llnl.gov

Karen Nelson
J. Craig Venter Institute
kenelson@jcvi.org

Kelly Nevin
University of Massachusetts, Amherst
dlovley@microbio.umass.edu

Lee Newberg
Wadsworth Center, NYSDOH
lee.newberg@wadsworth.org

Basil Nikolau
Iowa State University
dimmas@iastate.edu

Trent Northen
Lawrence Berkeley National Laboratory
trnorthen@lbl.gov

Pavel Novichkov
Lawrence Berkeley National Laboratory
PSNovichkov@lbl.gov

Mari Nyyssonen
Lawrence Berkeley National Laboratory
mjnyyssonen@lbl.gov

Yasuhiro Oda
University of Washington
yasuhiro@uw.edu

Rachel Ogorzalek Loo
University of California, Los Angeles
JLoo@chem.ucla.edu

T. C. Onstott
Princeton University
tullis@princeton.edu

John Orban
University of Maryland College Park
jorban@umd.edu

Victoria Orphan
Caltech
vorphan@gps.caltech.edu

Andrei Osterman
Sanford-Burnham Medical Research Institute
osterman@sanfordburnham.org

Ross Overbeek
University of Chicago
stevens@cs.uchicago.edu

Steve Padgette
Monsanto Company
steve.padgette@monsanto.com

Kanchana Padmanabhan
North Carolina State University
kpadman@ncsu.edu

Himadri Pakrasi
I-CARES, Washington University in St. Louis
pakrasi@wustl.edu

Anthony Palumbo
Oak Ridge National Laboratory
palumboav@ornl.gov

Dehal Paramvir
Lawrence Berkeley National Laboratory
psdehal@lbl.gov

Morey Parang
Oak Ridge National Laboratory
parangm@ornl.gov

Byung-Hoon Park
Oak Ridge National Laboratory
parkbh@ornl.gov

Joshua Park
DOE Joint BioEnergy Institute (JBEI)
klsale@lbl.gov

Sivakumar Pattathil
University of Georgia, CCRC
siva@ccrc.uga.edu

Markus Pauly
University of California, Berkeley
mpauly69@berkeley.edu

Matteo Pellegrini
University of California, Los Angeles
matteop@mcdb.ucla.edu

Dale Pelletier
Oak Ridge National Laboratory
pelletierda@ornl.gov

Zhaohua Peng
Mississippi State University
zp7@ra.msstate.edu

Gary Peter
University of Florida
gfpeter@ufl.edu

Loukas Petridis
Oak Ridge National Laboratory
petridisl@ornl.gov

Jennifer Pett-Ridge
Lawrence Livermore National Laboratory
pettridge2@llnl.gov

Chris Petzold
Lawrence Berkeley National Laboratory
cjpetzold@lbl.gov

Somsak Phattarasukol
University of Washington
csh5@uw.edu

Ruth Pietri
Brookhaven National Laboratory
rpietri@bnl.gov

Madeleine Pincu
University of California, Irvine
mpincu@uci.edu

Sai Venkatesh Pingali
Oak Ridge National Laboratory
pingalis@ornl.gov

David Pletcher
DOE Joint BioEnergy Institute (JBEI)
Lawrence Berkeley National Laboratory, Technology Division
pdadams@lbl.gov

Jonathan Plett
INRA-Nancy, France
jm.plett@nancy.inra.fr

Mircea Podar
Oak Ridge National Laboratory
podarm@ornl.gov

Wei-Jun Qian
Pacific Northwest National Laboratory
weijun.qian@pnl.gov

Pablo Rabinowicz
U.S. Department of Energy
pablo.rabinowicz@science.doe.gov

Arthur Ragauskas
Georgia Institute of Technology
michele.yager@chemistry.gatech.edu

Srivatsan Raman
Harvard Medical School
sraman@genetics.med.harvard.edu

Hema Ramanna
DOE BioEnergy Science Center
gilnap@ornl.gov

Sridhar Ranganathan
Pennsylvania State University
sur152@psu.edu

Prabhasa Ravikirthi
Pennsylvania State University
pxr173@psu.edu

Jennifer Reed
University of Wisconsin-Madison
reed@engr.wisc.edu

Karin Remington
NIH
NIGMS
remingka@nigms.nih.gov

Michael Resch
DOE BioEnergy Science Center
National Renewable Energy Laboratory
michael.resch@nrel.gov

