DOE/SC-0142

DOE GENOMIC SCIENCE

SYSTEMS BIOLOGY FOR ENERGY AND ENVIRONMENT

2012

DOE Genomic Science Awardee Meeting X

U.S. DEPARTMENT OF ENERGY
Office of Science

February 26-29, 2012

U.S. DEPARTMENT OF
ENERGY
Office of Science

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
Joseph Graber
Acting Director, Biological Systems Science Division
U.S. Department of Energy
Office of Biological and Environmental Research
301-903-1239
joseph.graber@science.doe.gov

An electronic version of this document is available at the Genomic Science website:
http://genomicscience.energy.gov/pubs/2012abstracts/

DOE/SC-0142

U.S. DEPARTMENT OF
ENERGY
Office of Science

2012

Genomic Science Awardee Meeting X

Bethesda, Maryland
February 26-29, 2012

[Revised: March 1, 2012]

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, D.C.
February 26, 2012

Dear Colleagues:

Welcome to the 2012 Department of Energy (DOE) Genomic Science Program Awardee Meeting. This workshop brings together researchers supported by the program, representatives from DOE, 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 research and technology development. This program provides the foundation for biological solutions to DOE mission challenges in energy, environment, and climate. The objectives of the program are to:

- *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.*

The Genomic Science program continues to support groundbreaking research by individual investigators, interdisciplinary research teams, and larger-scale integrated research centers. Important scientific contributions by all of these research configurations will be highlighted at this year's meeting. The DOE Bioenergy Research Centers have completed their fourth year of operations and will report on exciting new results that are paving the way for the conversion of cellulosic biomass to biofuels. Plenary sessions will also feature discussions on advances in genome-enabled research linking microbial community function to ecosystem-scale processes, development of novel tools and techniques to extend systems biology to multispecies assemblages, and new lines of research aimed at targeted functional modification of plants and microbes.

In the past year, the Genomic Science program has engaged in a series of discussions aimed at building on our strong record of breakthrough research in synthetic biology and developing fundamental understanding of the design rules governing biological systems. In addition to numerous research presentations on these topics, this year's meeting will feature a presentation on the recently published report from the DOE Biosystems Design Workshop, which discusses the current state of the field and critical knowledge gaps that must be addressed to accelerate progress. In keeping with this theme, this year's keynote presentation will be from Dr. James Collins of Boston University, who will present a lecture titled "Life Redesigned: The Emergence of Synthetic Biology."

The Genomic Science program is committed to advancing research by providing community access to unique technological capabilities at DOE Office of Science (SC) user facilities. Plenary talks will highlight BER-funded research at two such facilities: the DOE Joint Genome Institute and the SLAC National Accelerator Laboratory. DOE will also present the output of a recent workshop considering the development of novel analytical capabilities for high-impact biological sciences at new DOE SC national user facilities. Finally, Dr. Adam Arkin of Lawrence Berkeley National Laboratory will discuss the initiation of the Genomic Science program's Systems Biology Knowledgebase, a cyberinfrastructure that will serve as a community resource for data integration, visualization, and modeling.

Topical breakout sessions at this year's meeting display the broad range of research interests being pursued in the Genomic Science program. These include development of next-generation analytical tools for plant and microbial systems biology, development of new computational biology capabilities, and consideration of societal concerns relevant to program research. This year, we are also pleased to present a breakout session highlighting research on photosynthetic systems supported by DOE's Office of Basic Energy Sciences. This program supports research at the interface of biology, chemistry, and biophysics and should be of strong interest to Genomic Science program grantees. We will again feature a series of student research presentations during lunch hours and evening poster sessions to highlight new results and promote in-depth discussion of program research.

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.
Associate Director of Science for Biological and Environmental Research (BER)
Office of Science

Contents

Welcome iii

Contents v

Agenda vi

Hotel Map xiii

Abstract Titles xv

Workshop Abstracts 1

- **Bioenergy Research Centers** 3
  - Joint BioEnergy Institute (JBEI) 3
  - Great Lakes Bioenergy Research Center (GLBRC) 14
  - BioEnergy Science Center (BESC) 30
- **Biofuels Research** 47
- **Biological Hydrogen Production** 67
- **Biological Systems Research on the Role of Microbial Communities in Carbon Cycling** 79
- **Systems Biology Strategies and Technologies for Understanding Microbes, Plants, and Communities** 107
- **Knowledgebase and Computing for Systems Biology** 175
- **Communication and Ethical, Legal, and Societal Issues** 207

Appendix 1: Participants 211

Appendix 2: Websites 219

- Genomic Science Program Websites 219
- Bioenergy Research Center Websites 219
- DOE User Facilities 219
- Project Websites 219

Author Index 221

Institutional Index 231

Agenda

Correct as of February 7, 2012

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 12:30 p.m. Wednesday.

Sunday, February 26

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

Monday, February 27

7:00-8:30 am	**Registration**	
7:30-8:30 am	**Continental Breakfast**	
8:30–9:00 am	**Welcome and Introduction to the Meeting**	
	8:30-8:45 am	**Sharlene Weatherwax**, Associate Director DOE Office of Biological and Environmental Research
	8:45-9:00 am	**Joseph Graber,** Genomic Science Program Team Lead DOE Office of Biological and Environmental Research
9:00–11:00 am	**Plenary Session: DOE Bioenergy Research Centers** **Location:** Grand Ballroom ABCD **Moderator:** Cathy Ronning, DOE BER	
	Speakers:	
	9:00–9:30 am	**Jim Tiedje**, GLBRC/Michigan State University “Metagenomics of the Prairie and Biofuel Crop Rhizospheres”
	9:30–10:00 am	**Art Ragauskas**, BESC/Georgia Institute of Technology “Characterizing the Mechanisms of Reduced Recalcitrance of Biomass”
	10:00–10:30 am	**Aindrila Mukhopadhyay**, JBEI/Lawrence Berkeley National Laboratory “Engineering Advanced Microbes for Biofuel Production”
10:30–11:00 am	Break	
11:00 am–12:00 pm	**Keynote Presentation:** **Jim Collins**, Boston University “Life Redesigned: The Emergence of Synthetic Biology”	
12:00-2:00 pm	**Lunch**	
12:30–2:00 pm	**Lunchtime Student Oral Presentations** **Location:** Grand Ballroom ABCD **Moderator:** Libby White Speakers to be Announced	

Breakout Sessions

2:00–5:00 pm **Breakout Session A: Plants and Their Environment: Biology, Molecular Interactions, and Homeostasis**
Location: Forest Glen Room (Lower Level)
Moderators: Cathy Ronning and Prem Srivastava, DOE BER

Description of Session: Study of plant genomics and homeostatic processes in living plants can be based on *in situ* chemical reactions and the environmental effects of perturbations of these chemical reactions. Molecular plant-environment (e.g. microbe) interactions and metabolism, and the use of radioactive tracers to detect and study perturbations of homeostatic reactions will be emphasized in this session.

Speakers:

2:00–2:10 pm **Session Introduction—Cathy Ronning**

2:10–2:40 pm **Jean Greenberg**, University of Chicago
"Tracking the Movement of Bio-Active Peptides in *Arabidopsis* and Poplar"

2:40–3:10 pm **Richard Ferrieri**, Brookhaven National Laboratory
"The Role of Radiochemistry in Systems Biology to Study the Interrelationship between Hormone Signaling and Plant Root Development"

3:10–3:30 pm Break

3:30–3:50 pm **Jocelyn Rose**, Cornell University
"Peeling Apart the Structural and Functional Complexity of the Cuticularized Plant Cell Wall"

3:50–4:10 pm **Leland Cseke**, University of Alabama, Huntsville
"Merging Bottom-Up with Top-Down Research in Symbiosis and Nutrient Cycling"

4:10–5:30 pm **Roundtable Discussion—Prem Srivastava**
Open Discussions on the Current State of the Art and Future Direction

2:00–5:00 pm **Breakout Session B: Enabling Tools and Applications for the DOE Systems Biology Knowledgebase**
Location: Brookside Room A/B (Lower Level)
Moderator: Susan Gregurick, DOE BER

Description of Session: The Systems Biology Knowledgebase (Kbase) will be a community-driven software framework enabling the data-driven prediction of microbial, plant, and biological community function in an environmental context. Extensible and scalable, Kbase also will feature open architecture, source code, and open development. The intent of Kbase is to provide access to a distributed, scalable computing resource for data-intensive analysis and will support a large user community with tools and services for **Microbes:** Reconstruction and prediction of metabolic and gene expression regulatory networks for 100 to 1,000 microbes to manipulate microbial function; **Plants:** Integration of phenotypic and experimental data and metadata for 10 key plants related to DOE missions to predict biomass properties from genotype and assemble regulatory data to enable analysis, cross-comparisons, and modeling; and **Communities:** Modeling metabolic processes within 10 to 100 microbial communities with DOE relevance and mine metagenomic data to identify unknown genes. This session will highlight some of the modeling tools and capabilities funded by DOE that could enhance the Knowledgebase.

Speakers:

2:00-2:30 pm **Kimmen Sjolander**, University of California, Berkeley
"The PhyloFacts 3.0 Phylogenomic Encyclopedia of Microbial Gene Families: New Developments and Plans"

2:30-3:00 pm	**Peter Karp**, SRI International “Introducing Metabolic Engineering and Atom Mapping Capabilities into Pathway Tools”
3:00-3:30 pm	**Chris Henry**, Argonne National Laboratory “Enhancing the SEED Framework for Curation and Analysis of Genomic Data and Genome-Scale Metabolic Models”
3:30-3:45 pm	Break
3:45-4:15 pm	**Mark Gerstein**, Yale University “Tools and Approaches for Integrating Multiple Genetic and Cellular Networks”
4:15-4:45 pm	**Daniel Segre**, Boston University “From Genome-Scale to Ecosystem-Level Models of Metabolism”
4:45-5:00 pm	**Discussion of Gaps and Opportunities in Enabling Methods for the Knowledgebase**

2:00–5:00 pm **Breakout Session C: Biological Structure Research in the Genomic Science Program**
Location: Glen Echo Room (Lower Level)
Moderator: Roland Hirsch, DOE BER

Description of Session: Research in systems biology requires application of a wide range of technologies, each of which reveals aspects of the processes that occur in living cells and organisms. The DOE synchrotron light sources and neutron beam facilities provide many experimental capabilities that are being used in research projects in the Genomic Science program. The talks in this breakout session will explain how specific beamlines at those facilities are being used in GSP projects. The speakers will seek to make the technologies understandable for GSP biologists by focusing on the biological science that is enabled by the beamline experiments.

Speakers:

2:00–2:05 pm	**Session Introduction**
2:05–2:40 pm	**Sol Gruner**, Cornell University **Session Keynote:** “Biomolecules Under Pressure: Why it Matters”
2:40–3:00 pm	**Frank Collart**, Argonne National Laboratory Representing: Argonne Structural Biology Center “Binding Profiles and Crystal Structures of Bacterial Solute Binding Proteins for Transport of Aromatic Products of Lignin Degradation”
3:00–3:20 pm	**Zöe Fisher**, Los Alamos National Laboratory Representing: Los Alamos Protein Crystallography Station “Elucidating the Proton Transfer Mechanism of Carbonic Anhydrase Using Joint Neutron/X-Ray Crystallography”
3:20–3:40 pm	Break
3:40–4:00 pm	**George Phillips**, University of Wisconsin Representing: Berkeley Small Angle X-Ray Scattering Program “Characterization of Cellulose Deconstruction Enzymes by Small-Angle X-Ray Scattering and X-Ray Diffraction”
4:00–4:20 pm	**Gang Cheng**, Lawrence Berkeley National Laboratory Representing: ORNL Center for Structural Molecular Biology “Understanding the Effect of Ionic Liquid Treatment on the Structures of Lignins in Solutions by Small Angle Neutron Scattering”

4:20–4:40 pm **Terry Hazen**, University of Tennessee
Representing: Berkeley Synchrotron Infrared Structural Biology Program
"Label-Free Monitoring of Chemical Reactions in Cells During Stress-Adaptive Response"

4:40–5:00 pm **Blake Simmons**, Sandia National Laboratory
Representing: Berkeley National Center for X-Ray Tomography
"Understanding the Lipid Trigger in Algae using X-Ray Tomography and Spectroscopy"

5:30–8:00 pm **Poster Session**
Location: Grand Ballroom EFGH

Tuesday, February 28

8:30–10:00 am **Plenary Session: Small Cogs Turn Big Wheels: Microbial Communities and the Carbon Cycle**
Location: Grand Ballroom ABCD
Moderator: Joe Graber, DOE BER

Speakers:

8:30–9:00 am **Steve Allison**, University of California, Irvine
"Linking Microbial Enzyme Genes with Community Responses to Drought and Nitrogen"

9:00–9:30 am **Dave Myrold**, Oregon State University
"Meta-omics Analysis of Microbial Carbon Cycling Responses to Altered Rainfall Inputs in Native Prairie Soils"

9:30–10:00 am **Cheryl Kuske**, Los Alamos National Laboratory
"Patterns of Soil Community Response to Elevated Atmospheric CO_2 Across Terrestrial Ecosystems"

10:00–10:30 am Break

10:30 am–12:00 pm **Plenary Session: DOE User Facilities and Community Resources**
Location: Grand Ballroom ABCD
Moderator: Dan Drell, DOE BER

Speakers:

10:30–11:00 am **Eddy Rubin**, Joint Genome Institute, Lawrence Berkeley National Laboratory
"Science at the JGI"

11:00–11:30 am **Adam Arkin**, Lawrence Berkeley National Laboratory
"The DOE Systems Biology Knowledgebase"

11:30 am-12:00 pm **Keith Hodgson**, SLAC National Accelerator Laboratory
"Recent Developments with the LCLS X-Ray FEL at SLAC and Prospects for Future Science"

12:00-12:15 pm **DOE Report Update:** "Biosystems Design: Report from the July 2011 Workshop"
Pablo Rabinowicz, DOE BER

12:15-12:30 pm **DOE Report Update:** "Applications of New DOE National User Facilities in Biology: Report from the May 2011 Workshop"
Roland Hirsch, DOE BER

12:30-2:00 pm Lunch

1:00–2:00 pm **Lunchtime Student Oral Presentations**
Location: Grand Ballroom ABCD
Moderator: Libby White, DOE BER
Speakers to be Announced

Breakout Sessions

2:00–5:00 pm **Breakout Session D: DOE Office of Basic Energy Sciences: Photosynthetic Systems Research**
Location: Brookside Room A/B (Lower Level)
Moderator: Gail McLean, DOE Office of Basic Energy Sciences

Description of Session: This session highlights projects in the Photosynthetic Systems program in DOE's Office of Basic Energy Sciences (BES). This program supports basic research in natural photosynthesis and brings together biology, chemistry, biochemistry, and biophysics to uncover the fundamental science of the biological capture of sunlight and its conversion to and storage as chemical energy.

Speakers:

2:00–2:10 pm **Session Introduction**

2:10–2:40 pm **Kevin Redding**, Arizona State University
"Routing of Electrons in the Photosynthetic Firmicute, *Heliobacterium modesticaldum*"

2:40–3:10 pm **Christine Kirmaier**, Washington University in St. Louis
"Controlling Electron Transfer Pathways in Photosynthetic Reaction Centers"

3:10–3:40 pm **K.V. Lakshmi**, Rensselaer Polytechnic Institute
"The Mechanism of Solar Water Oxidation in Nature: A Cross Species Comparison of the Structure of the S-State Photochemical Intermediates of Photosystem II"

3:40–4:00 pm Break

4:00–4:30 pm **Ann McDermott**, Columbia University
"Shifting Shapes: NMR Studies of Conformational Flexibility in Light Harvesting Complexes"

4:30–5:00 pm **Robert Burnap**, Oklahoma State University
"Genetic Regulatory Circuits Integrating the Light and Dark Reactions of Oxygenic Photosynthesis"

2:00–5:00 pm **Breakout Session E: Integrating Societal Considerations/Impacts into BER-Funded Research**
Location: Forest Glen Room (Lower Level)
Moderator: Libby White, DOE BER

Description of Session: As BER's Genomic Science program's research progresses, DOE will continue to incorporate into such research a component that addresses societal considerations/impacts. This session will focus on ongoing efforts at DOE and other organizations to look at societal considerations/impacts, as well as gaps/potential future ELSI research needs.

Speakers:

2:00-2:05 pm **Session Introduction**
Libby White, BER

2:05-2:25 pm **Amy Wolfe**, Oak Ridge National Laboratory
"Societal Implications of Science and Technology Research and Development Undertaken at U.S. DOE Research Centers"

2:25-2:45 pm	**Sarah Carter**, Venter Institute “Managing the Risks of Synthetic Biology Assessing the U.S. Regulatory System”
2:45-3:05 pm	**Nathan Hillson**, Lawrence Berkeley National Laboratory “Assessing and Mitigating the Risks of Large-Scale Metabolic Engineering”
3:05-3:20 pm	Break
3:20-3:40 pm	**Paula Olsiewski**, Sloan Foundation “Overview of Sloan Foundation’s Synbio-ELSI Program”
3:40-4:00 pm	**Dave Rejeski**, Wilson Center “Summary of Fall 2010 Workshop and Other Related Wilson Center Efforts”
4:00–4:30 pm	**Tom Murray**, The Hastings Center “The Evolution of ELSI: Where it is Today and How it Could Respond to Synthetic Biology”
4:30–5:00 pm	**Wrap-up Discussion/Panel Discussion of Previous Speakers**

2:00–5:00 pm **Breakout Session F: Innovative Analytical and Imaging Technology for Plants and Microbes**
Location: Glen Echo Room (Lower Level)
Moderators: Arthur Katz and Dean Cole, DOE BER

Description of Session: The Genomic Science program supports basic research that includes the application and development of a variety of imaging and analytical technologies. The biological challenge for these technologies remains extending their capabilities in order to simultaneously measure multiple chemical and biological species at multiple scales within complex, heterogeneous cellular and environmental systems. One critical step will be increasing temporal and spatial resolution. This breakout session will introduce current capabilities of key technologies and relate how they can be used to address significant biological problems of interest to the BER community.

Speakers:

2:00-2:30 pm	**Mitch Doktycz**, Oak Ridge National Laboratory “Microfluidic Technologies for Characterizing Plant-Microbe Interfaces”
2:30–3:00 pm	**Haw Yang**, Princeton University “Visualizing Molecular Reactivity in Context”
3:00–3:30 pm	**Paul Bohn**, Notre Dame University “Applications of In Situ Raman Microscopy and Spectroscopy to Spatially and Temporally Distributed Processes in Complex Multi-Organismal Systems”
3:30–3:45 pm	Break
3:45–4:15 pm	**Kenneth Hammel**, University of Wisconsin “Fungal Biodegradative Oxidants in Lignocellulose: Fluorescence Mapping and Correlation with Gene Expression”
4:15–4:45 pm	**Drew Weisenberger**, Jefferson Laboratory “PET Radiotracer Imaging in Plant Biology”
4:45–5:00 pm	Observations and Comments

5:30–8:00 pm **Poster Session**
Location: Grand Ballroom EFGH

Wednesday, February 29

8:30–10:00 am **Plenary Session: Extending Systems Biology to the Community Scale**
Location: Grand Ballroom ABCD
Moderator: Roland Hirsch, DOE BER

Speakers:

8:30–9:00 am **Margie Romine**, Pacific Northwest National Laboratory
"Genome Sequence-Enabled Investigations of Microbial Interactions in Mat Communities"

9:00–9:30 am **Chris Marx**, Harvard University
"Evolution of Cooperation in Synthetic, Multi-Species Microbial Consortia"

9:30–10:00 am **Dave Stahl**, University of Washington
"Adaptive and Evolutionary Dimensions of Microbial Communities"

10:00–10:30 am **Break**

10:30–11:50 am **Plenary Session: DOE Early Career Research Program**
Location: Grand Ballroom ABCD
Moderator: John Houghton, DOE BER

Speakers:

10:30–10:50 am **Susannah Tringe**, Lawrence Berkeley National Laboratory
"Microbial Communities in Restored Freshwater Wetlands"

10:50–11:10 am **Sam Hazen**, University of Massachusetts
"Optimizing Plant-Microbial Systems for Energy—Mapping Feedstock Quality Genes in *Brachypodium distachyon*"

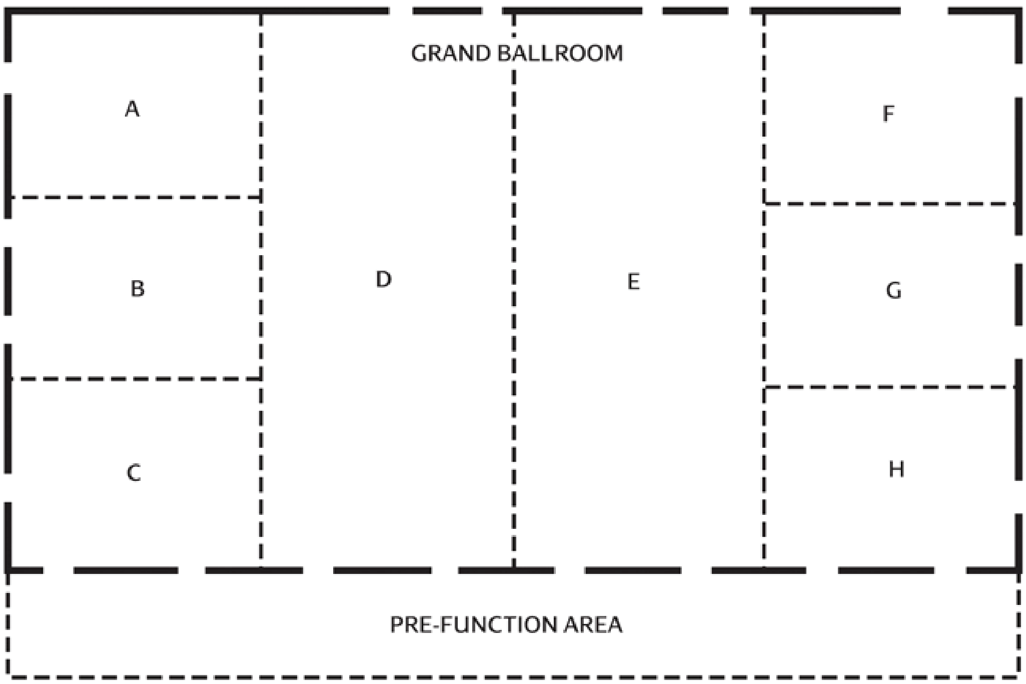
11:10-11:30 am **Yongqin Jiao**, Lawrence Livermore National Laboratory
"Understanding Uranium Resistance and Mineralization by *Caulobacter crescentus*"

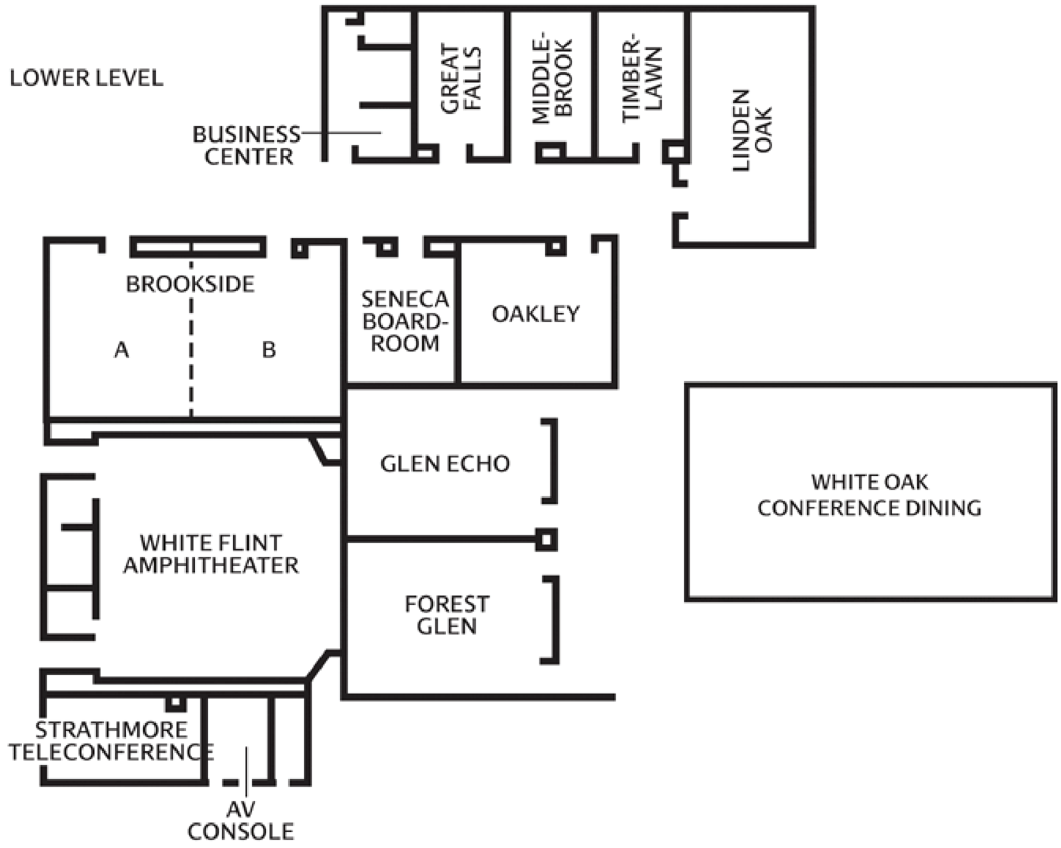
11:30-11:50 am **Mary Dunlop**, University of Vermont
"Engineering Robust Hosts for Microbial Biofuel Production"

12:00 pm **Closeout and Adjournment**

Hotel Map

MAIN LEVEL


GRAND BALLROOM
A
B
C
D
E
F
G
H
PRE-FUNCTION AREA


LOWER LEVEL
BUSINESS CENTER
GREAT FALLS
MIDDLEBROOK
TIMBERLAWN
LINDEN OAK
BROOKSIDE
A
B
SENECA BOARDROOM
OAKLEY
GLEN ECHO
WHITE OAK CONFERENCE DINING
WHITE FLINT AMPHITHEATER
FOREST GLEN
STRATHMORE TELECONFERENCE
AV CONSOLE

Abstract Titles

Poster **Page**

Bioenergy Research Centers 3

Joint BioEnergy Institute (JBEI)

1 **The Challenge of Enzyme Cost in the Production of Lignocellulosic Biofuels** 3
Daniel Klein-Marcuschamer* (DKlein@lbl.gov), Piotr Oleskowicz-Popiel, Blake A. Simmons, and **Harvey W. Blanch**

2 **Structural Comparison of Plant Glycosyltransferases** 3
Sara Fasmer Hansen, Andy DeGiovanni, Peter McInerney, Masood Hadi, Ryan McAndrew, Jose Henrique Pereira, Paul Adams, and **Henrik Vibe Scheller*** (hscheller@lbl.gov)

3 **A Genome-Wide Survey of Switchgrass Genome Structure and Organization** 4
Peijian Cao, Jerry Jenkins, Laura Bartley, Jane Grimwood, Jeremy Schmutz, Daniel Rokhsar, Manoj Sharma,* (MKSharma@lbl.gov), and **Pamela C. Ronald**

4 **MASCP Gator and ModHunter, Bioinformatics Tools for Identifying Post Translational Modifications in *Arabidopsis*** 4
Joshua Heazlewood* (JHeazlewood@lbl.gov), Hiren Joshi, Andrew Carroll, and Harriet Parsons

5 **Biosynthesis and Incorporation of Side-Chain-Truncated Lignin Monomers to Reduce Lignin Polymerization and Enhance Saccharification** 4
Aymerick Eudes* (ageudes@lbl.gov), Jin Sun Kim, Anthe George, Purba Mukerjee, Brigitte Pollet, Peter I. Benke, Fan Yang, Prajakta Mitra, Lan Sun, Ozgul Persil-Cetinkol, Salem Chabout, Grégory Mouille, Ludivine Soubigou-Taconnat, Sandrine Balzergue, Seema Singh, Bradley M. Holmes, Aindrila Mukhopadhyay, Jay D. Keasling, Blake A. Simmons, Catherine Lapierre, John Ralph, and **Dominique Loque**

6 **Engineering of Plants with Decreased Xylan and Lignin Contents and Increased Cell Wall Density** 5
Jane Lau, Fan Yang, Pia Damm Petersen, Ling Zhang, Prajakta Mitra, Lan Sun, Lina Prak, Berit Ebert, Yves Verhertbruggen, Jin Sun Kim, Anongpat Suttangkakul, Manfred Auer, Henrik Vibe Scheller, and **Dominique Loque*** (dloque@lbl.gov)

7 **Functional and Structural Characterization of Rice Cellulose Synthase-Like f6 Loss-of-Function Mutants** 5
Andreia Michelle Smith-Moritz, Yves Verhertbruggen, Jeemeng Lao, Lina Prak, Manfred Auer, Henrik V. Scheller, Zhao Hao, Hoi-Ying N. Holman, Pamela C. Ronald, Joshua L. Heazlewood, and **Miguel E. Vega-Sanchez*** (mevega-sanchez@lbl.gov)

8 **Feedstock Agnostic Pretreatment Technology** 6
Seema Singh* (SSingh@lbl.gov), Kevin George, Rohit Arora, Chenlin Li, Ian Mathews, Jay Keasling, Taek-Soon Lee, and Blake Simmons

9 **Understanding the Interactions of Cellulose with Ionic Liquids and Ionic Liquid/Water Binary Mixture: A Molecular Dynamics Study** 6
Hanbin Liu, Gang Cheng, Hanbin Liu* (hliu2@lbl.gov), Mike Kent, Vitalie Stavila, Kenneth L. Sale, Blake A. Simmons, and **Seema Singh**

Poster **Page**

10 **A Systematic Pipeline for Biomass Characterization Using Aligned Mechanical Stress Analysis, Polarized Raman Microspectroscopy and Scanning Electron Microscopy** 7
Lan Sun, Jacob Katsnelson, Patanjali Varanasi* (PVaranasi@lbl.gov), Rita Sharma, Manoj Sharma, Miguel Vega-Sanchez, Marcin Zemla, David Larson, Pamela Ronald, Blake Simmons, Paul Adams, Seema Singh, and **Manfred Auer**

11 **Interaction of Endoglucanases with Amorphous Cellulose Revealed by Quartz Crystal Microbalance and Neutron Reflectivity** 7
Gang Cheng* (gcheng@lbl.gov), Supratim Datta, Chao Wang, Zelin Liu, Jaclyn K. Murton, Page Brown, Michael Jablin, Manish Dubey, Jaroslaw Majewski, Candice Halbert, James Browning, Alan Esker, Brian J. Watson, Haito Zhang, Steven W. Hutcheson, Dale Huber, Blake A. Simmons, and **Michael S. Kent**

12 **Nanostructure-Initiator Mass Spectrometry (NIMS): High-Throughput Enzyme Activity Assays for Biofuel Development** 8
Xiaoliang Cheng* (xiaoliangcheng@lbl.gov), Wolfgang Reindl, Kai Deng, Benjamin Bowen, Bergeman Lai, John M. Gladden, Steven W. Singer, April Wong, Terry C. Hazen, Brian Fox, Kenneth Sale, Blake A. Simmons, Anup K. Singh, Jay Keasling, Paul D. Adams, and **Trent R. Northen**

13 **Tuning Cellulase Activity Using Carbohydrate Binding Modules** 8
Richard Heins* (raheins@lbl.gov), Vimalier Reyes-Ortiz, Huu Tran, Edward Kim, Nathan Hillson, Rajat Sapra, Kenneth Sale, Blake Simmons, Danielle Tullman-Ercek, and **Masood Hadi**

14 **Tracing the Determinants of Dual-Substrate Specificity in a Diverse Subfamily of Family 5 Glycoside Hydrolases** 9
G.D. Friedland* (gdfriedland@lbl.gov), Z. Chen, J. Pereira, S.A. Reveco, R. Chan, J.I. Park, M.P. Thelen, P.D. Adams, A.P. Arkin, J.D. Keasling, H.W. Blanch, B.A. Simmons, K.L. Sale, D. Chivian, and **S. Chhabra**

15 **Metagenomics, Metabolic Reconstruction, and High-Resolution Proteomics of Biomass Degradation in a Thermophilic Bacterial Community** 9
Patrik D'haeseleer* (PDHaeseleer@lbl.gov), John Gladden, Joshua Park, Alyssa Redding, Chris Petzold, Martin Allgaier, Dylan Chivian, Steve Singer, Terry Hazen, and Blake Simmons

16 **Synthetic Biology Design Automation Tool Suite** 10
T.S. Ham, Z. Dmytriv, H.A. Plahar, J. Chen, D. Densmore, R.D. Rosengarten, **N. Hillson*** (njhillson@lbl.gov), N. Elsbree, G. Linshiz, and J.D. Keasling

17 **Transferring Ionic Liquid Tolerance from the Rain Forest to *E. coli*** 10
Thomas Rüegg, Patrik D'haeseleer, Sharon Borglin, Kristen DeAngelis, Hannah Woo, Erika Lindquist, Jane Khudyakov, Blake Simmons, and **Michael Thelen*** (MPThelen@lbl.gov)

18 **Engineering *E. coli* to Convert Plant Biomass Into Fuels** 11
Greg Bokinski* (GEBokinski@lbl.gov), Pamela Peralta-Yahya, Anthe George, Bradley M. Holmes, Eric Steen, Jeffrey Dietrich, Taek Soon Lee, Danielle Tullman-Ercek, Christopher Voigt, Blake A. Simmons, and **Jay Keasling**

19 **Engineering *Escherichia coli* for Improved Production of FA and FAEE** 11
Fuzhong Zhang* (fuzhongzhang@lbl.gov), Pouya Javidpour, Suzanne Ma, Greg Friedland, Dylan Chivian, Adam Arkin, Swapnil Chhabra, and **Jay D. Keasling**

20 **Metabolic Engineering of Mevalonate Pathway in *E. coli* for Isoprenoid Fuel Production** 12
Taek Soon Lee* (TSLee@lbl.gov), Kevin W. George, Rafael Rosengarten, Nathan Hillson, David Garcia, Suzanne Ma, Swapnil Chhabra, Alyssa Redding-Johanson, Tanveer Batth, Christopher Petzold, Paul Adams, and Jay D. Keasling

Poster **Page**

21 **Rational Engineering of Xylose Co-Utilization in Yeast for Advanced Biofuels Production** 12
Mario Ouellet, Heather Szmidt, Paul D. Adams, Jay D. Keasling, and **Aindrila Mukhopadhyay*** (amukhopadhyay@lbl.gov)

22 **Engineering of Bacterial Methyl Ketone Synthesis for Biofuels** 13
Ee-Been Goh* (egoh@lbl.gov), Jay D. Keasling, and **Harry R. Beller**

23 **Structure of a Three-Domain Sesquiterpene Synthase: A Prospective Target for Advanced Biofuels Production** 13
Pamela P. Peralta-Yahya, Ryan McAndrew* (RPMcandrew@lbl.gov), Andy DeGiovanni, Jose H. Pereira, Masood Z. Hadi, Jay D. Keasling, and **Paul D. Adams**

24 **A Two-Scale C-Based Method for Metabolic Flux Measurement and Prediction** 13
Vinay Satish Kumar, Daniel Weaver, Adam Arkin, and **Hector Garcia Martin*** (hgmartin@lbl.gov)

25 **Mutually Consistent Metabolic Flux and Metabolite Concentration Prediction** 14
Vinay Satish Kumar* (VSKumar@lbl.gov), Hector Garcia Martin, Dylan Chivian, and **Adam P. Arkin**

Great Lakes Bioenergy Research Center (GLBRC)

26 **Development of Crucial Tools for Lignin Research** 14
Yimin Zhu* (zhu6@wisc.edu), Allison Mohammadi, Timothy Pearson, Ruili Gao, Yuki Tobimatsu, Matthew Regner, Fachuang Lu, Hoon Kim, **John Ralph**, Michael Hahn, Fang Chen, Richard A. Dixon, and Steve Decker

27 **Identification of Enzymes that Produce Acylated Monolignols: Progress in the Pursuit of Zip-Lignin™** 15
Saunia Withers, Yimin Zhu* (zhu6@wisc.edu), Fachuang Lu, Hoon Kim, Ji-Young Park, Shawn Mansfield, **Curtis Wilkerson**, and **John Ralph**

28 **Characterization of the *Arabidopsis* Digestibility Mutant *4 COUMARATE CO-A LIGASE 1*** 15
Carl-Erik Tornqvist, Tanya Falbel, Frances Goglio, Nick Santoro, Cliff Foster, and **Sara E. Patterson*** (spatters@wisc.edu)

29 **Identification and Characterization of Direct Regulators of Secondary Wall Cellulose Synthases (CESAs)** 15
Won-Chan Kim* (wonchan@msu.edu), Jae-Heung Ko, Joo-Yeol Kim, and **Kyung-Hwan Han**

30 **Metabolic and Transcriptional Changes During Induced Senescence in Maize** 16
Rajandeep Sekhon* (rsekhon@glbrc.wisc.edu), Kevin Childs, Nicholas Santoro, Cliff Foster, Robin Buell, Natalia de Leon, and **Shawn Kaeppler**

31 **Systems-Level Discovery and Characterization of Cellulolytic Enzymes from the Wood Wasp Symbiont *Streptomyces*** 16
Taichi E. Takasuka* (ttakasuka@glbrc.wisc.edu), Adam J. Book, **Cameron R. Currie**, and **Brian G. Fox**

32 **Identification of Novel Biomass Deconstructing Enzymes via Phylogenetic and Structural Analyses** 17
Adam J. Book* (ajbook@wisc.edu), Ragothaman M. Yennamalli, Taichi E. Takasuka, Gina R. Lewin, Brian G. Fox, George N. Phillips Jr., and **Cameron R. Currie**

33 **Strategies to Enable the Assembly of Soil Metagenomes** 17
Adina Howe* (howead@msu.edu), Jason Pell, Arend Hintze, James M. Tiedje, and **C. Titus Brown**

34 **Developing a Toolbox for Soil Metagenomes** 18
Adina Howe* (howead@msu.edu), Erick Cardenas, Jordan Fish, Jiarong Guo, Aaron Garoutte, Jason Pell, Arend Hintze, James Cole, Patrick Chain, Susannah Tringe, Janet Jansson, C. Titus Brown, and **James M. Tiedje**

Poster **Page**

35 **Microbial Communities in Agricultural Soils have the Potential to Increase Atmospheric Concentrations of Greenhouse Gases** 19
Tracy K. Teal* (tkteal@msu.edu), Vicente Gomez-Alvarez, and **Thomas M. Schmidt**

36 **Microarray Design for Linking Agroecosystem and Gene Function in Soil Microbial Communities** 20
David Duncan* (dsduncan@wisc.edu), Richard LeDuc, James Tiedje, and **Randall Jackson**

37 **A Genomic Comparison of Two Gram Positive Aerobes Reveals Different Strategies for Carbohydrate Degradation** 21
David Mead* (dmead@lucigen.com), Julie Boyum, Lynne Goodwin, and Philip Brumm

38 **Comparative Multiomic Analysis of an *Escherichia coli* Ethanologen During Fermentation of Lignocellulosic Hydrolysates and Synthetic Hydrolysate Reveals Effects of Multiple Inhibitors and Stress Responses on Ethanol Production** 22
Yaoping Zhang* (yzhang@glbrc.wisc.edu), Michael S. Schwalbach, David H. Keating, Mary Tremaine, William Bothfeld, Jeffrey A. Grass, Cameron Cotten, Jennifer L. Reed, Leonardo da Costa Sousa, Mingjie Jin, Venkatesh Balan, James Ellinger, Bruce Dale, Patricia J. Kiley, and **Robert Landick**

39 **Transplantation of *Cellvibrio japonicus* Biomass Deconstruction System into *Escherichia coli*** 22
Rembrandt J.F. Haft* (rhaft@wisc.edu), Jeffrey G. Gardner, David H. Keating, and **Robert Landick**

40 **Improved Enzyme Mixtures for Biomass Deconstruction** 23
Dina Jabbour* (jabbour@msu.edu), John S. Scott-Craig, Goutami Banerjee, and **Jonathan D. Walton**

41 **Three-Dimensional Structures of Enzymes Involved in Cellulose Degradation** 24
Kate E. Helmich, Goutami Banerjee, Christopher M. Bianchetti, Nathaniel L. Elsen, John S. Scott-Craig, Robert W. Smith, Thomas J. Rutoski, Brian G. Fox, Jonathan D. Walton, and **George N. Phillips Jr.*** (phillips@biochem.wisc.edu)

42 **Phenotypic and Genomic Characterization of Natural Isolates of *Saccharomyces cerevisiae* for Growth Tolerance in Lignocellulosic Hydrolysates** 24
Trey K. Sato* (tksato@glbrc.wisc.edu), Jason Shao, Dana J. Wohlbach, Li Hinchman, James McCurdy, Enhai Xie, Adam Halstead, Wendy Schackwitz, Christa Pennacchio, Bruce E. Dale, Blake A. Simmons, David B. Hodge, Venkatesh Balan, Kyria Boundy-Mills, **Yury Bukhman**, and **Audrey P. Gasch**

43 **Co-Fermentation of Glucose, Xylose, and Cellobiose by the Beetle-Associated Yeast, *Spathaspora passalidarum*** 25
Tanya M. Long, Yi-Kai Su, Alan Higbee, Jennifer Headman, Laura B. Willis, and **T.W. Jeffries*** (twjeffri@wisc.edu)

44 **Metabolic Engineering of Bacteria for Sustainable Production of Fatty Acid Derived Products** 26
Rebecca M. Lennen, J. Tyler Youngquist, Daniel Mendez-Perez, Matthew B. Begemann, Spencer W. Hoover, Daniel E. Agnew, and **Brian F. Pfleger*** (pfleger@engr.wisc.edu)

45 **Constraint-Based Analysis of Microbial Metabolism and Regulation for Improving Biofuel Production** 26
Cameron Cotten* (ccotten@wisc.edu), Joonhoon Kim, Christos Maravelias, and **Jennifer L. Reed**

46 **Enhancing Energy Density in Crop Biomass by Redirecting Photosynthate into Triacylglycerols in Vegetative Tissues** 27
Sanjaya* (sanjaya@msu.edu), Timothy Durrett, Mike Pollard, John Ohlrogge, Yair Shachar-Hill, and **Christoph Benning**

47 **Improving the Sustainability of Bioenergy Crops Through Arbuscular Mycorrhization** 28
Arijit Mukherjee* (amukherjee@wisc.edu), Rajandeep Sekhon, Shawn Kaeppler, and **Jean-Michel Ané**

Poster **Page**

48 **Unifying Physics and Biology to Simulate Microbial Denitrification and Nitrous Oxide Flux with the EPIC Model** 28
R.C. Izaurralde* (cesar.izaurralde@pnnl.gov), W.B. McGill, J.R. Williams, R.P. Link, D.H. Manowitz, and D.E. Schwab

49 **Lipid Production in *Rhodobacter sphaeroides*: From Photosynthetic Membranes to Biofuels** 29
Kimberly C. Lemmer* (kchristensen@glbrc.wisc.edu) and **Timothy J. Donohue**

50 **Using Genomic Investigations of the Photosynthetic Bacterium *Rhodobacter sphaeroides* to Determine the Origin of the Increased H_2 Production Capacity of a Mutant Strain** 29
Wayne S. Kontur* (wskontur@wisc.edu), Wendy Shackwitz, Joel Martin, Shweta Deshpande, Hope Tice, Kurt LaButti, Natalia Ivanova, **Daniel R. Noguera** (noguera@engr.wisc.edu), and **Timothy J. Donohue** (tdonohue@bact.wisc.edu)

BioEnergy Science Center (BESC)

51 **BioEnergy Science Center Education and Outreach—Overview** 30
S. Kral, J. Griffin, H. Schulson, S. Fowler, B. Davison,* (davisonbh@ornl.gov), **J. Westpheling*** (janwest@uga.edu), W. Robinson,* and **Paul Gilna** (BESC PI)

52 **BESC Knowledgebase, Tools, and Laboratory Information Management System** 31
Mustafa Syed* (syedmh@ornl.gov), Susan Holladay, Tatiana Karpinets, Morey Parang, Guru Kora, Doug Hyatt, Sheryl Martin, Leslie Galloway, Byung Park, Ying Xu, **Ed Uberbacher**, Brian Davison, and **Paul Gilna** (BESC PI)

53 **Computational Biology to Target Plant Pathways and Experimental Data** 32
Ying Xu* (xyn@bmb.uga.edu), Eberhard Voit, Yanbin Yin, Wen-Chi Chou, Shan Wang, Xizeng Mao, Michael Udvardi, Maor Bar-Peled, Jerry Tuskan, Brian Davison, and **Paul Gilna** (BESC PI)