Virginia Rich
University of Arizona
vrich@email.arizona.edu

Kathryn Richmond
DOE Great Lakes Bioenergy Research Center (GLBRC)
kerichmond@glbrc.wisc.edu

William Riehl
Boston University
briehl@bu.edu

Elijah Roberts
University of Illinois
erobert3@illinois.edu

Andrea Rocha
University of South Florida
amrocha@mail.usf.edu

Dmitry Rodionov
Sanford-Burnham Medical Research Institute
rodionov@burnham.org

Margaret Romine
Pacific Northwest National Laboratory
margie.romine@pnl.gov

Cathy Ronning
U.S. Department of Energy
catherine.ronning@science.doe.gov

Gail Rosen
Drexel University
gailr@ece.drexel.edu

Eddy Rubin
DOE Joint Genome Institute
emrubin@lbl.gov

Rajib Saha
Pennsylvania State University
rus184@psu.edu

Kenneth Sale
DOE Joint BioEnergy Institute (JBEI)
klsale@lbl.gov

Kourosh Salehi-Ashtiani
New York University
ksa3@nyu.edu

Scott Saleska
University of Arizona
saleska@email.arizona.edu

Nagiza Samatova
North Carolina State University
Oak Ridge National Laboratory
samatovan@ornl.gov

Sanjaya Sanjaya
DOE Great Lakes Bioenergy Research Center (GLBRC)
sanjaya@msu.edu

Gautam Sarath
USDA-ARS
Gautam.Sarath@ars.usda.gov

Michael Saunders
Stanford University
saunders@stanford.edu

Christopher Schadt
Oak Ridge National Laboratory
schadtcw@ornl.gov

Amy Schaefer
University of Washington
amyschae@uw.edu

Courtney Schaffert
University of Nebraska Medical Center
cschaffert@unmc.edu

Henrik Scheller
Lawrence Berkeley National Laboratory
hscheller@lbl.gov

Jonathan Schilling
University of Minnesota
schilling@umn.edu

Skye Schmidt
University of Delaware
sschmidt@udel.edu

Gerrit Schut
University of Georgia
gerti@uga.edu

Michael Schwalbach
DOE Great Lakes Bioenergy Research Center (GLBRC)
mschwalbach@glbrc.wisc.edu

Daniel Serge
Boston University
dsegre@bu.edu

Margrethe Serres
Marine Biological Laboratory
mserres@mbl.edu

Joao Setubal
Virginia Bioinformatics Institute
setubal@vbi.vt.edu

Migun Shakya
Oak Ridge National Laboratory
shakyam@ornl.gov

Manoj Sharma
JBEI
mksharma@ucdavis.edu

Henry Shaw
Lawrence Livermore National Laboratory
shaw4@llnl.gov

Yaoqing Shen
University of California, Berkeley
emeralsh@berkeley.edu

Louis Sherman
Purdue University
lsherman@purdue.edu

Shengjing Shi
University of California, Berkeley
shengjing.shi@berkeley.edu

Lucas Showman
Iowa State University
lshowman@iastate.edu

Christopher Shymansky
DOE Joint BioEnergy Institute (JBEI)
CMShymansky@lbl.gov

Jessica Sieber
University of Oklahoma
jessica_rhea@ou.edu

Blake Simmons
DOE Joint BioEnergy Institute (JBEI)
basimmo@sandia.gov

Steven Singer
Lawrence Berkeley National Laboratory
SWSinger@lbl.gov

Anup Singh
Sandia National Laboratories
aksingh@sandia.gov

Gary Siuzdak
Lawrence Berkeley National Laboratory
Scripps
siuzdak@scripps.edu

Kimmen Sjolander
University of California, Berkeley
kimmen@berkeley.edu

Steven Slater
DOE Great Lakes Bioenergy Research Center (GLBRC)
scslater@glbrc.wisc.edu

Andreia Smith
DOE Joint BioEnergy Institute
University of California, Department of Plant Pathology
mevega-sanchez@lbl.gov