54 **Enhanced Pathway Visualization in Pathway-Genome Database (PGDB) Captures Subcellular Localization of Metabolites and Enzymes: The Nucleotide-Sugar Biosynthetic Pathways of *Populus trichocarpa*** 33
Ambarish Nag, Tatiana Karpinets, Christopher H. Chang, Maor Bar-Peled,* (mao@ccrc.uga.edu), **Debra Mohnen**, and **Paul Gilna** (BESC PI)

55 **A Workflow for Archiving and Integrating Diverse Data for Structural Characterization of Biomass** 33
William S. York (will@ccrc.uga.edu), Malcolm O'Neill, Trina Saffold, Mehdi Allahyari, Breeanna Urbanowicz* (breeanna@uga.edu), Maria Peña, and **Paul Gilna** (BESC PI)

56 **C Labeling and NMR Analysis: Critical Tools in the Development of Next Generation Biofuel Platforms** 34
Marcus Foston, Reichel Samuels, Rui Katahira, Erica Gjersing, Mark Davis, Heather L. McKenzie, Charles E. Wyman, **Arthur J. Ragauskas*** (Arthur.ragauskas@chemistry.gatech.edu), and **Paul Gilna** (BESC PI)

57 **High- and Medium-Throughput Characterization of Biomass to Understand Causes of Recalcitrance in Plant Cell Walls** 34
Angela Ziebell* (Angela.Ziebell@nrel.gov), Robert Sykes, Steve Decker, Erica Gjersing, Crissa Doeppke, Geoff Turner, Cody Law, Melissa Carlile, **Mark Davis**, and **Paul Gilna** (BESC PI)

58 **High Resolution Characterization of Biomass** 35
Ali Passian* (passianan@ornl.gov), Laurene Tetard, **Brian Davison**, and **Paul Gilna** (BESC PI)

59 **How Do Bacteria Eat Biomass? Multi-scale Imaging and Modeling of Cells, Cell Walls, and Cellulosomes** 36
Michael F. Crowley* (michael.crowley@nrel.gov), Bryon S. Donohoe, Peter N. Ciesielski, Yannick J. Bomble, Loukas Petridis, Jeremy C. Smith, **Michael E. Himmel**, and **Paul Gilna** (BESC PI)

Poster **Page**

60 Real-Time Imaging of Plant Cell Wall Structure at Nanometer Scale, with Respect to Cellulase Accessibility and Degradation Kinetics 36
Yu-San Liu, Yining Zeng, John O. Baker, **Shi-You Ding*** (Shi.you.Ding@nrel.gov), and **Paul Gilna** (BESC PI)

61 A Novel APAP1 Glycoconjugate Indicates a New Role for Pectin in Plant Cell Walls 37
Li Tan* (tan@ccrc.uga.edu), Melani Atamodjo, Stefan Eberhard, Sivakumar Pattathil, Xiang Zhu, Michael G. Hahn, Alan Darvill, Marcia Kieliszewski, **Debra Mohnen**, and **Paul Gilna** (BESC PI)

62 Genome-Scale Discovery of Cell Wall Biosynthesis Genes in *Populus* 38
Steve P. DiFazio* (spdifazio@mail.wvu.edu), Wellington Muchero, G.T. Slavov, J. Martin, W. Schackwitz, E. Rodgers-Melnick, C.P. Pennacchio, U. Hellsten, L. Pennacchio, L.E. Gunter, P. Ranjan, D. Rokhsar, **G.A. Tuskan**, and **Paul Gilna** (BESC PI)

63 Comparative Glycomics Provide Insights into Cell Wall Components that Affect Biomass Recalcitrance .. 38
Michael G. Hahn* (hahn@ccrc.uga.edu), Sivakumar Pattathil, Shishir P.S. Chundawat, Jaclyn D. DeMartini, Jeffrey Miller, Virginia Brown, Sindhu Kandemkavil, Bruce E. Dale, **Charles Wyman**, and **Paul Gilna** (BESC PI)

64 Comparative Proteomics Unveils Functional Signatures of Cellulose Formation (*Populus*) and Deconstruction (Cellulolytic Microorganisms) at a Cellular Level 39
Richard J. Giannone* (giannonerj@ornl.gov), Paul Abraham, Rachel Adams, Andrew Dykstra, **Gerald A. Tuskan**, Robert L. Hettich, and **Paul Gilna** (BESC PI)

65 Development of a High-Throughput Genetic Transformation System for Switchgrass (*Panicum virgatum* L.) 40
Zeng-Yu Wang* (zywang@noble.org), Chunxiang Fu, Hiroshi Hisano, Frank Hardin, Xirong Xiao, **Richard A. Dixon**, and **Paul Gilna** (BESC PI)

66 Cell Wall Characteristics and Ethanol Fermentation Studies of PvMYB4-Over-Expressing Switchgrass Plants 40
Hui Shen* (hshen@noble.org), Charleson R. Poovaiah, Angela Ziebell, Erica Gjersing, Robert Sykes, Fang Chen, Jonathan R. Mielenz, Timothy J. Tschaplinski, Mark Davis, Neal Stewart Jr., **Richard A. Dixon**, and **Paul Gilna** (BESC PI)

67 Consolidated Bioprocessing Conversion of Genetically Modified Switchgrass 41
Kelsey Yee, Choo Hamilton, Miguel Rodriguez Jr., Scott Hamilton-Brehm, Chunxiang Fu, Zeng-Yu Wang, **Jonathan R. Mielenz*** (mielenzjr@ornl.gov), and **Paul Gilna** (BESC PI)

68 Down-Regulation of the Caffeic Acid *O*-methyltransferase Gene in Switchgrass Reveals a Novel Monolignol Analog 42
Timothy J. Tschaplinski* (tschaplinstj@ornl.gov), Robert F. Standaert, Nancy Engle, Madhavi Z. Martin, Amandeep K. Sangha, Jerry M. Parks, Jeremy C. Smith, Reichel Samuel, Yunqiao Pu, Art J. Ragauskas, Choo Y. Hamilton, Chunxiang Fu, Zeng-Yu Wang, Brian H. Davison, Richard A. Dixon, Jonathan R. Mielenz, and **Paul Gilna** (BESC PI)

69 Flowthrough Pretreatment to Characterize Biomass Deconstruction 43
Heather L. McKenzie, **Charles E. Wyman*** (cewyman@engr.ucr.edu), Marcus B. Foston, Jung Seokwon, Arthur Ragauskas, Nancy L. Engle, Joshua F. Emory, Bruce A. Tomkins, Timothy Tschaplinski, Gerald Tuskan, Gary J. Van Berkel, and **Paul Gilna** (BESC PI)

70 Plant Biomass-Degrading Loci Play a Role as Determinants for Lignocellulose Degradation From the Extremely Thermophilic Genus, *Caldicellulosiruptor* 43
Sara E. Blumer-Schuette* (seblumer@ncsu.edu), Richard J. Giannone, Jeffrey V. Zurawski, Inci Ozdemir, Dhaval B. Mistry, Qin Ma, Yanbin Yin, Ying Xu, Farris L. Poole II, Irina Kataeva, Mike W.W. Adams, Scott D. Hamilton-Brehm, James G. Elkins, Frank W. Larimer, Miriam L. Land, Loren Hauser, Robert W. Cottingham, Robert L. Hettich, **Robert M. Kelly**, and **Paul Gilna** (BESC PI)

Poster **Page**

71 **Identifying Potential Detoxification Mechanisms for Furan and Aromatic Aldehydes in Thermophilic, Anaerobic Bacteria Using 2D LC-MS/MS Proteomics** 44
James G. Elkins* (elkinsjg@ornl.gov), Scott D. Hamilton-Brehm, Richard J. Giannone, Nancy L. Engle, Steven D. Brown, Timothy J. Tschaplinski, Robert L. Hettich, and **Paul Gilna** (BESC PI)

72 **Methylation by a Unique α-class N4-Cytosine Methyltransfersase is Required for DNA Transformation of *Caldicellulosiruptor bescii* DSM6725: Use for Construction of Mutants That Affect Biomass Utilization** 45
Daehwan Chung* (chung31@uga.edu), Min Cha, Joel Farkas, Estefania Olivar, **Janet Westpheling**, and **Paul Gilna** (BESC PI)

73 **Improving Ethanol Production in *Clostridium thermocellum*** 45
Daniel G. Olson* (daniel.g.olson@dartmouth.edu), Adam Guss, Yu Deng, Jon Lo, Aaron Argyros, Chris Herring, A. Joe Shaw, Douwe van der Veen, Ranjita Biswas, **Lee R. Lynd**, and **Paul Gilna** (BESC PI)

74 **Thermophilic Isobutanol Production in *Geobacillus thermoglucosidasius*** 45
Paul P. Lin, Jennifer Takasumi, Elizabeth Felnagle, **James C. Liao*** (liaoj@seas.ucla.edu), and **Paul Gilna** (BESC PI)

75 **Engineering Improved Cellulases for CBP Microbes** 46
Qi Xu, Mo Chen, Michael G. Resch, Markus Alahuhta, Roman Brunecky, Vladimir V. Lunin, John W. Brady, Michael F. Crowley, Yannick J. Bomble* (yannick.bomble@nrel.gov), **Michael E. Himmel**, and **Paul Gilna** (BESC PI)

Biofuels Research 47

76 **Phylogenetic Distribution of Potential Cellulases in Bacteria** 47
Renaud Berlemont* (rberlemo@uci.edu) and **Adam Martiny**

77 **Engineering Robust Hosts for Microbial Biofuel Production** 47
Mary Harrison and **Mary J. Dunlop*** (mjdunlop@uvm.edu)

78 **Engineering Ethylmalonyl-CoA Pathway in *Methylobacterium extorquens* AM1 for Butanol Production: Identification of a Regulator Activating the Expression of a Key Gene** 48
Bo Hu* (bohu@uw.edu), **David Beck**, **Robert Synovec**, and **Mary Lidstrom**

79 **Integration of Carbon, Nitrogen, and Oxygen Metabolism in *Escherichia coli*** 49
Yi-Fan Xu* (yifanxu@Princeton.EDU), Christopher D. Doucette, Junyoung O. Park, Nathaniel Roquet, Daniel Amador-Noguez, Marshall L. Reaves, Xiao-jiang Feng, Herschel Rabitz, David J. Schwab, Ned S. Wingreen, and **Joshua D. Rabinowitz**

80 **Metabolomics of Clostridial Biofuel Production** 50
Daniel Amador-Noguez, Ludmilla Aristilde, Caleb Bastian, Herschel Rabitz, Leonid Kruglyak, and **Joshua D. Rabinowitz,*** (joshr@genomics.princeton.edu)

81 **Experimental Systems Biology Approaches for Clostridia-Based Bioenergy Production: The Metabolite Stress-Response System in Solventogenic Clostridia** 50
E. Terry Papoutsakis* (epaps@udel.edu), Blake C. Meyers, Kelvin H. Lee, Maciek R. Antoniewicz, Costas D. Maranas, H. Huang, and **Cathy H. Wu**

82 **Population Level Analysis of Mutations Underlying Improvements in Biofuel Production by *Clostridium phytofermentans*** 51
Supratim Mukherjee, Stephen Godin, and **Jeffrey L. Blanchard*** (blanchard@microbio.umass.edu)

Poster **Page**

83 **Understanding Fundamental Aspects of Butanol Production by *Clostridium beijerinckii*** 52
Caroline B. Milne, Yi Wang, Beth Papanek, Yong-Su Jin, Nathan D. Price, and **Hans P. Blaschek*** (blaschek@illinois.edu)

84 **Applying the Biology of Brown Rot Fungi to Consolidated Bioprocessing—Early Career Program** 52
Robert A. Blanchette, Shona M. Duncan, Timothy Filley, Joel Jurgens, Justin Kaffenberger* (kaffe002@umn.edu), and **Jonathan S. Schilling**

85 **Development of Quantum Dot Probes for Studies of Synergy Between Components of the Wood-Degrading Fungal Enzymes** 53
Haw Yang, Kevin Welsher, Chih-Hao Hsia, Daniel Montiel, Robin Rohwer* (rrr24@psu.edu), **B. Tracy Nixon*** (btn1@psu.edu), and **Ming Tien**

86 **Biodegradative Oxidant Production by Fungi in Lignocellulose: In Situ Mapping and Gene Expression** 54
Premsagar Korripally, Peter Kitin, Christopher G. Hunt, Carl J. Houtman, **John Ralph**, and **Kenneth E. Hammel*** (kehammel@wisc.edu)

87 **Multi-Species, Multi-Gene Co-Regulation: Finding DNA *Cis*-Regulatory Elements for Biofuel Pathways**55
Lee Ann McCue, **Lee A. Newberg**,* (lee.newberg@wadsworth.org), William A. Thompson, and **Charles E. Lawrence**

88 **Optimizing Plant-Microbial Systems for Energy—Mapping Feedstock Quality Genes in *Brachypodium distachyon*** 56
Scott J. Lee and **Samuel P. Hazen*** (hazen@bio.umass.edu)

89 **A Systems Biology, Whole-Genome Association Analysis of the Molecular Regulation of Biomass Growth and Composition in *Populus deltoides*** 56
Leandro Gomide Neves* (gomide@ufl.edu) and **Matias Kirst**

90 **Next Generation Protein Interactomes for Plant Systems Biology and Biomass Feedstocks Research** 57
Mary Galli, Shelly Wanamaker, and **Joseph R. Ecker*** (ecker@salk.edu)

91 **Modeling of Cellulose, Hemicellulose and Lignin-Carbohydrate Complex Formation and Regulation to Understand Plant Cell Wall Structure** 58
Vincent Chiang* (vchiang@ncsu.edu), Ron Sederoff, David Muddiman, Cranos Williams, Fikret Isik, Joel Ducoste, Chris Smith, John Ralph, and Sunkyu Park

92 **Regulation of Neutral Lipid Accumulation in Vegetative Tissues of Plants** 59
Kent D. Chapman* (chapman@unt.edu), John M. Dyer, and Robert T. Mullen

93 **Investigations of Clusters of Cellobiose-Acid-(H_2O)$_n$** 60
Madeleine Pincu* (mpincu@uci.edu) and **Robert B. Gerber**

94 **Production and NMR Analysis of Deuterated Cellulose and Lignocellulosic Biomass and Its Utilization for Neutron Scattering Studies** 60
Barbara R. Evans* (evansb@ornl.gov), Marcus Foston, Art Ragauskas, Hugh M. O'Neill, Juhong He, Sai Venkatesh Pingali, Volker Urban, Paul A. Langan, and **Brian H. Davison**

95 ***In situ* Small-Angle Neutron Scattering and Computer Simulation Investigate Lignin Aggregation During Biomass Pretreatment** 61
Loukas Petridis* (petridisl@ornl.gov), Sai Venkatesh Pingali, Volker Urban, William T. Heller, Hugh M. O'Neill, Marcus Foston, Art Ragauskas, Barbara R. Evans, Paul A. Langan, Jeremy C. Smith, and **Brian H. Davison**

Poster **Page**

96 **Study of Plant Cell Wall Polymers Affected by Metal Accumulation Using Stimulated Raman Scattering Microscopy** 62
Shi-You Ding* (Co-PI: Shi.you.Ding@nrel.gov), and **X. Sunney Xie** (PI: xie@chemistry.harvard.edu)

97 **Defining Determinants and Dynamics of Cellulose Microfibril Biosynthesis, Assembly, and Degradation** 63
Jocelyn Rose,* (jr286@cornell.edu) Thomas Brutnell, Seth DeBolt, Detlef Smilgies, and Larry Walker

98 **Reduced Lignification Enhances Cellulose Accessibility in Wood Cell Walls** 63
Catherine I. Lacayo, Shi-You Ding, Mona Hwang, and **Michael P. Thelen*** (mthelen@llnl.gov)

99 **Live Cell Identification of Redox Regulated Dithiol Sensors by Chemical Probes** 64
Aaron T. Wright* (aaron.wright@pnnl.gov), Natalie C. Sadler, Na Fu, Eric A. Hill, Matthew R. Melnicki, Thomas C. Squier, Allan E. Konopka, Diana J. Bigelow, Erika M. Zink, Michael P. Lewis, Suchitra Datta, Dian Su, Wei-Jun Qian, Don A. Bryant, and **Alexander S. Beliaev**

100 **Effects of Nutrient Limitation, Light Quantity and Quality on Growth Dynamics and Physiology of *Synechococcus* sp. PCC 7002** 65
Allan E. Konopka* (allan.konopka@pnnl.gov), Grigoriy E. Pinchuk, Eric A. Hill, Leo A. Kucek, Matthew R. Melnicki, Trang Vu, Jennifer L. Reed, and **Alex S. Beliaev**

Biological Hydrogen Production........ 67

101 **Characterization of H_2 Producing Microbial Communities of the Wood-Feeding Beetle *Odontotaenius disjunctus* by Multi-Scale Measurement of Metabolic Function** 67
Javier A. Ceja-Navarro* (JCNavarro@lbl.gov), Xavier Mayali, Peter K. Weber, Mari Nyyssönen, Zhao Hao, Hoi-Ying Holman, Meredith M. Blackwell, **Jennifer Pett-Ridge**, and **Eoin L. Brodie**

102 **Anoxic Carbon and Nitrogen Metabolism in Photosynthetic Microbial Mats as Revealed by Metatranscriptomic and NanoSIMS Studies** 68
Steven W. Singer* (SWSinger@lbl.gov), Luke C. Burow, Dagmar Woebken, Ian P.G. Marshall, Brad M. Bebout, Leslie Prufert-Bebout, Jennifer Pett-Ridge, Tori Hoehler, Alfred M. Spormann, and **Peter K. Weber**

103 **Development and Validation of Novel Techniques to Understand and Address Limitations to Algal Hydrogen and Biofuels Photoproduction** 68
Alexandra Dubini (alexandra.dubini@nrel.gov), Matthew Wecker (matt.wecker@nrel.gov), and **Maria L. Ghirardi*** (maria.ghirardi@nrel.gov)

104 **Development of *Cyanothece* as a Model Organism for Photobiological Hydrogen Production** 69
Anindita Bandyopadhyay, Jana Stöckel, Xueyang Feng, Yinjie Tang, Sujata Mishra, Xiaohui Zhang, Louis A. Sherman, Uma K. Aryal* (Uma.Aryal@pnnl.gov), Stephen J. Callister, Janani I. Shutthanandan, Carrie D. Nicora, Ljiljana Pasa-Tolic, Dave W. Koppenaal, Richard D. Smith, and **Himadri B. Pakrasi**

105 **Contrasting the Metabolic Capabilities of *Cyanothece* 51142 and *Synechocystis* 6803** 70
Rajib Saha* (rus184@psu.edu), Bertram M. Berla, Thomas J. Mueller, Thanura Elvitigala, Lawrence E. Page, Louis Sherman, **Himadri B. Pakrasi**, and Costas D. Maranas

106 **Photophysiology and Energetics of Light-Driven H_2 Production by *Cyanothece sp.* ATCC 51142** 71
Matthew R. Melnicki* (matthew.melnicki@pnnl.gov), Eric A. Hill, Leo A. Kucek, Grigoriy E. Pinchuk, Allan E. Konopka, and **Alex S. Beliaev**

107 **Genomic, Genetic, and Cytological Analysis Reveals Biased Inheritance of PatN, a Negative Regulator of Heterocyst Differentiation in *Nostoc punctiforme*** 72
Douglas Risser, Harry Christman, Elsie Campbell, Brett Phinney, Wan-Ling Chiu, and **John C. Meeks*** (jcmeeks@ucdavis.edu).

* Presenting author

Poster **Page**

108 **Biohydrogenesis in the Thermotogales** 73

J.S. Notey* (jsnotey@ncsu.edu), A.D. Frock, M. El-Sayed, A. Petrus, S. Swithers, N. Boucher, D. White, R. Singh, **P.H. Blum**, **K.M. Noll**, and **R.M. Kelly**

109 **Mutational Analysis of the Enzymes Involved in the Metabolism of Hydrogen by *Pyrococcus furiosus*** 75

Gerrit J. Schut* (gerti@uga.edu), William J. Nixon, Gina L. Lipscomb, **Michael W.W. Adams**, and **Robert A. Scott**

110 **Pathway of Fermentative H_2 Production by Sulfate-Reducing Bacteria** 75

Kimberly L. Keller* (kellerkl@missouri.edu), Steve D. Brown, Barbara J. Rapp-Giles, Iris Porat, Adam Deutschbauer, Jennifer Kuehl, Timothy J. Tschaplinski, and **Judy D. Wall**

111 **Pathways and Regulatory Network of Hydrogen Production from Cellulose by *Clostridium thermocellum*** 76

Ryan Severins, Jonathan Millen, Katherine Chou, Ji Hye Jo, Yuan Zhang, Wendy D.M. Wang, Robert J. Zagursky, Pin-Ching Maness, Yi Qu, Mary Lipton, and **J.H. David Wu*** (davidwu@che.rochester.edu)

262A **The Ins and Outs of Algal Metal Transport** 77

Crysten E. Blaby-Haas and **Sabeeha S. Merchant*** (merchant@chem.ucla.edu)

Biological Systems Research on the Role of Microbial Communities in Carbon Cycling 79

112 **Approaches to Systematically Examine Carbon Flux in Microbial Communities Using 'Omics' and Stable Isotope Probing** 79

Annika Mosier, Nicholas Justice, Daniela Aliaga Goltsman, Christopher Miller, Chongle Pan, Ben Bowen, Brian Thomas, Steven Singer, Trent Northen, Robert Hettich, and **Jill Banfield*** (jbanfield@berkeley.edu)

113 **Estimation of Hydrogen Isotope Enrichment in Proteins and Lipids in a Microbial Community** 80

Ben Bowen* (BPBowen@lbl.gov), Curt Fischer, Chongle Pan, Bob Hettich, Trent Northen, and **Jillian F. Banfield**

114 **Microbial Food Web Mapping: Linking Carbon Cycling and Community Structure in Soils Through Pyrosequencing-Enabled Stable Isotope Probing** 80

Nathalie J.A. Curlevski,* Ashley N. Campbell, and **Daniel H. Buckley** (dbuckley@cornell.edu)

115 **From Community Structure to Function: Metagenomics-Enabled Predictive Understanding of Microbial Communities to Climate Warming at the Temperate Grassland Ecosystems in Oklahoma** .81

Jizhong Zhou* (jzhou@ou.edu), Liyou Wu, Kai Xue, Lei Cheng, Mengting Yuan, Jin Zhang, Ye Deng, Joy D. Van Nostrand, Zhili He, Ryan Penton, Jim Cole, **James M. Tiedje**, Rosvel Bracho-Garrillo, **Edward A.G. Schuur**, Chengwei Luo, **Konstantinos Konstantinidis**, Xia Xu, Dejun Li, and **Yiqi Luo**

116 **From Community Structure to Functions: Metagenomics-Enabled Predictive Understanding of Microbial Communities to Climate Warming in Alaskan Tundra** 83

Edward A.G. Schuur* (tschuur@ufl.edu), **Jizhong Zhou**, Liyou Wu, Kai Xue, Lei Cheng, Mengting Yuan, Jin Zhang, Ye Deng, Joy D. Van Nostrand, **Zhili He**, Ryan Penton, Jim Cole, **James M. Tiedje**, Rosvel Bracho-Garrillo, Chengwei Luo, **Konstantinos Konstantinidis**, Xia Xu, Dejun Li, and **Yiqi Luo**

117 **Microbial Communities Generating Greenhouse Gases in Thawing Permafrost** 84

Rhiannon Mondav, Suzanne B. Hodgkins, Virginia I. Rich, Carmody K. McCalley, Tyler Logan, Rick Wehr, Changsheng Li, Steve Frolking, Jeffrey Chanton, Patrick M. Crill, Scott R. Saleska, and **Gene W. Tyson*** (g.tyson@uq.edu.au)

Poster **Page**

118 **The Impact of Global Warming on the Carbon Cycle of Arctic Permafrost: An Experimental and Field Based Study** 85
T. Vishnivetskaya, N. Mykytczuk, J. Allan, K. Cheng, K. Chourey, R. Hettich, A. Layton, X. Liu, J. Murphy, T.J. Phelps, S. Pfiffner, G. Saarunya, B. Stackhouse, C.Y.M. Lau, R. Sanders, S. Myneni, L. Whyte, R. Wilhem, W. Pollard, G. Lamarche-Gagnon, M. Kidd, S. Elliott, P. Hatcher, R.L. Sleighter, P. Bennett, L. Ziolkowski, and **T.C. Onstott*** (tullis@princeton.edu)

119 **IsoGenie: Microbes, Ecosystem Biogeochemistry, and Climate Change: Dissecting Methane Flux in a Thawing Permafrost Peatland in Northern Sweden** 86
Carmody K. McCalley,* Suzanne B. Hodgkins,* Rhiannon Mondav, Tyler Logan, Rick Wehr, David D. Nelson, J. Barry McManus, Mark S. Zahniser, **Changsheng Li**, **Steve Frolking**, **Jeffrey Chanton**, **Virginia Rich**, **Patrick M. Crill**, **Gene Tyson**, and **Scott R. Saleska** (saleska@email.arizona.edu)

120 **The Impact of Precipitation and Nitrogen Deposition Changes on the Microbial Community and Greenhouse Gas Cycles of a Southern California Grassland** 87
Emma L. Aronson* (earonson@uci.edu) and **Steven D. Allison**

121 **Reconciling Phylogeny and Function During Plant Litter Decomposition by High-Throughput Functional Metagenomics** 88
Mari Nyyssönen* (mjnyyssonen@lbl.gov), Claudia Weihe, Michael L. Goulden, Kathleen K. Treseder, Jennifer B.H. Martiny, Adam C. Martiny, **Steven D. Allison**, and Eoin L. Brodie

122 **Atmospheric Nitrogen Deposition and Microbial Mechanisms Enhancing Soil Carbon Storage** 89
Sarah D. Eisenlord* (seisen@umich.edu), Ivan P. Edwards, Elizabeth E. Entwistle, **Kurt S. Pregitzer**, **Cheryl R. Kuske**, and **Donald R. Zak**

123 **Microbial Response to Modified Precipitation Patterns in Tallgrass Prairie Soil: Molecular Mechanisms, Activity Rates, and Organic Matter Dynamics** 90
Lydia H. Zeglin* (Lydia.zeglin@oregonstate.edu), Maude M. David, Peter J. Bottomley, Robert L. Hettich, Janet K. Jansson, Ari Jumpponen, Charles W. Rice, Susannah G. Tringe, Nathan C. VerBerkmoes, and **David D. Myrold**

124 **Microbial Communities in Restored Wetland Sediments** 91
Shaomei He* (she@lbl.gov), Mark Waldrop, Lisamarie Windham-Myers, Stephanie Malfatti, Tijana Glavina del Rio, and **Susannah G. Tringe**

125 **Great Prairie Grand Challenge Soil Metagenome Sequencing Project** 92
Janet K. Jansson* (jrjansson@lbl.gov), Titus Brown, Patrick Chain, James Cole, Maude David, Jordan Fish, Adina Howe, Regina Lamendella, Kostas Mavromattis, Susannah Tringe, and James M. Tiedje

126 **Development of Integrated "Omics" Approach for Assessing Microbial Cycling of Carbon in Prairie Soil Using a Model Soil Bacterium: *Arthrobacter chlorophenolicus*** 92
Maude M. David***** (MMDavid@lbl.gov), Lydia Zeglin, Peter Bottomley, Robert Hettich, Kristen Corrier, Ari Jumpponen, Konstantinos Mavromatis, Charles Rice, Manesh Shah, Susannah Tringe, Nathan VerBerkmoes, David Myrold, and **Janet K. Jansson**

127 **Microbial Mediation of Litter Decomposition in Soil: The Role of Plant Roots** 93
Shengjing Shi* (shengjing.shi@berkeley.edu), Donald Herman, Erin Nuccio, Romain Barnard, Jeremy Bougoure, Jennifer Pett-Ridge, Eoin Brodie, Trent Northen, Zhili He, Jizhong Zhou, and **Mary Firestone**

128 **Linking Microbial Identity and Succession Patterns to Uptake of Plant-Derived Carbon in the Rhizosphere** 94
Jeremy Bougoure* (bougoure1@llnl.gov), Erin Nuccio, Xavier Mayali, Shengjing Shi, Donald Herman, Zhili He, Jizhong Zhou, Ulisses Nunes da Rocha, Eoin Brodie, Jennifer Pett-Ridge, and **Mary Firestone**

Poster **Page**

129 **Soil Fungal Community Responses to Long-Term Elevated CO_2 and N Deposition Conditions in a Temperate Pine Forest (DOE Duke Forest FACE site)** 95
Carolyn F. Weber, Andrea Porras-Alfaro, Kuan-Liang Liu, Gang Xie, Monica Moya Balasch, La Verne Gallegos-Graves, John Dunbar, Rytas Vilgalys, and **Cheryl R. Kuske*** (kuske@lanl.gov)

130 **Arid Land Ecosystem Responses to Long-Term Elevated Atmospheric CO_2 in a Large Manipulated Field Experiment** 96
Blaire Steven, La Verne Gallegos-Graves, Jayne Belnap, R. David Evans, John Dunbar, and **Cheryl R. Kuske*** (kuske@lanl.gov)

131 **Direct Interspecies Electron Transfer in Methanogenic Environments: Genome-Scale Analysis of Mechanisms in Defined Co-Cultures and Natural Aggregates** 96
Amelia-Elena Rotaru* (arotaru@microbio.umass.edu), Pravin Mala Shresta, Fanghua Liu, Pier-Luc Tremblay, Jessica Smith, Dan Carlo Flores, Mollie Murnane, Harish Nagarajan, Karsten Zengler, and **Derek R. Lovley**

132 **Integrated Genome-Scale Modeling of Syntrophic Consortia Reveals Microbial Community Dynamics and Mechanisms of Electron Transfer** 98
Harish Nagarajan* (nh@ucsd.edu), Mallory Embree, Adam M. Feist, Amelia Elena Rotaru, Pravin Shrestha, Bernhard Ø. Palsson, **Derek R. Lovley**, and **Karsten Zengler**

133 **Systems-Level Insights into Alternate Methane Cycling Modes in a Freshwater Lake via Community Transcriptomics, Metabolomics, and nanoSIMS Analysis** 99
M.G. Kalyuzhnaya, S. Yang, D.A.C. Beck, J. Bickford, M.C. Konopka, L. Chistoserdova, V.J. Orphan, and **M.E. Lidstrom*** (lidstrom@u.washington.edu)

134 **Vitamin Biosynthesis and Regulation in Marine Algae** 99
Darcy McRose,* Shuangchun Yan, Emily Reistetter, and **Alexandra Z. Worden** (azworden@mbari.org)
PI: Alexandra Z. Worden, Monterey Bay Aquarium Research Institute and UCSC
Co-PIs: Stephen Callister, Richard Smith, Joshua Stuart

135 **Computational and Experimental Approaches to Systems Biology of Marine Green Algae** 100
Marijke J. van Baren*, Peter Waltman, Daniel Sam, Emily Reistetter, Samuel Purvine, Charles Ansong, Stephen Callister, Joshua Stuart, and **Alexandra Z. Worden** (azworden@mbari.org)
PI: Alexandra Z. Worden, Monterey Bay Aquarium Research Institute and UCSC
Co-PIs: Stephen Callister, Richard Smith, Joshua Stuart

136 **Linking Phylogeny and Function of Methanotrophic Archaeal-Bacterial Consortia in Deep-Sea Methane Seeps Using *in situ* Targeted Metagenomics, Molecular Ecology, and Stable Isotope Tracer Experiments** 101
Joshua A. Steele* (jsteele@gps.caltech.edu), Rachel S. Poretsky, Anne E. Dekas, Shulei Sun, Abigail Saxena, Patricia L. Tavormina, Mark H. Ellisman, and **Victoria J. Orphan**

137 **Correlative Compositional Imaging and Protein Profiling in Marine Anaerobic Methane Oxidizing Microbial Communities** 102
Shawn E. McGlynn* (smcglynn@caltech.edu), Jeff J. Marlow, Grayson Chadwick, Tomas J. Deerinck, Shulei Sun, Karuna Chourey, Chongle Pan, Robert L. Hettich, Mark H. Ellisman, and **Victoria J. Orphan**

138 **How Monomer Availability Impacts the Scramble for Carbon During Plant Polymer Decomposition: Using Pure Cultures as a Predictive Model System** 103
Shifang Hsu, Nivedita Clark, and **Christopher B. Blackwood*** (cblackwo@kent.edu)

139 **Methanogenic Archaea and the Global Carbon Cycle: A Systems Biology Approach to the Study of *Methanosarcina* Species** 104
Matthew N. Benedict, Matthew C. Gonnerman, Ping Ma, Rachel J. Whitaker, Nathan D. Price* (nprice@systemsbiology.org), and **William W. Metcalf**

Poster **Page**

140 **Identification of S-layer Proteins in the Methanosarcinaceae** 104
R.P. Gunsalus* (robg@microbio.ucla.edu), D. Francoleon, R. Loo, J. Loo, L. Rohlin, U. Kim, and M. Arbing

Systems Biology Strategies and Technologies for Understanding Microbes, Plants, and Communities........ 107

141 **Influence of Habitat on Diversity and Evolution of Surface Ocean Microbes** 107
Andrew E. Allen* (aallen@jcvi.org), Lisa Z. Allen, Jonathan H. Badger, Christopher L. Dupont, John P. McCrow, Ian T. Paulsen, Mathangi Thiagarajan, Doug B. Rusch, Jed Fuhrman, Kenneth H. Nealson, Shibu Yooseph, and **J. Craig Venter**

142 **Influence of Nutrients and Currents on the Genomic Composition of Microbes across an Upwelling Mosaic** 107
Lisa Zeigler Allen, Andrew E. Allen* (aallen@jcvi.org), Eric E. Allen, Jonathan H. Badger, John P. McCrow, Ian T. Paulsen, Liam D.H. Elbourne, Mathangi Thiagarajan, Doug B. Rusch, Kenneth H. Nealson, Shannon J. Williamson, and **J. Craig Venter**

143 **Genomic Insights to SAR86, an Abundant and Uncultivated Marine Bacterial Lineage**........ 108
Chris L. Dupont* (cdupont@jcvi.org), Douglas B. Rusch, Shibu Yooseph, Mary-Jane Lombardo, R. Alexander Richter, Ruben Valas, Mark Novotny, Joyclyn Yee-Greenbaum, Jeremy D. Selengut, Dan H. Haft, Aaron L. Halpern, Roger S. Lasken, Kenneth Nealson, Robert Friedman, and **J. Craig Venter**

144 **Regulation of Nitrogen Metabolism in the Model Marine Diatom *Phaeodactylum tricornutum***........ 108
Christopher L. Dupont* (cdupont@jcvi.org), James McCarthy, Adriano Nunes-Nesi, Sabrina Kleeson, Alisdair Fernie, and **Andrew E. Allen**

145 **Responses of and Interactions Between Nitrifying Bacteria to Environmental Changes: A Systems Level Approach** 110
Peter Bottomley* (peter.bottomley@oregonstate.edu), Jeff Chang, Yannming Di, José Perez, Luis Sayavedra-Soto, and **Daniel Arp**

146 **Regulons, Uptake, and Salvage Pathways that Mediate Metabolite Exchange in Hot Spring Microbial Mats**........ 110
Dmitry A. Rodionov* (rodionov@burnham.org), Frank Collart, Aaron T. Wright, Margie Romine, Jim Fredrickson, and **Andrei L. Osterman**

147 **Methanotrophic-Mediated Metal Binding: Effects on *In Situ* Microbial Community Structure and the Sustainability of Subsurface Water Systems** 112
Jeremy D. Semrau* (jsemrau@umich.edu), Alan A. DiSpirito, and Nathan Bandow

148 **Improved Understanding of Microbial Iron and Sulfate Reduction Through a Combination of Bottom-Up and Top-Down Functional Proteomics Assays**........ 113
Ruth E. Richardson* (rer26@cornell.edu), Annette R. Rowe, Peter Kelly, Hening Lin, Sheng Zhang, Robert Sherwood, Lee Ann McCue, Michael Wilkins, and Stephen J. Callister

149 ***In situ* Spectroscopy on Intact *Leptospirillum ferrooxidans* Reveals that Reduced Cytochrome 579 is an Intermediate in the Iron Respiratory Chain** 114
Robert C. Blake II* (rblake@xula.edu) and Megan N. Griff

150 **DOE JGI Plant Genomics Program**........ 114
Jeremy Schmutz* (jschmutz@hudsonalpha.org), Jarrod Chapman, David Goodstein, Jane Grimwood, Uffe Hellsten, Jerry Jenkins, Gerald A. Tuskan, Jim Bristow, Kerrie Barry, Daniel S. Rokhsar, and **Edward M. Rubin**

Poster **Page**

151 **Genome Improvement of DOE JGI Flagship Plant Genomes** ... 115
David Sims* (dsims@hudsonalpha.org), Jerry Jenkins, Wellington Muchero, Gerald A. Tuskan, Jim Bristow, Kerrie Barry, Daniel S. Rokhsar, Jeremy Schmutz, Jane Grimwood, and **Richard M. Myers**

152 **Fidelity and Dynamics of DNA Methylation in Plants** ... 116
Qin Yao* (qyao@bnl.gov), Changjun Liu, John Shanklin, Chuan He, and **John Dunn**

153 **PET Radiotracer Imaging in Plant Biology** ... 117
Drew Weisenberger* (drew@jlab.org)

154 **Pan-omics Informatics and Advanced Analysis Pipelines** ... 117
Gordon A. Anderson* (gordo@pnnl.gov), Ronald J. Moore, Joshua N. Adkins, Kenneth J. Auberry, Mikhail E. Belov, Kristin E. Burnum, Kevin L. 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 J. Orton, Ljiljana Paša-Tolić, 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 T. Wright, Rui Zhang, Rui Zhao, and **Richard D. Smith** (rds@pnnl.gov)

155 **Proteomics and Pan-omics Measurements for Comprehensive Systems Characterization of Biological Systems** ... 118
Mary S. Lipton* (mary.lipton@pnnl.gov), **Thomas O. Metz*** (thomas.metz@pnl.gov), Kristin E. Burnum, Carrie D. Nicora, Qibin Zhang, Stephen J. Callister, Scott R. Kronewitter, Aaron T. Wright, Ronald J. Moore, Zeping Hu, Xiaoli Gao, Weijun Qian, Erin S. Baker, Therese R.W. Clauss, Karl K. Weitz, Daniel J. Orton, Matthew E. Monroe, Anil K. Shukla, Rui Zhao, Roslyn Brown, Joshua N. Adkins, **Gordon A. Anderson**, and **Richard D. Smith** (rds@pnnl.gov)

156 **Initial Application of a New LC-IMS-TOF MS Platform to Complex Environmental Samples** ... 118
Erin Shammel Baker* (erin.baker@pnnl.gov), Yehia M. Ibrahim, William F. Danielson III, Kristin E. Burnum, Matthew E. Monroe, Kevin L. Crowell, Daniel J. Orton, Gordon W. Slysz, Brian J. Anderson, Carrie D. Nicora, Mary S. Lipton, Gordon A. Anderson, and **Richard D. Smith** (rds@pnnl.gov)

157 **Integrating Transcriptomics and Proteomics Data for Identifying Mechanisms Regulating the Response of Microbes to Environmental Change** ... 120
H. Steven Wiley* (steven.wiley@pnnl.gov), Ronald C. Taylor, Meng L. Markillie, Mary S. Lipton, Gretta Serres, and Margie Romine

158 **Quantitative Proteomic Profiling of Site-Specific Redox Modifications, Proteolytic Processing, and N-glycosylation** ... 121
Dian Su, Jong-Seo Kim, Lu Wang, Uma K. Aryal, Yi Qu, Aaron T. Wright, T.C. Squier, Mary S. Lipton, Richard D. Smith, Ziyu Dai, Scott E. Baker, and **Wei-Jun Qian*** **(**weijun.qian@pnnl.gov)

159 **Optimization of Microbial Community Proteomics** ... 122
W.R. Cannon* (william.cannon@pnl.gov), A. Hugo, D.J. Baxter, A. Kalyanaraman, and B.-J. Webb-Robertson

160 **PUNCS: Towards Predictive Understanding of Nitrogen Cycling in Soils** ... 123
Robert Sanford* (rsanford@illinois.edu), Allana Welsh, Kostas Konstantinidis, Joanne Chee-Sanford, Kirsti Ritalahti, Sukhwan Yoon, and **Frank Löffler**

161 **Spatioelectrochemistry: The Molecular Basis for Electron Flow Within Metal-Reducing Biofilms** ... 124
Daniel R. Bond* (dbond@umn.edu), Ying Liu, **Caleb Levar**, and Janet Rollefson

162 **Metabolic Dynamics of Starvation Induced Dormancy in *Saccharomyces cerevisiae*** ... 125
Niels Klitgord* (njklitgord@lbl.gov), Fredrik Alsin, Theodore Peters, Matt Youngblood, Robert E. Hughes, and **Trent Northen**

Poster **Page**

163 **Engineering Specialized Metabolism in a Single Cell Type** 126
Anna Berim, Adam Schmidt, B. Markus Lange, Michael Scanlon, Carol A. Soderlund, and **David R. Gang*** (gangd@wsu.edu)

164 **Characterization of Metabolites from Phototrophic Microbial Mats in Biogeochemically Unique Environments** 127
Young-Mo Kim* (young-mo.kim@pnnl.gov), Stephen R. Lindemann, David M. Ward, Sergey Stolyar, Thomas O. Metz, **Margaret F. Romine**, and **James K. Fredrickson**

165 **Carbon Flow in Lithotrophic Acid Hot Springs Microbial Communities, Yellowstone National Park** 127
Helen Kreuzer* (Helen.kreuzer@pnnl.gov), James Moran, Ryan Jennings, Laura Whitmore, Mark Kozubal, Bill Inskeep, and **Jim Fredrickson**

166 **Metabolic and Community Modeling of Phototroph-Heterotroph Community Interactions** 129
William Cannon* (william.cannon@pnl.gov), Gretta Serres, Sebastian Jaramillo, and Don Bryant

167 **Metabolic Coupling and Carbon Flux Regulation in Phototroph-Heterotroph Associations** 130
Alex S. Beliaev* (alex.beliaev@pnnl.gov), Thomas C. Squier* (thomas.squier@pnnl.gov), Diana J. Bigelow, Grigoriy E. Pinchuk, Sergey M. Stolyar, Baowei Chen, Na Fu, Yijia Xiong, E.A. Hill, and **Jim K. Fredrickson**

168 **Microbial Diveristy and Biogeochemical Function of the Phototrophic Microbial Mats of Epsomitic Hot Lake, Washington** 131
Stephen R. Lindemann* (stephen.lindemann@pnnl.gov), James J. Moran* (james.moran@pnnl.gov), Alice Dohnalkova, David W. Kennedy, Sergey M. Stolyar, Paul D. Majors, H. Steven Wiley, Allan E. Konopka, and **James K. Fredrickson**

169 **Microbial Interactions in Hot Spring Phototrophic Mats and Isolates** 133
David M. Ward* (umbdw@montana.edu), Christian Klatt, Donald A. Bryant, Zhenfeng Liu, and Sergey Stolyar

170 **Predicting Autotroph—Heterotroph Metabolic Interactions from Genome Sequence** 134
Margrethe (Gretta) Serres* (mserres@mbl.edu), Elizabeth Landorf, Frank Collart, Sebastian Jaramillo Riveri, and **Margie Romine**

171 **Chemical Profiling of Group B Vitamin Transport and Protein Interactions in Microbes and Microbial Communities** 135
Aaron T. Wright* (Aaron.Wright@pnnl.gov), Phillip K. Koech, Andrei L. Osterman, Dmitry A. Rodionov, and **Margie Romine**

172 **Proteomic Characterization of Microbes and Microbial Communities** 136
Mary S. Lipton* (mary.lipton@pnnl.gov), Laurey Steinke, Roslyn Brown, Ronald J. Moore, Therese R.W. Clauss, Gordon Slysz, Gordon A. Anderson, David M. Ward,* Christian Klatt, Donald A. Bryant, William P. Inskeep, Mark Kozubal, Zack Jay, Jacob Beam, and Ryan Jennings

173 **Integrating Geochemical, Microscopic, and –Omics Analyses to Understand Microbial Interactions in High-Temperature Chemotrophic Communities** 137
William P. Inskeep* (binskeep@montana.edu), Mark Kozubal, Zack Jay, Jacob Beam, Ryan Jennings, Hans Bernstein, Laura Whitmore, Ross Carlson, Roslyn Brown, Mary Lipton, Helen Kreuzer, James Moran, Alice Dohnalkova, Margie Romine, James Fredrickson, Susannah Tringe, and Tanja Woyke