Dick Smith
Pacific Northwest National Laboratory
rds@pnl.gov

James Smith
DOE BioEnergy Science Center
gilnap@ornl.gov

James Smith
University of Georgia
jamess@uga.edu

Jeremy Smith
Oak Ridge National Laboratory
smithjc@ornl.gov

Kevin Sowers
University of Maryland Biotechnology Institute
sowers@umbc.edu

Alfred Spormann
Stanford University
spormann@stanford.edu

Thomas Squier
Pacific Northwest National Laboratory
thomas.squier@pnl.gov

Goundla Srinivas
Oak Ridge National Laboratory
smithjc@ornl.gov

Prem Srivastava
U.S. Department of Energy
prem.srivastava@science.doe.gov

Gary Stacey
University of Missouri
staceyg@missouri.edu

Robert Stack
U.S. Department of Energy
robert.stack@science.doe.gov

Brandon Stackhouse
Princeton University
tullis@princeton.edu

David Stahl
University of Washington
dastahl@u.washington.edu

Laurey Steinke
University of Nebraska Medical Center
lsteinke@unmc.edu

Rick Stevens
Argonne National Laboratory
University of Chicago
stevens@anl.gov

Marvin Stodolsky
U.S. Department of Energy
marvin.stodolsky@science.doe.gov

Sergey Stolyar
Pacific Northwest National Laboratory
sergey.stolyar@pnl.gov

Steven Strauss
Oregon State University
steve.strauss@oregonstate.edu

Poornima Sukumar
Oak Ridge National Laboratory
sukumar.poornima@ornl.gov

Zarath Summers
University of Massachusetts, Amherst
zsummers@microbio.umass.edu

Patrick Suthers
Pennsylvania State University
suthers@engr.psu.edu

Yilmaz Suzan
Sandia National Laboratory
syilmaz@sandia.gov

Jonathan Sweedler
Univeristy of Illlinois
jsweedle@illinois.edu

Nicole Taheri
Stanford University

John Tainer
Lawrence Berkeley National Laboratory
sara@scripps.edu

Yuhong Tang
DOE BioEnergy Science Center
ytang@noble.org

Michael Teresinski
U.S. Department of Energy
michael.teresinski@science.doe.gov

Hazen Terry
Lawrence Berkeley National Laboratory
tchazen@lbl.gov

Michael Thelen
DOE Joint BioEnergy Institute (JBEI)
Lawrence Livermore National Laboratory
mthelen@llnl.gov

Ines Thiele
University of Iceland
ithiele@hi.is

Brian Thomas
University of California, Berkeley
bcthomas@berkeley.edu

Kimberly Thomas
University of California, Los Angeles
kimberly.l.thomas-1@ou.edu

Michael Thorgersen
University of Georgia, Athens
mthorger@uga.edu

James Tiedje
Michigan State University
tiedjej@msu.edu

Christian Tobias
Agricultural Research Service
christian.tobias@ars.usda.gov

Trudy Torto-Alalibo
Virginia Bioinformatics Institute
trudy@vbi.vt.edu

Chung-Jui Tsai
University of Georgia
cjtsai@uga.edu

Timothy Tschaplinski
Oak Ridge National Laboratory
tschaplinstj@ornl.gov

Qichao Tu
University of Oklahoma
qichao@ou.edu

Gerald Tuskan
Oak Ridge National Laboratory
gtk@ornl.gov

Edward Uberbacher
DOE BioEnergy Science Center
gilnap@ornl.gov

Toshiyuki Ueki
University of Massachusetts
tueki@microbio.umass.edu

Patanjali Varanasi
JBEI
pvaranasi@lbl.gov

Miguel Vega-Sanchez
DOE Joint BioEnergy Institute (JBEI)
mevega-sanchez@lbl.gov

Nathan VerBerkmoes
Oak Ridge National Laboratory
verberkmoesn@ornl.gov

Wilfred Vermerris
University of Florida
wev@ufl.edu

Rytas Vilgalys
Duke University
fungi@duke.edu

John Vogel
USDA-ARS, Western Regional Research Center
john.vogel@ars.usda.gov

Eberhard Voit
Georgia Tech
DOE BioEnergy Science Center
eberhard.voit@bme.gatech.edu