174 **In Situ Correlated Molecular Imaging of Chemically Communicating Microbial Communities** 138
R. Masyuko, C.M. Driscoll, **P.W. Bohn*** (pbohn@nd.edu), J. Shrout, and J.V. Sweedler

175 **From Genomes to Metabolomes: Studying Mechanisms of Interspecies Interaction Using the Archaeal System *Ignicoccus-Nanoarchaeum*** 138
Richard Giannone, Brian Bothner, Valérie Copié, Robert Hettich, and **Mircea Podar*** (mpodar@utk.edu)

* Presenting author

Poster **Page**

176 **System-Scale Modeling of Mycorrhizal Symbiosis** 139
Peter Larsen* (plarsen@anl.gov), Geetika Trivedi, Avinash Sreedasyam, Leland Cseke, and **Frank Collart**

177 **Plant-Microbe Interfaces: Mining Genomic Signatures of Species-Specific Microbe Interactions in *Populus* Using High Density SNP Arrays and Whole Genome Resequencing** 140
Wellington Muchero* (mucherow@ornl.gov), Jessy L. Labbé, Ranjan Priya, Lee Gunter, **Gerald A. Tuskan**, and **Mitchel J. Doktycz**

178 **Plant-Microbe Interfaces: Emerging Research Directions** 140
Jay Chen* (chenj@ornl.gov), Jessy Labbé, Jennifer Morrell-Falvey, Sarah Melton, Jim Guo, Udaya C. Kalluri, Dave Weston, Dale Pelletier, **Gerald A. Tuskan**, and **Mitchel J. Doktycz**

179 **Plant-Microbe Interfaces: Revealing Regularities in Large Datasets with Biological Annotations Using Networks and Rules of their Association** 141
Tatiana Karpinets* (karpinetstv@ornl.gov), Byung Park, Guruprasad Kora, Mustafa Syed, Christopher Schadt, Gerald Tuskan, Dale Pelletier, **Edward Uberbacher**, and **Mitchel J. Doktycz**

180 **Plant-Microbe Interfaces: Dynamic Data Analytics Through Integrated Knowledgebase and LIMS** 142
Guruprasad Kora* (koragh@ornl.gov), Susan Holladay, Tatiana Karpinets, Mustafa Syed, Christopher Schadt, **Edward Uberbacher**, and **Mitchel J. Doktycz**

181 **Plant-Microbe Interfaces: Genotype-by-Environment Interactions Drive Root-Associated Microbiome Composition in Natural Populations of *Populus deltoides*** 143
Christopher W. Schadt* (schadtcw@ornl.gov), Migun Shakya, Neil Gottel, Hector Castro, Zamin Yang, Marilyn Kerley, Gregory Bonito, Dale Pelletier, Tatiana Karpinets, Guruprasad Kora, Jessy Labbé, Edward Uberbacher, Wellington Muchero, Gerald Tuskan, Mircea Podar, Rytas Vilgalys, and **Mitchel J. Doktycz**

182 **Plant-Microbe Interfaces: Characterization of IAA7.1 in *Populus* and its Role in Plant-Microbe Interactions** 143
Poornima Sukumar* (sukumarp@ornl.gov), Tatiana Karpinets, Karuna Chourey, Gregory Hurst, Jay Chen, **Gerald A. Tuskan**, **Udaya C. Kalluri**, and **Mitchel J. Doktycz**

183 **Plant-Microbe Interfaces: Sequencing of Twenty-One *Pseudomonas* Genomes and Twenty-Three Genomes from Diverse Bacteria Isolated from *Populus deltoides* Rhizospheres and Endospheres** 144
Sagar M. Utturkar* (sutturka@utk.edu), **Steven D. Brown**, Dawn M. Klingeman, Courtney M. Johnson, Stanton L. Martin, Miriam L. Land, Tse-Yuan S. Lu, Christopher W. Schadt, **Dale A. Pelletier**, and **Mitchel J. Doktycz**

184 **Plant-Microbe Interfaces: Whole Genome Sequence Analysis of Bacterial Strains Isolated From the *Populus* Microbiome** 145
Dale A. Pelletier* (pelletierda@ornl.gov), Tse-Yuan Lu, Christopher W. Schadt, Tatiana Karpinets, Sagar Utturkar, Dawn Klingeman, Steven D. Brown, Patricia Lankford, Marilyn Kerley, Amy Schaefer, Caroline S. Harwood, and **Mitchel J. Doktycz**

185 **Plant-Microbe Interfaces: Dynamics of Bacterial Microbiome of *Populus deltoides*** 146
Migun Shakya,* (shakyam@ornl.gov), Neil R. Gottel, Hector F. Castro, Zamin Yang, Marilyn Kerley, **Mircea Podar**, Mitchel J. Doktycz, and **Christopher W. Schadt**

186 **Plant-Microbe Interfaces: Quorum Sensing Systems are Prevalent in the *Populus* Microbiome** 146
Amy L. Schaefer* (amyschae@uw.edu), **Dale A. Pelletier**, Sagar Utturkar, Steven Brown, Colin R. Lappala, Ryan Morlen, Daniel B. Mueller, E. Peter Greenberg, and **Caroline S. Harwood**

187 **Plant-Microbe Interfaces: *Populus deltoides* Supports Distinct Fungal and Bacterial Root Associates From Other Ectomycorrhizal Tree Hosts** 147
Gregory M. Bonito* (gmb2@duke.edu), **Christopher W. Schadt**, Gerald A. Tuskan, Mitchel J. Doktycz, and **Rytas Vilgalys**

Poster **Page**

188 Plant-Microbe Interfaces: Genome-Wide Identification of *Populus* Small Proteins Responsive to Mycorrhizal Symbiosis 148
Ting Li, Jonathan Plett, Chu-Yu Ye, Gregory B. Hurst, Jessy L. Labbé, Wellington Muchero, Priya Ranjan, Hengfu Yin, Gerald A. Tuskan, Timothy J. Tschaplinski, **Francis Martin**, and **Xiaohan Yang*** (yangx@ornl.gov)

189 Plant-Microbe Interfaces: Extending Single Plant-Microbe Co-Expression and Metabolic Networks to Community Scales 149
David J. Weston (westondj@ornl.gov), Dale A. Pelletier, Abhijit A. Karve,* Jennifer L. Morrell-Falvey, Timothy J. Tschaplinski, Sara S. Jawdy, Lee E. Gunter, Nancy L. Engle, Jessy L. Labbé, Tse-Yuan S. Lu, Chris W. Schadt, **Gerald A. Tuskan**, and **Mitchel J. Doktycz**

190 Plant-Microbe Interfaces: Proteomics at the Plant-Microbe Interface 150
Gregory B. Hurst* (hurstgb@ornl.gov), Karuna Chourey, Xiaohan Yang, Ting Li, Chongle Pan, Zhou Li, Rachel M. Adams, Robert L. Hettich, Patricia K. Lankford, Keiji G. Asano, Tatiana V. Karpinets, Udaya C. Kalluri, Poornima Sukumar, David J. Weston, Sara M. Allen, Dale A. Pelletier, Jonathan Plett, Francis Martin, Timothy J. Tschaplinski, **Gerald A. Tuskan**, and **Mitchel J. Doktycz**

191 Nutrient Cycling for Biomass: Interactive Proteomic/Transcriptomic Networks for Global Carbon Management Processes within Poplar-Mycorrhizal Interactions 151
Leland J. Cseke* (csekel@uah.edu), Frank M. Collart, Peter E. Larsen, Stephen Barnes, and Helen Kim

192 Global Gene Expression Profiling of Switchgrass Following Inoculation with *Burkholderia phytofirmans* Strain PsJN 151
Alejandra Lara-Chavez* (alejandra.lara@ialr.org), Seonhwa Kim, Scott Lowman, Yuhong Tang, Barry Flinn,, John Seiler, Jerzy Nowak, and **Chuansheng Mei**

193 ENIGMA Biotechnology: Systems Approaches to Studying Microbial Communities 153
Richard Baran, Benjamin Bowen, Morgan Price, Christopher Petzold, Simon Allen, Susan J. Fisher, Haichuan Liu, Steve Hall, Evelin D. Szakal, H. Ewa Witkowska, Adam A. Arkin, Gareth P. Butland, Swapnil Chhabra, John Tainer, Greg Hura, Mark Biggin, Kenneth Downing, Robert Glaeser, Manfred Auer, Gary Siuzdak, Adam M. Deutschbauer, Paul Adams, and **Trent Northen*** (trnorthen@lbl.gov)

194 ENIGMA Biotechnology: Metabolic Profiling of Bacterial Mutant Libraries to Link Metabolite Utilization to Genotype 153
Richard Baran* (RBaran@lbl.gov), Benjamin Bowen, Morgan Price, Adam A. Arkin, Adam Deutschbauer, and **Trent Northen**

195 ENIGMA – Biotechnology: High-Throughput Chemogenomic Fitness Profiling of ENIGMA Relevant Microbes Using a Next Generation Barcode Sequencing Assay 154
Adam M. Deutschbauer, Jennifer He, Morgan Price, Jennifer V. Kuehl, Kelly Wetmore, Nancy L. Liu, Sonia Reveco, **Swapnil R. Chhabra**, Grant M. Zane, Tom R. Juba, **Judy D. Wall**, **John-Marc Chandonia**, **Trent R. Northen**, **Adam P. Arkin**, and **Gareth P. Butland*** (GPButland@lbl.gov)

196 ENIGMA Biotechnology: Identification of Differentially Expressed Metabolites in *Desulfovibrio* and Communities Using Meta-Analysis and the METLIN Database 155
Richard Baran, Kevin Cho, Benjamin Bowen, Paul Adams, Adam A. Arkin, Adam M. Deutschbauer, Linh Hoang, Ewa Kalisiak, Nitin S. Baliga, Zhengjiang Zhu, Ralf Tautenhahn, Trent R. Northen, and **Gary Siuzdak*** (siuzdak@scripps.edu)

197 ENIGMA Biotechnology: Membrane Protein Complexes—Their Roles in *Desulfovibrio vulgaris* Stress Response and in the Establishment and Maintenance of Communities 156
Peter J. Walian* (PJWalian@lbl.gov), Simon Allen, Max Shatsky, Lucy Zeng, Evelin D. Szakal, Haichuan Liu, Bonita Lam, Jil T. Geller, Kristina L. Hillesland, Steven C. Hall, Susan J. Fisher, **Matthew W. Fields**, **David A. Stahl**, **Terry C. Hazen**, **Steven E. Brenner**, **Adam M. Deutschbauer**, **Trent R. Northen**, **John-Marc Chandonia**, **H. Ewa Witkowska**, **Mark D. Biggin**, **Bing K. Jap**, and **Paul D. Adams**

Poster **Page**

198 **ENIGMA Biotechnology: Biofilm Imaging: From Protein Complexes to Intact Microbial Communities** 157
M. Auer, Amita Gorur* (agorur@lbl.gov), P. Arbeleaz, N. Baliga, D.A. Ball, M. Biggin, J.M. Chandonia, S. Chhabra, R. Csencsits, K.H. Downing, M.W. Fields, J.T. Geller, R. Glaeser,, T.C. Hazen, G.L. Hura, T. Juba, B. Lam, C.M. Leung, J. Liphardt, J. Malik, J.P. Remis, S. Reveco, J. Tainer, A. Tauscher, J. Wall, A. Deutschbauer, T. Northen, A. Arkin, and **P.D. Adams**

199 **ENIGMA Computation: An Integrated Framework of Databases and Software for Exploring Genes, Genomes, Proteomes, and Networks in Microbes and Communities** 158
J. Christopher Bare* (Christopher.Bare@systemsbiology.org), Paramvir S. Dehal, Karen Foley, Keith Keller, Morgan N. Price, Wei-Ju Wu, Serdar Turkarslan, Jason Baumohl, Marcin P. Joachimiak, John T. Bates, David J. Reiss, Pavel S. Novichkov, **Adam P. Arkin**, **Nitin S. Baliga**, and **Paul D. Adams**

200 **ENIGMA Computation: Inferring Correlation Networks from Genomic Survey Data** 159
Yonatan Friedman and **Eric J. Alm*** (ejalm@mit.edu)

201 **ENIGMA Computation: Multiscale Modeling of Gene Regulatory Networks Across Evolutionary Timescales by Integrating "Top-Down" and "Bottom-Up" Approaches** 159
Pavel Novichkov,* David J. Reiss* (dreiss@systemsbiology.org), Marcin Joachimiak, Alexey Kazakov, Serdar Turkaslan, **Inna Dubchak**, **Adam P. Arkin**, and **Nitin S. Baliga**

202 **ENIGMA Computation: Evolutionary and Experimental Evidence-Based Functional Annotation of Genes** 160
Morgan N. Price, **Adam P. Arkin**, Ben Bowen, Barbara E. Engelhardt, Ameet Talkwalkar, Jeffrey M. Yunes, Michael L. Souza, Gaurav Pandey, Susanna Repo, Michael I. Jordan, **Steven E. Brenner**,* (brenner@compbio.berkeley.edu), Nitin Baliga, and Paul D. Adams

203 **ENIGMA Computational Core Group: Progress and Directions** 161
David Reiss, Serdar Turkarslan, Jason Baumohl, Paramvir Dehal, Marcin Joachimiak, Keith Keller, Morgan Price, Max Shatsky, Alexey Kazakov, Pavel Novichkov, Yonatan Friedman, Sonia Timberlake, Michael Souza, John-Marc Chandonia, Steven Brenner, Inna Dubchak, Adam P. Arkin, **Eric Alm**, **Nitin Baliga*** (Nitin.Baliga@systemsbiology.org), and Paul D. Adams

204 **ENIGMA Environmental Overview: Field to Lab to Field** 162
Terry C. Hazen* (tchazen@utk.edu), **Matthew W. Fields**, Jizhong Zhou, Joy Van Nostrand, Dwayne A. Elias, Jennifer J. Mosher, Steven D. Brown, Tommy J. Phelps, Mircea Podar, Anthony V. Palumbo, Richard A. Hurt, Anup Singh, **Eric Alm**, Mark B. Smith, Dominique C. Joyner, Romy Chakraborty, Boris Faybishenko, Kara DeLeon, Jil Geller, Bonita Lam, Tamas Torok, Julian Fortney, Sharon E. Borglin, David Stahl, Judy Wall, **Adam P. Arkin**, and **Paul D. Adams** (PDAdams@lbl.gov)

205 **ENIGMA Environmental Microbiology: Microbial Community Dynamics in Groundwater and Surrogate Sediments During HRC® Biostimulation of Cr(VI)-Reduction** 163
K.B. De Leon, B.D. Ramsay, D.R. Newcomer, B. Faybishenko, T.C. Hazen, M.W. Fields* (matthew.fields@erc.montana.edu), and **Paul D. Adams** (PDAdams@lbl.gov)

206 **ENIGMA Environmental: Succession of Hanford Groundwater Microbial Communities During Lactate Amendment and Electron-Acceptor Limitation** 164
Dwayne A. Elias* (eliasda@ornl.gov), Jennifer J. Mosher, Tommy J. Phelps, Mircea Podar, Richard A. Hurt, James H. Campbell, Meghan M. Drake, James G. Moberly, Christopher W. Schadt, Steven D. Brown, Terry C. Hazen, Adam P. Arkin, Anthony V. Palumbo, Boris A. Faybishenko, and **Paul D. Adams**

207 **ENIGMA Environmental Microbiology: Single-Cell Analysis Platforms for Genomic Analysis of Uncultivable Environmental Microbes** 164
Suzan Yilmaz* (syilmaz@sandia.gov), Peng Liu, Robert J. Meagher, Yooli Light, Adam P. Arkin, Terry C. Hazen, **Anup K. Singh**, and Paul D. Adams

Poster **Page**

208 **ENIGMA Environmental Microbiology: Metagenomics-Enabled Understanding of Microbial Communities at DOE Contaminated Sites** 165

Jizhong Zhou* (jzhou@ou.edu), Zhili He, Ye Deng, Aifen Zhou, Ping Zhang, Qichao Tu, Joy Van Nostrand, Hao Yu, Zhou Shi, James Voordeckers, Yongjin Lee, Rong Song, Liyou Wu, **Matthew W. Fields**, **David A. Stahl**, **Judy D. Wall**, **Terry C. Hazen**, **Adam P. Arkin**, and **Paul D. Adams** (PDAdams@lbl.gov)

209 **ENIGMA Environmental Microbiology: Effect of Nitrate Stress on Metal-Reducing Microbes and Results of Nitrate Push/Pull Field Tests at Hanford 100H** 166

Romy Chakraborty* (rchakraborty@lbl.gov), Sharon E. Borglin, Boris Faybishenko, Paramvir Dehal, **Adam P. Arkin**, **Terry C. Hazen**, **Matthew W. Fields**, Jil Geller, Julian Fortney, Dominique Joyner, Mark Conrad, and **Paul D. Adams** (PDAdams@lbl.gov)

210 **ENIGMA Microbiology Physiology: Dissecting the Physiology and Community Interactions of Environmental Bacteria** 167

Adam P. Arkin* (APArkin@lbl.gov), Kelly Bender, David Stahl, Aindrila Mukhopadhyay, Jan Liphardt, Swapnil Chhabra, Adam Deutschbauer, Michael W.W. Adams, Steve Yannone, Mark Biggin, Paramvir Dehal, Marcin Joachimiak, Kimberly Keller, Birte Meyer, James Moberly, Morgan Price, Lara Rajeev, Trent Northen, Matthew W. Fields, Dwayne Elias, Nitin Baliga, **Judy Wall**, and Paul Adams

211 **ENIGMA Microbial Physiology: Systematic Mapping of Two Component Response Regulators to Gene Targets in a Model Sulfate-Reducing Bacterium** 167

L. Rajeev, E. Luning, K. Keller, G. Zane, P. Dehal, M. Price, A. Arkin, J. Wall, and **A. Mukhopadhyay*** (amukhopadhyay@lbl.gov)

212 **ENIGMA Microbial Physiology: Evidence-Based Annotation of Gene Function in Metal-Reducing and Sulfate-Reducing Bacteria Using Genome-Wide Mutant Fitness Data** 168

Adam Deutschbauer* (AMDeutschbauer@lbl.gov), Grant M. Zane, Samuel Fels, Morgan Price, Jennifer V. Kuehl, Kelly Wetmore, Swapnil R. Chhabra, Thomas R. Juba, Gareth P. Butland, **Adam P. Arkin**, **Judy D. Wall**, and Paul Adams

213 **ENIGMA Laboratory: Evolutionary and Ecological Origins of Community Assembly, Stability, and Efficiency** 168

Birte Meyer* (birtemey@u.washington.edu), Kristina L. Hillesland, Jason Flowers, Nicolas Pinel, Nicholas Elliott, **Judy Wall**, Marcin Joachimiak, Aifen Zhou, **Joe Zhou**, Jennifer Kuehl, **Adam Deutschbauer**, Paramvir Dehal, Morgan Price, Zhili He, **Aindrila Mukhopadhyay**, Lara Rajeev, Edward Baidoo, **Trent Northen**, **Nitin Baliga**, **Mark Biggin**, Ming Dong, **Jap Bing**, Peter Walian, **Anup Singh**, Suzan Yilmaz, **Dwayne Elias**, **Matthew Fields**, Hector Garcia-Martin, **Terry Hazen**, **Adam Arkin**, and **David Stahl**

214 **ENIGMA Microbial Physiology: Cooperation Impacts Structure in a Syntrophic Biofilm of *Methanococcus maripaludis* and *Desulfovibrio vulgaris*** 169

Kristen A. Brileya, Matthew W. Fields* (matthew.fields@erc.montana.edu), Adam P. Arkin, and **Paul D. Adams** (PDAdams@lbl.gov)

215 **ENIGMA Microbial Physiology: Assimilatory and Dissimilatory Metallomics of *Desulfovibrio vulgaris* and *Pelosinus* Strain A11** 170

W. Andrew Lancaster* (lancaste@uga.edu), Israel Scott, Brian Vacarro, Angeli Lal Menon, Farris L. Poole, Jil Geller, Jennifer J. Mosher, Terry C. Hazen, Dwayne A. Elias, Michael W.W. Adams, Adam P. Arkin and Paul Adams

216 **ENIGMA Microbiology Physiology: Accurate, High-Throughput Identification of Stable Protein Complexes in *Desulfovibrio vulgaris*** 171

John-Marc Chandonia* (JMChandonia@lbl.gov), Maxim Shatsky, Ming Dong, Haichuan Liu, Lee Yang, Jil T. Geller, Megan Choi, Barbara Gold, Nancy L. Liu, Marjon Khairy, Sonia Reveco, Tom R. Juba, Bonita R. Lam, Evelin D. Szakal, Simon Allen, Sunil Kumar, Farris L. Poole, **Steven E. Brenner**, **Steven C. Hall**, **Susan J. Fisher**, **Michael Adams**, **Terry C. Hazen**, **Judy D. Wall**, **Swapnil Chhabra**, **Jian Jin**, **H. Ewa Witkowska**, **Adam P. Arkin**, **Gareth P. Butland**, **Mark D. Biggin**, and **Paul D. Adams**

Poster **Page**

217 **ENIGMA Microbiology Physiology: A Revised Bioenergetic Model of *Desulfovibrio vulgaris* Strain Hildenborough** 172
Romy Chakraborty* (RChakraborty@lbl.gov), Marcin Joachimiak, Aifen Zhou, Julian L. Fortney, Zhili He, Paramvir Dehal, Morgan R. Price, Jizhong Zhou, Terry C. Hazen, Adam P. Arkin, Judy D. Wall, Adam Deutschbauer, **Swapnil R. Chhabra**, and Paul Adams

218 **Berkeley Synchrotron Infrared Structural Biology (BSISB) Program: A Bioanalytical Facility for Characterizing Chemical Changes in Living Cells** 172
Hoi-Ying N. Holman* (hyholman@lbl.gov), Michael C. Martin, and Peter Nico

219 **Structural and Functional Properties of Transporter Solute Binding Proteins** 173
Sarah Zerbs* (szerbs@anl.gov), Lisa M. Miller, Ruth Pietri, Jamey C. Mack, Andrew Binkowski, Changsoo Chang, Marianne Cuff, Yao Fan, Karolina Michalska, Jureck Osipiuk, Kemin Tan, Andrzej Joachimiak, and **Frank Collart**

Knowledgebase and Computing for Systems Biology 175

220 **KBase: An Integrated Knowledgebase for Predictive Biology and Environmental Research** 175
Adam Arkin* (aparkin@lbl.gov), Robert Cottingham, Sergei Maslov, Rick Stevens, Dylan Chivian, Paramvir Dehal, Christopher Henry, Folker Meyer, Jennifer Salazar, Doreen Ware, David Weston, Brian Davison, and Elizabeth M. Glass

221 **The DOE Systems Biology Knowledgebase: Microbial Communities Science Domain** 176
Folker Meyer* (folker@anl.gov), Dylan Chivian, Andreas Wilke, Narayan Desai, Jared Wilkening, Kevin Keegan, William Trimble, Keith Keller, Paramvir Dehal, Robert Cottingham, Sergei Maslov, Rick Stevens, and **Adam Arkin**

222 **The DOE Systems Biology Knowledgebase: Microbial Science Domain** 177
Paramvir S. Dehal* (psdehal@lbl.gov), Chris S. Henry, Ben Bowen, Steven Brenner, Ross Overbeek, John-Marc Chandonia, Dylan Chivian, Pavel S. Novichkov, Keith Keller, **Adam P. Arkin**, Robert Cottingham, Sergei Maslov, and Rick Stevens

223 **The DOE Systems Biology Knowledgebase: Plant Science Domain** 178
Doreen Ware* (ware@cshl.edu), Sergei Maslov, Shinjae Yoo, Dantong Yu, Michael Schatz, James Gurtowski, Matt Titmus, Jer-ming Chia, Sunita Kumari, Andrew Olson, Shiran Pasternak, Jim Thomason, Ken Youens-Clark, Mark Gerstein, Gang Fang, Darryl Reeves, Pam Ronald, Chris Henry, Sam Seaver, David Weston, Priya Ranjan, Robert Cottingham, Rick Stevens, and **Adam Arkin**

224 **The KBase Architecture and Infrastructure Design** 179
Tom Brettin* (brettints@ornl.gov), Bob Olson, Ross Overbeek, Terry Disz, Bruce Parello, Shiran Pasternak, Folker Meyer, Michael Galloway, Steve Moulton, Dan Olson, Shane Canon, Dantung Yu, Shiran Paternak, Pavel Novichkov, Daniel Quest, Narayan Desai, Jared Wilkening, Miriam Land, Scott Deviod, **Adam Arkin**, Robert Cottingham, Sergei Maslov, and Rick Stevens

225 **The InterBRC Knowledgebase Data Registry** 180
Dylan Chivian* (DCChivian@lbl.gov), Guruprasad Kora, Mustafa Syed, Thomas Brettin, Keith Keller, Jason Baumohl, Yury Bukhman, Richard LeDuc, Adam Arkin, David Benton, **Steve Slater**, and **Edward Uberbacher**

226 **GGKbase, a Portable Knowledgebase for Analysis and Integration of "Omic" Data From Microbial Communities** 180
Brian C. Thomas,* Ken-ichi Ueda, Rahul Basu, Chongle Pan, Trent Northen, Benjamin Bowen,* and **Jill Banfield*** (jbanfield@berkeley.edu)

Poster **Page**

227 **The Ribosomal Database Project: Tools and Sequences for rRNA Analysis** 181
Benli Chai* (chaibenl@msu.edu), Qiong Wang, Jordan Fish, Donna McGarrell, C. Titus Brown, Yanni Sun, James M. Tiedje, and **James R. Cole**

228 **The NamesforLife Semantic Index of Phenotypic and Genotypic Data** 183
Charles T. Parker, Catherine Lyons, and **George M. Garrity*** (garrity@namesforlife.com)

229 **Development of Novel Random Network Theory-Based Approaches to Identify Network Interactions Among Nitrifying Bacteria** 184
Ye Deng* (dengye@glomics.com), Zhili He, Feifei Liu, Yujia Qin, Joy Van Nostrand, and **Jizhong Zhou**

230 **The PhyloFacts Phylogenomic Encyclopedia of Microbial Gene Families** 185
Yaoqing Shen, Bushra Samad, Ajithkumar Warrier, Ruchira Datta, and **Kimmen Sjölander*** (kimmen@berkeley.edu)

231 **Microbial ENergy Processes Gene Ontology (MENGO): New Gene Ontology Terms Describing Microbial Processes Relevant for Bioenergy** 186
Trudy Torto-Alalibo* (trudy@vbi.vt.edu), Endang Purwantini, Joao C. Setubal, Brett M. Tyler, and **Biswarup Mukhopadhyay**

232 **Gene Ontology Terms Describe Biological Production of Methane** 186
Endang Purwantini* (epurwant@vbi.vt.edu), Trudy Torto-Alalibo, Joao C. Setubal, Brett M. Tyler, and **Biswarup Mukhopadhyay**

233 **Toward System Biology KnowledgeBase on Transcriptional Regulation in Bacteria** 187
Pavel S. Novichkov* (PSNovichkov@lbl.gov), Dmitry A. Ravcheev, Alexey E. Kazakov, Semen A. Leyn, Adam P. Arkin, Inna Dubchak, and **Dmitry A. Rodionov*** (rodionov@burnham.org)

234 **Reference Collection of Transcriptional Regulons in Bacillales** 188
Semen A. Leyn* (sleyn@burnham.org), Marat D. Kazanov, **Pavel S. Novichkov**, and **Dmitry A. Rodionov**

235 **Community-Based Approach for Genome-Wide Regulon Annotation in Bacteria** 189
Dmitry A. Ravcheev* (dravcheev@sanfordburnham.org), Aaron A. Best, Mikhail S. Gelfand, **Pavel S. Novichkov**, and **Dmitry A. Rodionov**

236 **Computational Methodologies for Identification of Phenotype-Specific Biological Processes in Microbial Communities** 190
Nagiza F. Samatova (samatovan@ornl.gov), **Chongle Pan** (pcl@ornl.gov), **Robert (Bob) L. Hettich** (hettichrl@ornl.gov), **Jill Banfield** (jbanfield@berkeley.edu), Matthew C. Schmidt (mcschmid@ncsu.edu), William Hendrix (wthendri@ncsu.edu), Andrea M. Rocha (amrocha@mail.usf.edu), and Kanchana Padmanabhan* (kpadman@ncsu.edu)

237 **OptCom: A Multi-Level Optimization Framework for the Metabolic Modeling and Analysis of Microbial Communities** 191
Ali R. Zomorrodi* (zomorrodi@engr.psu.edu) and **Costas D. Maranas**

238 **MetRxn: A Knowledgebase of Metabolites and Reactions Spanning Metabolic Models and Databases** 192
Akhil Kumar* (azk172@psu.edu), Patrick F. Suthers, and **Costas D. Maranas**

239 **Computational Design of Bioenergy-Related Metabolic Pathways** 193
Mario Latendresse* (latendre@ai.sri.com), Mike Travers, and **Peter Karp**

240 **Bioenergy Curation in the MetaCyc Database of Metabolic Pathways** 194
Ron Caspi* (caspi@ai.sri.com), Deepika Weerasinghe, and **Peter Karp**

Poster **Page**

241 **Enhancing the SEED Framework for Curation and Analysis of Genomic Data and Genome-Scale Metabolic Models** 194
Christopher Henry* (chenry@mcs.anl.gov), **Ross Overbeek**, Robert Olson, Terrence Disz, Bruce Parello, Rob Edwards, Matt DeJongh, Aaron Best, Fangfang Xia, Scott Devoid, and **Rick Stevens**

242 **Tools and Approaches for Integrating Multiple Genetic and Cellular Networks** 195
Gang Fang, Darryl Reeves* (darryl.reeves@yale.edu), Koon-Kiu Yan, Nitin Bhardwaj, Declan Clarke, Chao Cheng, Pedro Alves, Sergei Maslov, and **Mark Gerstein**

243 **Biological Significance of Gene Modules in an *Arabidopsis thaliana* Co-Expression Network** 196
Darryl Reeves* (darryl.reeves@yale.edu), Gang Fang, and **Mark Gerstein**

244 **Dynamic Metabolic Model Building Based on the Ensemble Modeling Approach** 196
Jimmy G. Lafontaine Rivera* (lafonj@gmail.com), Ali R. Zomorrodi* (zomorrodi@engr.psu.edu), Thomas Wasylenko* (tmwasylenko@gmail.com), **Costas D. Maranas**, **Greg Stephanopoulos**, and **James C. Liao**

245 **Reliable Numerical Methods for FBA and FVA** 197
Yuekai Sun (yuekai@gmail.com)*, Ronan M.T. Fleming, and **Michael A. Saunders**

246 **A Systems Biology Knowledgebase and Analysis Platform for *De Novo* Phenotype Predictions, Integrated Omics Analysis, and Iterative Model Improvement** 197
Edward J. O'Brien* (ejobrien@ucsd.edu), Joshua A. Lerman, Roger L. Chang, Daniel R. Hyduke, Karsten Zengler, Bernhard Ø. Palsson, and **Michael A. Saunders**

247 **Elucidation of Distinct Transcriptional Regulatory Logic in Bacteria: Applications of a Constraints-Based Systems Biology Knowledgebase** 198
Ali Ebrahim* (aebrahim@ucsd.edu), Stephen A. Federowicz, Byung-Kwan Cho, **Karsten Zengler**, and **Bernhard Ø. Palsson**

248 **Multi-Scale Spatially Distributed Simulations of Microbial Ecosystems** 199
William J. Riehl, William Harcombe, Niels Klitgord, Amrita Kar* (akar@bu.edu), Pankaj Mehta, Nathaniel C. Cady, Christopher J. Marx, and **Daniel Segrè*** (dsegre@bu.edu)

249 **A Multi-Scale Approach to the Simulation of Lignocellulosic Biomass** 200
Goundla Srinivas, Glass Dennis, Sergiy Markutsya, Yana Kholod, Ajitha Devarajan, John Baluyut, Monica Lamm, Theresa L. Windus, Xiaolin Cheng* (chengx@ornl.gov), **Mark S. Gordon**, and **Jeremy C. Smith**

250 **A Multiscale Approach to the Simulation of Lignocellusic Biomass** 201
Monica H. Lamm* (mhlamm@iastate.edu), John Baluyut, Ajitha Devarajan, Yana Kholod, Sergiy Markutsya, Theresa L. Windus, and **Mark S. Gordon**

251 **Combining Whole Cell Stochastic Simulations with Systems Biology Approaches** 202
Elijah Roberts, Piyush Labhsetwar, John Cole, Nathan Price, and **Zaida Luthey-Schulten*** (zan@illinois.edu)

252 **Toolbox Model of Evolution of Metabolic Pathways on Networks of Arbitrary Topology** 202
Tin Yau Pang and **Sergei Maslov*** (maslov@bnl.gov)

253 **Enabling the Use of Externally-Built Alignments and Trees in ARB for Evolutionary Analysis** 203
Steve Essinger, Erin Reichenberger, **Chris Blackwood**, and **Gail Rosen*** (gailr@ece.drexel.edu)

Poster **Page**

254 **Numerical Optimization Algorithms and Software for Systems Biology: A Globally Convergent Algorithm for Computing Stable Non-Equilibrium Steady State Concentrations in Genome-Scale Networks** 203
Ronan M.T. Fleming* (ronan.mt.fleming@gmail.com), Ines Thiele, and **Michael A. Saunders**

259 **Development of Predictive Software Tools to Construct and Analyze Large Dynamical Networks for Systems Biology Knowledgebase** 204
Ravishankar R. Vallabhajosyula* (rrv@cfdrc.com), B. Prabhakarpandian, and Kapil Pant

262 **The GreenCut Resource, a Phylogenomically Derived Inventory of Proteins Specific To the Plant Lineage** 205
Rikard Fristedt, HiroakiYamasaki, Steven Karpowicz, Arthur Grossman, and **Sabeeha Merchant*** (merchant@chem.ucla.edu)

Communication and Ethical, Legal, and Societal Issues 207

255 **ELSI Pilot: Assessing and Mitigating the Risks of Large-Scale Metabolic Engineering** 207
J. Geller, S. Lee, D. Tarjan, C. Wu, T. Torok, T.C. Hazen, A.P. Arkin, and **N.J. Hillson*** (njhillson@lbl.gov)

256 **Managing the Risks of Synthetic Biology: Assessing the U.S. Regulatory System** 207
Sarah R. Carter* (scarter@jcvi.org), Michael Rodemeyer, and **Robert Friedman**

257 **Choices and Challenges in Translating Science and Technology from Concepts to Realities** 208
Amy K. Wolfe* (ami@ornl.gov), David J. Bjornstad, W. Christopher Lenhardt, Barry Shumpert, and Stephanie Wang

258 **Biological and Environmental Research Information System: A Multifaceted Approach to DOE Systems Research Communication** 208
Jennifer L. Bownas, Kris S. Christen, Holly L. Haun, Brett Hopwood, Sheryl A. Martin, Marissa D. Mills, Judy M. Wyrick, and **Betty K. Mansfield*** (mansfieldbk@ornl.gov)

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 12:30 p.m. Wednesday.

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

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

Biological Hydrogen Production: Systems Biology and Metabolic Engineering Approaches

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

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

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.

Joint BioEnergy Institute (JBEI)

1

The Challenge of Enzyme Cost in the Production of Lignocellulosic Biofuels

Daniel Klein-Marcuschamer* (DKlein@lbl.gov), Piotr Oleskowicz-Popiel, Blake A. Simmons, and **Harvey W. Blanch**

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

http://www.jbei.org

Project Goals: Determine reliable, documented, and open process models of biofuel processes that can be used to guide research, investment, and policy.

With the aim of understanding the contribution of enzymes to the cost of lignocellulosic biofuels, we constructed a technoeconomic model for the production of fungal cellulases. We found that the cost of producing enzymes was much higher than that commonly assumed in the literature, e.g. the cost contribution of enzymes to ethanol produced by the conversion of corn stover was found to be $0.68/gal if the sugars in the biomass could be converted at maximum theoretical yields, and $1.47/gal if the yields were based on saccharification and fermentation yields that have been previously reported in the scientific literature. We performed a sensitivity analysis to study the effect of feedstock prices and fermentation times on the cost contribution of enzymes to ethanol price.

2

Structural Comparison of Plant Glycosyltransferases

Sara Fasmer Hansen, Andy DeGiovanni, Peter McInerney, Masood Hadi, Ryan McAndrew, Jose Henrique Pereira, Paul Adams, and **Henrik Vibe Scheller*** (hscheller@lbl.gov)

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

http://www.jbei.org

Project Goals: The project aims to develop a fundamental understanding of the function of glycosyltransferases involved in cell wall biosynthesis. A fundamental understanding of the structure-function relationship of this class of enzymes will enable the development of tools for engineering of plants with modified cell walls and improved properties for processing to biofuels.

Plant cell walls are composed primarily of structural polysaccharides including cellulose, hemicelluloses and pectins. These complex polysaccharides are synthesized by glycosyltransferases (GTs)—a family of enzymes that transfer a sugar residue from an activated donor substrate, usually a nucleotide sugar, to an acceptor such as a growing oligosaccharide. GTs generally have narrow substrate specificity, and are highly stereo- and region-specific. The GTs involved in hemicellulose and pectin biosynthesis are membrane proteins located in the Golgi apparatus. Plants have a large number of such proteins, e.g. more than 300 in *Arabidopsis*, most of which have an unknown function.

Predicting the function of a putative GT based on sequence similarities is problematic and many closely related sequences have different catalytic activities. GTs appear to share a limited number of protein fold types and only two structural folds, GT-A and GT-B, have been identified to date. However, for many GT families—particularly those specific to plants—no structure has been solved, so it is not clear if other fold types exist.

Crystallization and structural comparison of the catalytic domains could help to find conserved motifs involved in substrate recognition of the many GTs in plants. We have selected a diverse group of rice and *Arabidopsis* GTs potentially involved in cell wall biosynthesis. Using bioinformatics and modeling, secondary structures were predicted for optimal construction of truncation variants suitable for crystallization. The protein variants were expressed in *E. coli* with fusion protein tags for improvement of solubility and expression and for purification. More than 70 proteins were expressed at high levels as soluble proteins and some were selected for initial crystallization efforts. Crystals have been obtained and results of the analysis will be reported.

Funding is provided by The Carlsberg Foundation and by the U.S. Department of Energy, Office of Science, through contract DE-AC02-05CH11231 with LBNL.

3

A Genome-Wide Survey of Switchgrass Genome Structure and Organization

Peijian Cao,[1] Jerry Jenkins,[2,4] Laura Bartley,[3] Jane Grimwood,[2] Jeremy Schmutz,[2,4] Daniel Rokhsar,[4] Manoj Sharma,[5,6]* (MKSharma@lbl.gov), and **Pamela C. Ronald**[5,6]

[1]China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute, Zhengzhou, China; [2]HudsonAlpha Institute of Biotechnology, Huntsville, Ala.; [3]Department of Botany and Microbiology, The University of Oklahoma; [4]DOE Joint Genome Institute, Walnut Creek, Calif.; [5]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.; and [6]Department of Plant Pathology, University of California, Davis

http://www.jbei.org

Project Goals: To elucidate switchgrass genome structure and function of cell wall-related enzymes.

The perennial grass, switchgrass (*Panicum virgatum* L.), is a promising bioenergy crop and the target of whole genome sequencing. We constructed two bacterial artificial chromosome (BAC) libraries from the AP13 clone of switchgrass to gain insight into the genome structure and organization, initiate functional and comparative genomic studies, and assist with genome assembly. Together representing 16 haploid genome equivalents of switchgrass, each library comprises 101,376 clones with an average insert size of 144 (Hind III-generated) and 110 kb (BstY I-generated). A total of 330,297 high quality BAC-end sequences (BES) were generated, accounting for 263.2 Mbp (16.4%) of the switchgrass genome. Analysis of the BES identified 279,099 known repetitive elements, >50,000 SSRs and 2,528 novel repeat elements, named switchgrass repetitive elements (SREs). Comparative mapping of 47 full-length BAC sequences and 330K BES revealed high levels of synteny with the grass genomes sorghum, rice, maize and Brachypodium. Our data indicate that the overall sequence composition of the switchgrass genome is most similar to that of rice and that the sorghum genome has retained larger microsyntenous regions with switchgrass. The resources generated in this effort will be useful for a broad range of applications.

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.

4

MASCP Gator and ModHunter, Bioinformatics Tools for Identifying Post Translational Modifications in *Arabidopsis*

Joshua Heazlewood* (JHeazlewood@lbl.gov), Hiren Joshi, Andrew Carroll, and Harriet Parsons

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

http://www.jbei.org

Project Goals: To develop a technique that can be used to support experimental approaches in identifying protein modifications involved in cell wall biosynthesis.

The understanding of the mechanisms of post-translational modification (PTM) is vital to elucidating the role of proteins within living organisms. To date, over 600 different types of post-translational modification have been catalogued. However, unlike the proteome, it is currently unfeasible to compute the protein modification repertoire for any system purely from the genome. Modern mass spectrometry is incredibly sensitive and results in a wealth of mass data points on the mass composition of the sample. This high accuracy technique enables the rapid identification of PTMs through delta mass calculations (comparing to an unmodified peptide mass). Phosphorylation has been fairly amenable to characterization using these techniques, and much data exists covering this. However, even with the ability to characterize some PTMs well, only about 25% of mass data from spectrometers match to unmodified peptides. This leaves a large search space in which information about PTMs could be found. However, the analysis of this data is non-trivial, and sophisticated computational techniques are needed to overcome the complex nature of calculations.

5

Biosynthesis and Incorporation of Side-Chain-Truncated Lignin Monomers to Reduce Lignin Polymerization and Enhance Saccharification

Aymerick Eudes[1]* (ageudes@lbl.gov), Jin Sun Kim,[1] Anthe George,[1] Purba Mukerjee,[2] Brigitte Pollet,[3] Peter I. Benke,[1] Fan Yang,[1] Prajakta Mitra,[1] Lan Sun,[1] Ozgul Persil-Cetinkol,[1] Salem Chabout,[3] Grégory Mouille,[3] Ludivine Soubigou-Taconnat,[4] Sandrine Balzergue,[4] Seema Singh,[1] Bradley M. Holmes,[1] Aindrila Mukhopadhyay,[1] Jay D. Keasling,[1] Blake A. Simmons,[1] Catherine Lapierre,[3] John Ralph,[2] and **Dominique Loque**[1]

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.; [2]Great Lakes Bioenergy Research Center, University of Wisconsin–Madison;

[3]IJPB-AgroParisTech, Versailles Cedex France; and [4]INRA-CNRS Evry, France

http://www.jbei.org

Project Goals: Lignin engineering to reduce its DP and cell wall recalcitrance to enzymatic hydrolysis.

Lignocellulosic biomass is utilized as a renewable feedstock in various agro-industrial activities. Lignin is an aromatic, hydrophobic and mildly-branched polymer integrally associated with polysaccharides within the biomass, which negatively affects their extraction and hydrolysis during industrial processing for sugars production. Engineering the monomer composition of lignins offers an attractive option towards new lignins with reduced recalcitrance. The presented work describes a new strategy developed in *Arabidopsis* for the overproduction of rare lignin monomers to reduce lignin polymerization degree (DP). Biosynthesis of these 'DP reducers' is achieved by expressing a bacterial hydroxycinnamoyl-CoA hydratase-lyase (HCHL) in lignifying tissues of *Arabidopsis* inflorescence stems. HCHL cleaves the propanoid side-chain of hydroxycinnamoyl-CoA lignin precursors to produce the corresponding hydroxybenzaldehydes so that plant stems expressing HCHL accumulate in their cell wall higher amounts of hydroxybenzaldehyde and hydroxybenzoate derivatives. Engineered plants with intermediate HCHL activity levels show no reduction in total lignin, sugar content, or biomass yield compared to wild-type plants. However, cell wall characterization of extract-free stems by thioacidolysis and by 2D-NMR revealed an increased amount of unusual C6C1 lignin monomers most likely linked to lignin as end-groups. Moreover the analysis of lignin isolated from these plants using size exclusion chromatography revealed a reduced molecular weight. Furthermore, these engineered lines show saccharification improvement of pretreated stem cell walls. Therefore, we conclude that enhancing the biosynthesis and incorporation of C6C1 monomers ('DP reducers') into lignin polymers represents a promising strategy to reduce lignin DP, and to decrease cell wall recalcitrance to enzymatic hydrolysis.

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.