Trang Vu
University of Wisconsin-Madison, Department of Chemical and Biological Engineering
tvu2@wisc.edu

Peter Walian
Lawrence Berkeley National Laboratory
PJWalian@lbl.gov

Judy Wall
University of Missouri
wallj@missouri.edu

Kuangyu Wang
North Carolina State University
kwang2@ncsu.edu

Zeng-Yu Wang
DOE BioEnergy Science Center
gilnap@ornl.gov

David M. Ward
Montana State University
umbdw@montana.edu

Sharlene Weatherwax
U.S. Department of Energy
sharlene.weatherwax@science.doe.gov

Peter Weber
Lawrence Livermore National Laboratory
weber21@llnl.gov

David Weston
Oak Ridge National Laboratory
westondj@ornl.gov

Janet Westpheling
Univeristy of Georgia
janwest@uga.edu

Libby White
U.S. Department of Energy
elizabeth.white@science.doe.gov

Owen White
University of Maryland
owhite@som.umaryland.edu

Steven Wiley
Pacific Northwest National Laboratory
steven.wiley@pnl.gov

Curtis Wilkerson
Michigan State University
wilker13@msu.edu

Halina Ewa Witkowska
University of California, San Francisco
witkowsk@cgl.ucsf.edu

Amy Wolfe
Oak Ridge National Laboratory
wolfeak@ornl.gov

Alexandra Worden
MBARI/UCSC
azworden@mbari.org

Aaron Wright
Pacific Northwest National Laboratory
aaron.wright@pnl.gov

JH David Wu
University of Rochester
davidwu@che.rochester.edu

Charles Wyman
DOE BioEnergy Science Center
gilnap@ornl.gov

Fangfang Xia
Argonne National Laboratory
fangfang.xia@gmail.com

Xiaoliang Sunney Xie
Harvard University
xie@chemistry.harvard.edu

Qi Xu
DOE BioEnergy Science Center
gilnap@ornl.gov

Ying Xu
University of Georgia
xyn@bmb.uga.edu

Kai Xue
University of Oklahoma
kx36@ou.edu

Koon-Kiu Yan
Yale University
koon-kiu.yan@yale.edu

Fan Yang
DOE Joint BioEnergy Institute (JBEI)
Lawrence Berkeley National Laboratory
fyang@lbl.gov

Liming Yang
NIH
lyang@mail.nih.gov

Shuo Yang
University of Wisconsin-Madison
syang@chem.wisc.edu

Xiaohan Yang
Oak Ridge National Laboratory
yangx@ornl.gov

Steve Yannone
Lawrence Berkeley National Laboratory
SMYannone@lbl.gov

Safak Yilmaz
University of Wisconsin-Madison
yilmaz@cae.wisc.edu

Sung Ho Yoon
Institute for Systems Biology
syoon@systemsbiology.org

William York
DOE BioEnergy Science Center
will@ccrc.uga.edu

Shireen Yousef
U.S. Department of Energy
shireen.yousef@science.doe.gov

Donald Zak
University of Michigan
drzak@umich.edu

Lydia Zeglin
Oregon State University
lydia.zeglin@oregonstate.edu

Karsten Zengler
University of California, San Diego
kzengler@ucsd.edu

Sarah Zerbs
Argonne National Laboratory
fcollart@anl.gov

He Zhang
University of Wisconsin-Madison,
BACTER, GLBRC
hzhang32@wisc.edu

Fengfeng Zhou
University of Georgia
ffzhou@csbl.bmb.uga.edu

Jizhong Zhou
University of Oklahoma
jzhou@ou.edu

Alireza Zomorrodi
Pennsylvania State University
zomorrodi@engr.psu.edu

# Appendix 2: Websites

## Genomic Science Program Websites

- Genomic Science program http://genomicscience.energy.gov
- Office of Biological and Environmental Research Genomic Science http://www.sc.doe.gov/ober/BSSD/genomics.html
- This book http://genomicscience.energy.gov/pubs/2011abstracts/
- Plant Feedstock Genomics Research http://genomicscience.energy.gov/research/DOEUSDA/
- Publications
    - *Systems Biology Knowledgebase Implementation Plan* (September 2010) http://genomicscience.energy.gov/compbio/
    - *New Frontiers in Characterizing Biological Systems* (October 2009) http://genomicscience.energy.gov/characterization/
    - *Sustainability of Biofuels: Future Research Opportunities* (March 2009) http://genomicscience.energy.gov/biofuels/sustainability/
    - *Carbon Cycling and Biosequestration: Integrating Biology and Climate Through Systems Science* (December 2008) http://genomicscience.energy.gov/carboncycle/