6

Engineering of Plants with Decreased Xylan and Lignin Contents and Increased Cell Wall Density

Jane Lau,[1] Fan Yang,[1] Pia Damm Petersen,[1,2] Ling Zhang,[1] Prajakta Mitra,[1] Lan Sun,[1] Lina Prak,[1] Berit Ebert,[1] Yves Verhertbruggen,[1] Jin Sun Kim,[1] Anongpat Suttangkakul,[1] Manfred Auer,[1] Henrik Vibe Scheller,[1,3] and **Dominique Loque**[1]* (dloque@lbl.gov)

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.; [2]Department of Plant Biology and Biotechnology, University of Copenhagen, Frederiksberg, Denmark; and [3]Department of Plant and Microbial Biology, University of California, Berkeley

http://www.jbei.org

Project Goals: Using synthetic biology tools to redesign the regulation of secondary cell wall biosynthesis to reduce biomass recalcitrance and increase biomass density and sugar yield.

Plant biomass for bioenergy purposes is composed largely of secondary cell walls, about a third of which is hemicellulose and up to a quarter is lignin, a strong recalcitrant aromatic polymer. Lignin embeds cell wall polysaccharides and inhibits enzymatic degradation, and hemicellulose is mainly composed of xylans, pentoses and polymers which are less desirable than hexoses for fermentation. Unfortunately, none of these polymers can be easily removed without impacting cell wall cohesion thus integrity of vessels and their water and nutrient transport function. In order to improve plant biomass quality with optimized hexose/pentose ratio, reduced lignin content or enhanced cell wall deposition, we developed modular strategies to spatially and temporally fine-tune the deposition of xylan or lignin to vessel by disconnecting their biosynthetic regulation from the network controlling secondary cell wall deposition in fiber cells. With this approach, we generated healthy plants with reduced xylan or lignin and enhanced cell wall deposition in fiber cells, which resulted in significant improvements in sugar releases after various pretreatments.

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.

7

Functional and Structural Characterization of Rice Cellulose Synthase-Like f6 Loss-of-Function Mutants

Andreia Michelle Smith-Moritz,[1] Yves Verhertbruggen,[1] Jeemeng Lao,[1] Lina Prak,[1] Manfred Auer,[1] Henrik V. Scheller,[1] Zhao Hao,[2] Hoi-Ying N. Holman,[2] Pamela C. Ronald,[1] Joshua L. Heazlewood,[1] and **Miguel E. Vega-Sanchez**[1]* (mevega-sanchez@lbl.gov)

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif. and [2]Berkeley Infrared Structural Biology Program, Lawrence Berkeley National Laboratory, Berkeley, Calif.

http://www.jbei.org

Project Goals: Identify cell wall changes associated with lack of mixed linkage glucan accumulation in grasses.

Mixed-linkage glucan (MLG) is a cell wall polysaccharide containing a backbone of unbranched (1,3)- and (1,4) β-glucosyl residues. Based on its occurrence in plants and chemical characteristics, MLG has primarily been associated with the regulation of cell wall expansion due to its high and transient accumulation in young, expanding tissues. The Cellulose synthase-like F (CslF) subfamily of glycosyltransferases has previously been implicated in mediating the biosynthesis of this polymer. We have confirmed that the rice (*Oryza sativa*) CslF6 gene mediates the biosynthesis of MLG by overexpressing it in Nicotiana benthamiana. Rice cslf6 knockout mutants display a drastic decrease in MLG content (97% reduction in coleoptiles and virtually undetectable in other tissues) but otherwise grew normally during vegetative development, showing only a moderate decrease in both plant height and stem diameter. These results challenge previous assumptions on the role of MLG in grass cell wall structure and function. We present here preliminary results, using primarily Fourier Transform Mid-Infrared spectroscopy, that indicate that significant changes occur in cslf6 mutant cell walls in specific tissue and cell types of roots and coleoptiles. These results highlight possible structural modifications in plant polysaccharide organization that occur in response to the loss of MLG in grass cell walls.

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.

8

Feedstock Agnostic Pretreatment Technology

Seema Singh* (SSingh@lbl.gov), Kevin George, Rohit Arora, Chenlin Li, Ian Mathews, Jay Keasling, Taek-Soon Lee, and Blake Simmons

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

http://www.jbei.org

Project Goals: Development of feedstock agnostic pretreatment technology to support year-round operation on multiple local feedstocks to enable lignocellulosic derived advanced transportation fuel.

Efficient and cost-effective biomass pretreatment remains one of the most significant hurdles towards the realization of biofuels that can displace fossil fuels. Pretreatment represents one of the most significant costs from an operational perspective, and as such, JBEI is developing novel biomass pretreatments to help drive the overall costs of the biorefinery down. One cause of this expense, and limited deployment thus far, for the more common biomass pretreatments (e.g., dilute acids, autohydrolysis, dilute bases, organic solvents, steam explosion, lime) is that they are only effective on a narrow range of the available lignocellulosic feedstocks. For example, while dilute acid and ammonia fiber expansion may be relatively effective in pretreating grasses and corn stovers, they are not that effective in pretreating soft woods and hard woods. Additionally, no pretreatment exists today that is known to efficiently pretreat and liberate sugars from mixed feedstock streams (e.g., hardwoods, softwoods, grasses, and agricultural residues fed simultaneously). Year-round operation on multiple local feedstocks and operations that are not dependent on single feedstock availability and price may partially de-risk lignocellulosic derived transportation fuel. We have previously demonstrated that certain ionic liquids (e.g., 1-ethyl-3-methylimidazolium acetate) are very effective in pretreating a wide range of feedstocks, but have yet to demonstrate that this pretreatment technology can efficiently process mixed feedstocks. In that context, for the first time we have developed and demonstrated that ionic liquids can process a mixed feedstock input. Furthermore, we have demonstrated that the hydrolysates generated from this mixed feedstock are suitable for the production of advanced biofuels and/or biofuel precursors through microbial fermentation. These recent advancements in mixed feedstock processing using ionic liquid may support intercropping of feedstocks resulting in increased energy density per acre.

9

Understanding the Interactions of Cellulose with Ionic Liquids and Ionic Liquid/Water Binary Mixture: A Molecular Dynamics Study

Hanbin Liu,[1] Gang Cheng,[1] Hanbin Liu[1]* (hliu2@lbl.gov), Mike Kent,[1] Vitalie Stavila,[2] Kenneth L. Sale,[1] Blake A. Simmons,[1] and **Seema Singh**[1]

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

http://www.jbei.org

Project Goals: Understanding the mechanism of cellulose dissolution and regeneration in ionic liquid.

Biomass pretreatment using ionic liquids has received significant attention over the past 5 years. It has been reported that certain ionic liquids can dissolve cellulose at relatively high loadings, and that the cellulose can be precipitated through the addition of an anti-solvent. In this work, we perform molecular dynamics simulations to study cellulose dissolved in imidazolium-based ionic liquids at high biomass loading (20 wt%). The interactions of the [C2mim][OAc] with the Iβ cellulose structure at room temperature and the interactions within the cellulose structure at 120 ^{0}C were studied. The results show that both cation and anion of [C2mim][OAc] can easily penetrate into the cellulose crystal structure, but that the anion in particular forms strong hydrogen bonds with cellulose. Our results also show that

* Presenting author

the preferential conformation of the methylhydroxyl group of cellulose solvated in [C2mim][OAc] are in the gauche-trans (gt) conformation, in contrast to the dominant trans-gauche (tg) conformation of the cellulose Iβ found in water or after pretreatment with ammonia. Because of the gauche-trans (gt) conformation found mainly in the cellulose II crystal structure, we hypothesize that the regenerated cellulose from the similar pretreatment conditions are composed of the cellulose II structure. This hypothesis was verified by XRD experiments. MD simulations were also carried out to study fundamental intermolecular interactions that drive the subsequent regeneration of cellulose in complex mixtures of ionic liquids, water and cellulose. The structural analysis of cellulose with different concentration of ionic/water binary mixtures provides new insight into the molecular driving forces present in this ternary system.

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.

10

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

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

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.; [2]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; and [3]Department of Plant Pathology, University of California, Davis

http://www.jbei.org

Project Goals: Physical characterization of plant stems and leaves.

Cellulose and hemi-cellulose present in lignocellulosic biomass can be converted to simple sugars through enzymatic hydrolysis and hence to advanced biofuels. 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. A suite of biophysical tools were used to characterize differences in the rice mutants regarding mechanical strength, cell wall composition and fiber organization/orientation. Tensile stress testing followed by SEM imaging of the fractured plant material revealed substantial differences in mechanical strength and elasticity for mutant plants compared to wildtype as well as for plants culturing conditions (long day versus short day), suggesting high sensitivity for this method. To address the effect of mutagenized genes on cell wall fiber organization we developed a polarized Raman microspectroscopy approach, and have found clear differences in the degree of cell wall fiber orientation between mutant and wildtype plants. We are currently evaluating the correlation of both cell wall strength and organization on the saccharification yield.

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.

11

Interaction of Endoglucanases with Amorphous Cellulose Revealed by Quartz Crystal Microbalance and Neutron Reflectivity

Gang Cheng[1]* (gcheng@lbl.gov), Supratim Datta,[1] Chao Wang,[3] Zelin Liu,[3] Jaclyn K. Murton,[2] Page Brown,[2] Michael Jablin,[4] Manish Dubey,[4] Jaroslaw Majewski,[4] Candice Halbert,[5] James Browning,[5] Alan Esker,[3] Brian J. Watson,[6] Haito Zhang,[6] Steven W. Hutcheson,[6] Dale Huber,[2,7] Blake A. Simmons,[1] and **Michael S. Kent**[1]

[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.; [6]Department of Cell Biology and Molecular Genetics, University of Maryland, College Park; and [7]Center for Integrated Nanotechnologies, Albuquerque, N.M.

http://www.jbei.org

Project Goals: Unraveling of interactions between cellulases of different types and insoluble substrates is prerequisite for the design of more effective enzyme systems.

Intensive efforts are underway to lower the cost of enzymatic hydrolysis of cellulose to sugars. There are three general types of cellulases: endoglucanses hydrolyze internal bonds and produce chain ends, exoglucanases hydrolyze from the chain ends and release cellobiose, and b-glucosidases convert cellobiose units to glucose. These generic activities are likely to depend upon substrate characteristics. While each component plays an individual role, they work synergistically for highly efficient cellulose degradation. The full nature of that synergism, and its dependence on substrate characteristics, is not fully understood. A fundamental understanding of enzyme synergy would greatly aid the design of enzyme cocktails. In this work we are using quartz crystal microbalance (QCM) and neutron reflectivity (NR) measurements to reveal the actions of endoglucanases on amorphous cellulose. Amorphous cellulose is relevant to biomass pretreated with ionic liquids. QCD and NR are highly complementary and provide unprecedented detail into the effects of endoglucanses on film structure. Results are shown below for an enzyme cocktail from *T. viride* and from two endos that show qualitatively different behavior.

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.

12

Nanostructure-Initiator Mass Spectrometry (NIMS): High-Throughput Enzyme Activity Assays for Biofuel Development

Xiaoliang Cheng[1]* (xiaoliangcheng@lbl.gov), Wolfgang Reindl,[1] Kai Deng,[1] Benjamin Bowen,[2] Bergeman Lai,[3] John M. Gladden,[1] Steven W. Singer,[1] April Wong,[1] Terry C. Hazen,[1] Brian Fox,[3] Kenneth Sale,[1] Blake A. Simmons,[1] Anup K. Singh,[1] Jay Keasling,[1] Paul D. Adams,[1] and **Trent R. Northen**[1]

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.; [2]Life Sciences, Lawrence Berkeley National Laboratory, Emeryville, Calif.; and [3]Department of Biochemistry, University of Wisconsin–Madison

http://www.jbei.org

Project Goals: This project meets an urgent need for a highly specific activity screening approach and offers tremendous potential for the high-throughput identification and optimization of industrial enzymes and enabling application of biological approaches utilizing large libraries.

The efficient deconstruction of lignocellulosic biomass into biofuels represents a critical and formidable challenge. JBEI is addressing this challenge using a multifaceted approach that is highly dependent on enzyme discovery, optimization and synthetic biology. The optimization of deconstruction processes requires technologies for the high throughput screening and identification of glycoside hydrolase activities. The high sensitivity, specificity, and resolution of mass spectrometry make it well suited for the analysis of sugar molecules. However, the low throughput of conventional GC/MS and LC/MS precludes implementation for screening purposes. Here we present a multiplexed approach based on nanostructure-initiator mass spectrometry (NIMS) that allows for the rapid analysis of several glycolytic activities in parallel under diverse assay conditions. By forming colloids, it was possible to perform aqueous reactions in microwell plates despite the substrate analogs' hydrophobic perfluorinated tags. Our assay can be used both for the characterization of known enzymes (pH and temperature profiles, kinetic studies, ionic liquid tolerance), and the identification of yet unknown activities, even from complex biological samples (environmental and enrichment cultures). We are now integrating this assay with acoustic printing resulting in a 100-fold increase in throughput.

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.

13

Tuning Cellulase Activity Using Carbohydrate Binding Modules

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

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.; [2]Department of Chemical Engineering University of California, Berkeley; and [3]Joint Department of Bioengineering University of California, Berkeley and San Francisco

http://www.jbei.org

Project Goals: (see below)

Current commercial cellulases have not been optimized to saccharify biomass pretreated with ionic liquids. To address this, we have employed a modular approach to rapidly prototype and engineer highly robust cellulases by fusing thermophillic carbohydrate-binding modules (CBMs) to catalytic domains (CD). We demonstrate that the addition of CBMs enhances enzymatic activity compared to the catalytic domain alone at high temperatures when assayed on a model energy crop, switchgrass.

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.

14

Tracing the Determinants of Dual-Substrate Specificity in a Diverse Subfamily of Family 5 Glycoside Hydrolases

G.D. Friedland* (gdfriedland@lbl.gov), Z. Chen, J. Pereira, S.A. Reveco, R. Chan, J.I. Park, M.P. Thelen, P.D. Adams, A.P. Arkin, J.D. Keasling, H.W. Blanch, B.A. Simmons, K.L. Sale, D. Chivian, and **S. Chhabra**

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

http://www.jbei.org

Project Goals: To understand the determinants of specificity in a diverse family of glycoside hydrolases and engineer improved activity.

Enzymes are traditionally viewed as having exquisite substrate specificity; however, recent evidence supports the notion that many enzymes have evolved activities against a range of substrates. The diversity of activities across glycoside hydrolase family 5 (GH5) suggests that this family of enzymes may contain numerous members with activities on multiple substrates. In this study, we combined structure- and sequence-based phylogenetic analysis with biochemical characterization to survey the prevalence of dual-specificity for glucan- and mannan-based substrates in the GH5 family. Examination of amino acid profile differences between the subfamilies led to the identification and subsequent experimental confirmation of an active site motif indicative of dual-specificity. The motif enabled us to successfully discover several new dually-specific members of GH5 and this pattern is present in over seventy other enzymes, strongly suggesting that dual endoglucanase-mannanase activity is widespread in this family. In addition, reinstatement of the conserved motif in a wild type member of GH5 enhanced its catalytic efficiency on glucan and mannan substrates by 175% and 1,600%, respectively. Phylogenetic examination of other GH families further indicates that the prevalence of enzyme multi-specificity in GHs may be greater than has been experimentally characterization.

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.

15

Metagenomics, Metabolic Reconstruction, and High-Resolution Proteomics of Biomass Degradation in a Thermophilic Bacterial Community

Patrik D'haeseleer* (PDHaeseleer@lbl.gov), John Gladden, Joshua Park, Alyssa Redding, Chris Petzold, Martin Allgaier, Dylan Chivian, Steve Singer, Terry Hazen, and Blake Simmons

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

http://www.jbei.org

Project Goals: The Microbial Communities group at JBEI aims to develop a fundamental understanding of how microbial communities degrade targeted biomass feedstocks, and to utilize a targeted, function-based screening approach to genomics and proteomics to identify, isolate, and characterize new enzymes that are capable of efficiently degrading lignocellulosic feedstocks. Focusing on a thermophilic switchgrass-adapted enrichment community yields an order of magnitude more useful enzyme sequences compared to our previous work on a more complex community, 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.

A microbial enrichment culture with high biomass degrading activity was selected for metagenomic sequencing, annotated using JGI's IMG/M system, and binned into phylogenetic groups. Metabolic reconstructions were generated using Pathway Tools, allowing us to assign metabolic roles to the different members of the bacterial community. High resolution MS metaproteomics by EMSL was mapped to the community members to analyze differential expression of their metabolic pathways and identify highly expressed biomass degrading enzymes.

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.

16

Synthetic Biology Design Automation Tool Suite

T.S. Ham, Z. Dmytriv, H.A. Plahar, J. Chen, D. Densmore, R.D. Rosengarten, **N. Hillson*** (njhillson@lbl.gov), N. Elsbree, G. Linshiz, and J.D. Keasling

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

http://www.jbei.org

Project Goals: Our work collectively aims to reduce the time and cost required to pursue large scale cloning and DNA construction tasks, as well as to enable research scales otherwise unfeasible without the assistance of biological design automation software tools and process automation devices.

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 biological design and execution challenges, including the construction of large combinatorial libraries of engineered enzyme-variants and metabolic pathways. Towards addressing these challenges, we have developed a suite of foundational enabling technologies that include an on-line genetic component repository (JBEI-ICE), web-based biological design automation software tools (DeviceEditor, j5, and DNA Constructor), and an open-source biology-friendly robot programming language (PaR-PaR). JBEI-ICE is an open-source distributed platform that unifies and simplifies how genetic components are stored and managed. DeviceEditor offers a visual design canvas for spatially arranging abstractions of these genetic components, and integrates with j5 to automate the cost-optimal design of scar-less, multi-part DNA construction protocols to assemble the components together. DNA Constructor designs optimized protocols for hierarchically constructing related DNA molecules from DNA oligos, providing access to DNA sequences not yet physically available in JBEI-ICE. j5 and DNA Constructor exploit PaR-PaR to integrate with liquid-handling robotic platforms that automate the set up PCR and DNA assembly reactions.

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.

17

Transferring Ionic Liquid Tolerance from the Rain Forest to *E. coli*

Thomas Rüegg,[1] Patrik D’haeseleer,[1] Sharon Borglin,[1] Kristen DeAngelis,[1] Hannah Woo,[1] Erika Lindquist,[2] Jane Khudyakov,[1] Blake Simmons,[1] and **Michael Thelen**[1]* (MPThelen@lbl.gov)

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif. and [2]DOE Joint Genome Institute, Walnut Creek, Calif.

http://www.jbei.org

Project Goals: Microbes found in natural environments such as forest soils with fast decomposition rates produce highly efficient lignocellulolytic enzymes and are often stress-resistant due to adaptation in fluctuating environmental conditions. Using bacteria isolated from such environments, either directly in biofuel production or to improve existing laboratory strains by genetic engineering, can improve lignocellulose degradation and reduce microbial growth inhibition from toxic byproducts. Pretreatment of plant feedstock with ionic liquids (ILs) has significant advantages over current methods for deconstruction of lignocellulosic feedstocks; however, ILs are toxic to the microorganisms used subsequently for biomass saccharification and fermentation. Based on these considerations, one of our major goals at JBEI is to engineer biofuel microbes to tolerate ILs and chemical inhibitors.

At JBEI we are interested in using microbes that are tolerant to ionic liquids and other chemical inhibitors encountered during biofuel processing. Screening a tropical rain forest soil community for IL-tolerant cellulolytic bacteria identified a novel halotolerant anaerobe that grows in up to 0.5M (~8%) 1-ethyl-3-methylimidazolium chloride, or [C2mim]Cl. By creating a fosmid library containing genomic fragments from this bacterium, we discovered a predicted multidrug-efflux pump that promotes better tolerance to [C2mim]Cl in *E. coli* than in the rain forest isolate. IL-induced changes were found in the native bacterial membrane phospholipids, and in the significant differential expression of 1245 genes revealed by global transcriptomics (RNA-Seq) analysis and metabolic pathway reconstruction. The knowledge of these physiological responses provides us with a first step towards engineering microbial IL 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.

* Presenting author

18

Engineering *E. coli* to Convert Plant Biomass Into Fuels

Greg Bokinski[1]* (GEBokinski@lbl.gov), Pamela Peralta-Yahya,[1] Anthe George,[1] Bradley M. Holmes,[1] Eric Steen,[1] Jeffrey Dietrich,[1] Taek Soon Lee,[1] Danielle Tullman-Ercek,[1] Christopher Voigt,[2] Blake A. Simmons,[1] and **Jay Keasling**[1]

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif. and [2]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Mass.

http://www.jbei.org

Project Goals: We sought to demonstrate the production of three advanced (non-ethanol) biofuels directly from plant biomass without the use of added enzymes.

One approach to reducing the costs of advanced biofuel production from cellulosic biomass is to engineer a single microorganism to both digest plant biomass and produce hydrocarbons that have the properties of petrochemical fuels. We engineered *Escherichia coli* to grow using both the cellulose and hemicellulose fractions of several types of plant biomass pretreated with ionic liquids. Our engineered strains express cellulase, xylanase, beta-glucosidase and xylobiosidase enzymes under control of native *E. coli* promoters selected to optimize growth on model cellulosic and hemicellulosic substrates. Our strains grow using either the cellulose or hemicellulose components of ionic liquid pretreated biomass, or on both components when combined as a coculture. Both cellulolytic and hemicellulolytic strains were further engineered with three biofuel synthesis pathways to demonstrate the production of fuel substitutes or precursors suitable for gasoline, diesel, and jet engines directly from ionic liquid-treated switchgrass without externally-supplied hydrolase enzymes. This demonstration represents a major advance towards realizing a consolidated bioprocess. With improvements in both biofuel synthesis pathways and biomass digestion capabilities, our approach could provide an economical route to production of advanced 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.

19

Engineering *Escherichia coli* for Improved Production of FA and FAEE

Fuzhong Zhang* (fuzhongzhang@lbl.gov), Pouya Javidpour, Suzanne Ma, Greg Friedland, Dylan Chivian, Adam Arkin, Swapnil Chhabra, and **Jay D. Keasling**

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

http://www.jbei.org

Project Goals: Developing synthetic biology tools to increase titers and conversion yields for the microbial production of fatty acids and fatty acid-derived biofuels.