## Bioenergy Research Center Websites

- Overview http://genomicscience.energy.gov/centers/
- BioEnergy Science Center http://www.bioenergycenter.org
- Great Lakes Bioenergy Research Center http://www.greatlakesbioenergy.org
- Joint BioEnergy Institute http://www.jbei.org

## DOE User Facilities

- DOE Joint Genome Institute http://jgi.doe.gov
- Environmental Molecular Sciences Laboratory http://www.emsl.pnl.gov/emslweb/

## Project Websites

- Root Hair Systems Biology http://soyroothair.org
- Poplar Biomass Interactome Project http://xylome.vbi.vt.edu/
- Genetic Dissection of Bioenergy Traits in Sorghum http://www.sorghum.genome.ufl.edu
- Plant-Microbe Interfaces (PMI) http://PMI.ornl.gov
- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) http://enigma.lbl.gov
    - ENIGMA and Hydrogen Regulation http://baliga.systemsbiology.net/enigma/AMD
- Carbon Cycle Project http://geomicrobiology.berkeley.edu/pages/amd.html
- Geobacter Project http://www.geobacter.org
- UCLA-DOE Institute for Genomics and Proteomics http://www.doe-mbi.ucla.edu/overview
- MicrobesOnline http://microbesonline.org

**Project Websites Continued**

- Virtual Institute for Microbial Stress and Survival http://vimss.lbl.gov
- Gene 2 Function http://carb.umbi.umd.edu/g2f
- Ribosomal Database Project (RDP) http://rdp.cme.msu.edu
- Naïve Bayesian Classification Tool http://nbc.ece.drexel.edu
- GAGGLE http://gaggle.systemsbiology.net
- The openCOBRA Project http://opencobra.sourceforge.net
- BESC KnowledgeBase http://bobcat.ornl.gov/besc/
- National Center for X-ray Tomography http://ncxt.lbl.gov
- Biomass Conversion Research Laboratory, Michigan State University http://www.everythingbiomass.org
- Jackson Lab Grassland Ecology, University of Wisconsin-Madison http://agronomy.wisc.edu/jackson/
- Biofuel Feedstock Genomics Resource http://bfgr.plantbiology.msu.edu/
- Corn Functional Lab http://cropsci.illinois.edu/faculty/moose/lab/energy.html
- Buckler Lab for Maize Genetics http://www.maizegenetics.net
- Brown University Center for Computational Molecular Biology http://www.brown.edu/Research/CCMB/
- Harvard Molecular Technology Group and Lipper Center for Computational Genetics http://arep.med.harvard.edu/Geomicrobiology Group http://www.princeton.edu/southafrica/permafrost-project/
- $H_2$ Regulation http://baliga.systemsbiology.net/drupal/content/h2-regulation
- Systems Biology Research Group http://systemsbiology.ucsd.edu
- BER-PNNL Proteomics http://ober-proteomics.pnl.gov/
- PhyloFacts: Phylogenomic encyclopedias across the Tree of Life http://phylogenomics.berkeley.edu/phylofacts/
- The Model SEED http://seed-viewer.theseed.org/models/
- MENGO - Microbial ENergy processes Gene Ontology Project http://mengo.vbi.vt.edu/

# Author Index

**Indexed by Page Number**

## C



## D

## E



## F



## G



## H

## I



## J



## K

## L



## M

## N



## O



## P

## Q



## R



## S

## T

## U



## V



## W



## X



## Y

## Z

# Institutional Index

**Indexed by Page Number**

**Institutions determined by affiliations provided by authors**