Microbial production of chemicals is now an attractive alternative to chemical synthesis. However, there are very few strategies for engineering regulatory components and co-factor manipulation to improve product titers and conversion yields of heterologous pathways1. Nature has evolved sensors for a variety of intracellular and exogenous molecules, however the cognate regulators are rarely optimal for modulating engineered biosynthetic pathways. To demonstrate the utility of assimilating natural sensors and engineering regulators, we have developed a dynamic sensor-regulator system (DSRS) for the production of fatty acid ethyl esters (FAEEs) in *Escherichia coli*. DSRS detects a key intermediate in the fatty acid biosynthetic pathway and dynamically regulates expression of enzymes involved in FAEE production. The engineered DSRS optimized the host's metabolism, improved the genetic stability of the producing strain, and significantly enhanced the FAEE conversion yield. Manipulation of enzyme cofactor-specificity is an alternative engineering approach, especially in strategies that involve overexpression of cofactor-dependent enzymes. For operation and cost efficiency in an industrial context, anaerobic culture conditions would be preferred, but this raises the issue of NADH becoming more readily available than NADPH within the cell and poses a challenge for a key step in the fatty acid biosynthetic cycle: reduction mediated by the NADPH-dependent FabG enzyme. Through sequence alignment analysis and mutagenesis, we have identified *E. coli* FabG variants that potentially have a greater specificity for NADH than for NADPH. Here we describe our efforts in manipulating cofactor dependence of a highly conserved step in fatty acid biosynthesis.

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.

20

Metabolic Engineering of Mevalonate Pathway in *E. coli* for Isoprenoid Fuel Production

Taek Soon Lee* (TSLee@lbl.gov), Kevin W. George, Rafael Rosengarten, Nathan Hillson, David Garcia, Suzanne Ma, Swapnil Chhabra, Alyssa Redding-Johanson, Tanveer Batth, Christopher Petzold, Paul Adams, and Jay D. Keasling

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

http://www.jbei.org

Project Goals: Mevalonate pathway is one of the major biosynthetic pathways of isoprenoid fuel production, and the engineering of this pathway is a key approach to achieve higher production of these biofuels. Various engineering strategies and tools have been explored to identify the bottlenecks of the pathway and to understand the pathway enzymes better. Targeted proteomics and enzyme kinetics provide important information to achieve this goal, and a new design strategy using synthetic biology allows the combinatorial approach to find the best combination of the pathway genes for biofuel production. In this study, we present several engineering strategies of top portion and bottom portion of mevalonate pathway.

Mevalonate pathway has been explored and engineered as an important biosynthetic pathway for the production of isoprenoids in both *E. coli* and yeast. The engineering of mevalonate pathway to produce more IPP (isopentenyl diphosphate) and DMAPP (3, 3-dimethylallyl diphosphate) is a major approach to improve isoprenoid based fuel production. Targeted proteomics provides highly selective protein identification and inexpensive quantification of individual pathway proteins, and this tool can suggest bottlenecks of the metabolic pathway. The first bottleneck identified is HMG-CoA reductase (HMGR), which is one of the key enzymes in mevalonate pathway. By replacing the original NADPH-dependent HMGR into an NADH-dependent HMGR identified in the analysis of public genome database, we have improved the production about 50% higher and further improvement has been also achieved by increasing the intracellular NADH pool using formate dehydrogenase from Candida. Another bottleneck of the pathway is mevalonate kinase (MK) and phosphomevalonate kinase (PMK), and the kinetic study of these enzymes has been performed to understand the property of these enzymes better. Finally, based on the production and proteomics data we have acquired so far, we re-designed a large number of combinations of 8 separate genes in the mevalonate pathway under 3 operons in the same plasmid, which is pretty different from the original gene order in the operon and genetic context (or a relative position of individual gene within the pathway). We have prepared this combinatorial library of mevalonate pathway using newly developed BioCAD tool, j5, and robotics cloning tool. Using targeted proteomics and production profile, we can quantify mevalonate-pathway proteins for each variant to determine the effects of gene order on protein expression and actual biofuel production.

21

Rational Engineering of Xylose Co-Utilization in Yeast for Advanced Biofuels Production

Mario Ouellet, Heather Szmidt, Paul D. Adams, Jay D. Keasling, and **Aindrila Mukhopadhyay*** (amukhopadhyay@lbl.gov)

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.

http://www.jbei.org

Project Goals: Yeast Engineering for glucose/Xylose co-utilization and advanced biofuels production.

We previously engineered yeast (*S. cerevisiae*) to produce bisabolene, which after hydrogenation, is a good diesel replacement fuel. Economic feasibility of biofuel production from lignocellulosic feedstock will require efficient utilization of all the sugars available in plant hydrolysates. Despite its history of industrial use, yeast cannot naturally utilize pentoses efficiently, including Xylose, a major constituent of hydrolyzed lignocellulosic biomass. Several groups have tried to engineer xylose utilization in yeast for bioethanol production for the last 30 years with very limited success (e.g. slow xylose utilization following depletion of glucose). These attempts usually involved a combination of rational engineering and directed evolution to achieve improved phenotypes. However, the genetic bases of these selected phenotypic improvements have never been characterized. We sought to elucidate the genetic basis of strain phenotypic improvement after evolution on xylose in order to rationally transplant these traits into strains previously engineered for advanced biofuels production. The resulting strains should be capable of using biomass hydrolysates efficiently (i.e. consume glucose and xylose simultaneously in the proportions present) to produce a biofuel.

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.

22

Engineering of Bacterial Methyl Ketone Synthesis for Biofuels

Ee-Been Goh* (egoh@lbl.gov), Jay D. Keasling, and **Harry R. Beller**

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. The goal of this project was to produce methyl ketones from fatty acids for use as blending agents or substitutes for diesel fuel.

We have engineered *Escherichia coli* to overproduce saturated and monounsaturated aliphatic methyl ketones in the C11 to C15 (diesel) range; this group of methyl ketones includes 2-undecanone and 2-tridecanone, which are of importance to the flavor and fragrance industry and also have favorable cetane numbers (as we report here). We describe specific improvements that resulted in a 700-fold enhancement in methyl ketone titer relative to that of a fatty acid-overproducing *E. coli* strain, including the following: (a) overproduction of beta-ketoacyl-coenzyme A (CoA) thioesters achieved by modification of the beta-oxidation pathway (specifically, overexpression of a heterologous acyl-CoA oxidase and native FadB, and chromosomal deletion of fadA) and (b) overexpression of a native thioesterase (FadM). FadM was previously associated with oleic acid degradation, not methyl ketone synthesis, but outperformed a recently identified methyl ketone synthase (ShMKS2, a thioesterase from wild tomato) in beta-ketoacyl-CoA-overproducing strains tested. Whole-genome transcriptional (microarray) studies led to the discovery that FadM is a valuable catalyst for enhancing methyl ketone production. The use of a two-phase system with decane enhanced methyl ketone production by 4 to 7-fold in addition to increases from genetic modifications.

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.

23

Structure of a Three-Domain Sesquiterpene Synthase: A Prospective Target for Advanced Biofuels Production

Pamela P. Peralta-Yahya, Ryan McAndrew* (RPMcandrew@lbl.gov), Andy DeGiovanni, Jose H. Pereira, Masood Z. Hadi, Jay D. Keasling, and **Paul D. Adams**

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

http://www.jbei.org

Project Goals: The high similarity of AgBIS to diterpene synthases makes it an important link in understanding terpene synthase evolution. More practically, the AgBIS crystal structure is important in future protein engineering efforts. Through structural analysis, we can begin to engineer more stable enzymes for increased biofuel production.

The sesquiterpene bisabolene was recently identified as a biosynthetic precursor to bisabolane, an advanced biofuel with physico-chemical properties similar to D2 diesel. Here, we report the structure of AgBIS, a three-domain plant sesquiterpene synthase, crystallized in its apo form and bound to five different inhibitors. Structural and biochemical characterization of the AgBIS terpene synthase Class I active site leads us to propose a catalytic mechanism for the cyclization of farnesyl diphosphate into bisabolene via a bisabolyl cation intermediate.

24

A Two-Scale ^{13}C-Based Method for Metabolic Flux Measurement and Prediction

Vinay Satish Kumar, Daniel Weaver, Adam Arkin, and **Hector Garcia Martin*** (hgmartin@lbl.gov)

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

http://www.jbei.org

Project Goals: To measure and predict intracellular metabolic fluxes by using ^{13}C labeling experiments.

Systems biology aims to provide a predictive and quantitative understanding of cell behaviour as the outcome of the interaction of its comprising parts. Metabolic flux profiles (i.e. the number of molecules traversing each biochemical reaction encoded in its genome per unit time) are not only a key phenotypic characteristic but also embody the essence of this complexity since they represent the final functional output of the interactions of all the molecular machinery studied by all the other "omics" fields. Two of the most popular methods for studying metabolic fluxes are Flux Balance Analysis (FBA) and ^{13}C Metabolic Flux Analysis (^{13}C MFA), each of them displaying its own advantages and disadvantages. Here, we present a new method: Two scale ^{13}C Metabolic Flux Analysis (2S-^{13}CMFA), which combines the advantages of FBA and ^{13}C MFA. We showcase its applications and possibilities with data from the KEIO knockout collection.

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.

25

Mutually Consistent Metabolic Flux and Metabolite Concentration Prediction

Vinay Satish Kumar* (VSKumar@lbl.gov), Hector Garcia Martin, Dylan Chivian, and **Adam P. Arkin**

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

Project Goals: To develop next generation data-driven metabolic flux and metabolite concentration prediction procedures for microbial metabolic engineering applications.

Systems biology approaches aim to arrive at biological insight by leveraging complimentary insights provided by different –omics datasets. Of particular importance to elucidate cellular metabolism is the accurate and reliable estimation of metabolic fluxes and metabolite concentrations. Here, we develop a procedure mCAFE that integrates measured metabolite concentrations and ^{13}C based amino-acid isotope ratio data to predict mutually consistent estimates of internal fluxes and unmeasured concentrations. Methodologically, mCAFE builds on Metabolic Flux Analysis (MFA) by additionally incorporating a set of constraints that account for the nonequilibrium thermodynamic behavior of each reaction in the model. These constraints explicitly relate the flux ratios of each reaction (i.e., forward/reverse flux) with the ratios of the involved metabolites. We apply mCAFE to predict fluxes and concentrations in a customized model of *Escherichia coli* metabolism using the data available at the Keio multi-omics database and find that the estimates are substantially more reliable than those predicted by MFA respectively. Furthermore, mCAFE recapitulates 65% of the measured metabolite concentrations in a leave-one-out cross validation test therefore confirming its reliability in concentration prediction. Applying mCAFE to 25 different *E. coli* single gene deletion mutants reveals conserved covariances between metabolite concentrations and metabolic fluxes. Finally, mCAFE gives empirical basis for the extent of reversibility of reactions in *E. coli* under different conditions. With the anticipated increase in metabolomics and flux data for different organisms, we expect mCAFE to be the state of the art predictive procedure to harness these datasets to predict reliable estimates of internal fluxes and unmeasured metabolite concentrations.

Great Lakes Bioenergy Research Center (GLBRC)

26

Development of Crucial Tools for Lignin Research

Yimin Zhu[1,2]* (zhu6@wisc.edu), Allison Mohammadi,[1] Timothy Pearson,[1] Ruili Gao,[1] Yuki Tobimatsu,[1] Matthew Regner,[1] Fachuang Lu,[1,2,3] Hoon Kim,[1,2,3] **John Ralph**,[1,2,3] Michael Hahn,[4] Fang Chen,[5] Richard A. Dixon,[5] and Steve Decker[6]

[1]Department of Biochemistry, [2]DOE Great Lakes Bioenergy Research Center, [3]Wisconsin Bioenergy Initiative, University of Wisconsin, Madison; [4]Complex Carbohydrate Research Center, University of Georgia, Athens; [5]The Samuel Roberts Noble Foundation, Ardmore, Okla.; and [6]National Renewable Energy Laboratory, Golden, Colo.

Project Goals: The goal of our project is to develop crucial tools to address key challenges in lignin research.

Lignin is one of the main and essential biopolymers in vascular plants. It is among the prime barriers against effective utilization of agriculturally important plants, particularly their cell walls, in processes such as ruminant digestibility, biofuels production from lignocellulosics, and pulp and papermaking. One of the main problems in lignin research is the lack of powerful modern methods to answer fundamental and practical questions, such as the structural attributes of the complex biopolymer. This poster delineates our efforts in developing monoclonal lignin antibodies and polymer-supported lignin monomers and oligolignols to aid in localization and analysis of lignins and for delineating relative radical cross-coupling propensities of the various phenolics involved in lignification. We anticipate these tools will find a wide range of applications in cell wall research to help explore today's most pressing and recalcitrant problems in bioenergy research.

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

27

Identification of Enzymes that Produce Acylated Monolignols: Progress in the Pursuit of Zip-Lignin™

Saunia Withers,[1,2,3] Yimin Zhu[1,4]* (zhu6@wisc.edu), Fachuang Lu,[1,4,5] Hoon Kim,[1,4,5] Ji-Young Park,[6] Shawn Mansfield,[6] **Curtis Wilkerson**,[1,2,3] and **John Ralph**[1,4,5]

[1]DOE Great Lakes Bioenergy Research Center; [2]Department of Plant Biology, [3]Department of Biochemistry and Molecular Biology, Michigan State University, East Lancing; [4]Department of Biochemistry, [5]Wisconsin Bioenergy Initiative, University of Wisconsin, Madison; and [6]University of British Columbia, Vancouver, British Columbia, Canada

Project Goals: The project is aimed at overcoming the recalcitrance of biomass toward processing by altering lignin composition and structure.

Lignin is an essential polymer in plant cell walls, providing structural support to cells, strength to stems, lining to the vascular system, and many other functions. However, the strong interunit linkages in lignin and its cross-linking with other plant cell wall polymers make it the most important factor in the recalcitrance of lignocellulosic biomass to processing and enzyme digestion. Plants make lignin from a variety of monolignols including *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol. We have previously shown that the inclusion of a Zip-lignin™ replacement monomer, that is, a monomer conjugate such as coniferyl ferulate that introduces readily cleavable ester bonds into the lignin backbone, allows significantly decreased processing severity. Here we report our efforts on identifying genes producing transferase enzymes that produce acylated monolignols, and toward engineering plants to contain Zip-lignin™.

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

28

Characterization of the *Arabidopsis* Digestibility Mutant *4 COUMARATE CO-A LIGASE 1*

Carl-Erik Tornqvist,[1] Tanya Falbel,[1] Frances Goglio,[1] Nick Santoro,[2] Cliff Foster,[2] and **Sara E. Patterson**[1]* (spatters@wisc.edu)

[1]University of Wisconsin, Madison and [2]Michigan State University, East Lansing

http://glbrc.org

Project Goals: The overall aim of Area 1 of the Great Lakes Bioenergy Research Center (GLBRC) has focused on the improvement of plants for new sustainable sources of bioenergy. Specifically, our project involves identifying genes that are associated with cell wall digestibility, using the model plant *Arabidopsis*.

Through a reverse genetic screen of approximately 1,150 *Arabidopsis* T-DNA lines, with insertions in cell wall-associated genes, numerous outliers with increased digestibility (i.e. glucose and xylose yield per dry weight of sample) were isolated. WiscDsLox473B01 was identified as the mutant with the highest digestibility out of 102 outlier lines. This mutant develops and grows normally, although the stem tissue contains about 25% less lignin than wild type stem tissue. WiscDsLox473B01 has an insertion in the gene *At1g51680*, also known as *4 COUMARATE CO-A LIGASE* (*4CL1*). 4CL1 has been characterized to play an important role in the phenylpropanoid pathway, specifically monolignol synthesis. We will present results on digestibility and cell wall analytical assays and preliminary developmental characterization of this mutant and related family members.

29

Identification and Characterization of Direct Regulators of Secondary Wall Cellulose Synthases (CESAs)

Won-Chan Kim[1,2]* (wonchan@msu.edu), Jae-Heung Ko,[4] Joo-Yeol Kim,[1,2] and **Kyung-Hwan Han**[1,2,3]

[1]Michigan State University, East Lansing; [2]DOE-Great Lakes Bioenergy Research Center, East Lansing; [3]Chonnam National University, Gwangju; and [4]Kyung Hee University, Yongin

http://glbrc.msu.edu/

Project Goals: The overall goal of this project is to develop a knowledge base for biotechnological improvement of biomass feedstock. As a step toward achieving the goal, we pursue two specific objectives: (1) identification of transcriptional regulators that control the expression of the genes involved in secondary wall biosynthesis (*i.e.*, 'gene discovery') and (2) development of utility promoters that can drive transgene expression in a highly sink tissue-specific manner (i.e., 'biomass engineering').

Cellulose, the world's most abundant biopolymer, is a major feedstock for bioenergy. Therefore, genetic manipulation of cellulose synthesis in specific cells/tissues/organs of economically important crops is one of the top-priorities in current plant biotechnology research. Yet, biotechnological manipulation of cellulose biosynthesis has been one of the most challenging tasks, mainly because of the fact that cellulose biosynthesis is carried out by multi-member cellulose synthase complexes at the plasma membrane. In the secondary walls of *Arabidopsis* plants, three CESAs (CESA4, CESA7 and CESA8) are involved in cellulose biosynthesis. The most prudent approach to increase cellulose biosynthesis will be to simultaneously upregulate the genes encoding for all of the necessary cellulose synthases in the presence of increased substrate level. However, little is known about the transcription regulation of these CESA genes. In fact, no transcription factors that bind to the promoter of CESA gene has been reported yet. We have recently identified a direct regulator of CESAs, named Secondary wall Cellulose synthase Regulator 1 (SCR1), that can simultaneously upregulate the expression of all of the three secondary wall CESAs (U.S. Patent, in filing). Several lines of evidence support our hypothesis that SCR1 may function as one of the direct transcriptional regulators of secondary wall cellulose synthase genes CESA4, CESA7, and CESA8. Building upon this preliminary success, we are developing a novel approach, termed 'targeted regulation of cellulose biosynthesis,' which combines our current understanding of cellulose biosynthesis with the state-of-the-art biotechnology to materialize the concept of '*in planta*' or better yet '*in crop*' manipulation of cellulose biosynthesis.

The U.S. Department of Energy (DOE) via the DOE Great Lakes Bioenergy Research Center supported this work.

30

Metabolic and Transcriptional Changes During Induced Senescence in Maize

Rajandeep Sekhon[1]* (rsekhon@glbrc.wisc.edu), Kevin Childs,[2] Nicholas Santoro,[2] Cliff Foster,[2] Robin Buell,[2] Natalia de Leon,[1] and **Shawn Kaeppler**[1]

[1]DOE Great Lakes Bioenergy Research Center, University of Wisconsin–Madison, Madison and [2]DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing

http://www.glbrc.org/

Project Goals: The overall goal of this study is to understand molecular mechanisms underlying senescence and carbohydrate partitioning in maize. Understanding the senescence program can facilitate manipulation of the timing of onset of senescence and can provide approaches to extend the duration of active photosynthesis. Excess sugars accumulated following grain fill can improve the sugar content of maize stover and consequently ethanol production, or can be redirected to more energy dense compounds such as oils. This study supports a role for sugar accumulation in inducing senescence in maize inbred B73, and is consistent with the idea that providing a vegetative sink to utilize excess photosynthate can both extend the period of active photosynthesis as well as increase the energy density of vegetative tissue.

Plant senescence is a degradative process characterized by catabolism of proteins, lipids, and chlorophyll followed by remobilization of breakdown products to the sink. Delay of senescence, therefore, can extend photosynthetic productivity thereby leading to increased carbon fixation in form of grain and biomass yield. However, biochemical and molecular mechanisms regulating leaf senescence are not well understood. In the current study, pre-mature leaf senescence was induced by removal of seed sink through prevention of pollination, and differential metabolic and global transcriptional changes were assessed in leaves and internodes during normal grain filling period. One of the earliest metabolic changes associated with early senescence was hyper-accumulation of several carbohydrates including free, storage (starch), and structural (xylose) sugars in leaves, and to a lesser extent in internodes. This was associated with transcriptional up-regulation of genes involved in synthesis of storage and structural carbohydrates, and those involved in sugar transport. These observations strongly suggest that, in maize, the lack of sink resulting from pollination prevention leads to reprogramming of sugar partitioning, and the resulting hyper-accumulation of sugars induces early senescence. It is quite conceivable that onset of natural senescence, which coincides with abolition of seed sink due to cessation of grain filling, also results from accumulation of free sugars. Conversely, therefore, it is possible that the availability of an alternative sink (e.g. stalks, leave) for storage of excessive sugars after completion of grain filling could potentially delay the onset of natural senescence, and provide a means to increase the energy content of biofuel feedstocks. Overlaying expression data onto metabolic pathways revealed that, in addition to the known processes related to senescence, several novel pathways including those involved in cell wall biosynthesis were also up-regulated. Currently, we are exploring the molecular mechanisms underlying altered sugar status and early senescence by exploring natural genetic variation in maize. Senescence transcriptome and natural variation provides a framework for identification of novel genes involved in senescence and sugar partitioning which can be exploited for development of value-added biofuel feedstocks.

This project is funded through DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494)

31

Systems-Level Discovery and Characterization of Cellulolytic Enzymes from the Wood Wasp Symbiont *Streptomyces*

Taichi E. Takasuka[1,2]* (ttakasuka@glbrc.wisc.edu), Adam J. Book,[1,3] **Cameron R. Currie**,[1,3] and **Brian G. Fox**[1,2]

[1]Great Lakes Bioenergy Research Center and the Departments of [2]Biochemistry and [3]Bacteriology University of Wisconsin–Madison

Project Goals: Lignocellulosic biomass degradation is critical to carbon cycling in terrestrial ecosystems, and is of increasing scientific interest. Cellulolytic microbes deconstruct the recalcitrant polymers present in plant biomass into simple sugars. Our goal is to examine and learn cellulolytic capability from one of the highly cellulolytic bacterium isolated from the cellulosic environment, *Streptomyces*, by the systems-level approaches.

* Presenting author

Lignocellulose represents the largest reservoir of organic polymers in terrestrial ecosystems. Animals typically gain access to the energy stored in plant biomass through anoxic hydrolysis reactions provided by symbiotic gut microbes. Here we reveal an alternative mechanism, the aerobic deconstruction of hemicellulose and cellulose by *Streptomyces* sp. SirexAA-E (ActE), a bacterium associated with the feeding tunnels of the pinewood-boring wasp *Sirex noctilio*. Genome-wide expression profiling, proteomics, and biochemical assays show that this highly cellulolytic microbe produces an expanded repertoire of hydrolytic and oxidative enzymes with high specific activity for deconstruction of cellulose, mannan, xylan, and biomass. The rapid deconstruction of plant polysaccharides provided by aerobic bacteria that associate with herbivorous insects and propagate externally likely represents a substantial, and previously underappreciated contribution to biomass utilization in natural environments. This work provides the first extensive systems biology characterization of the capabilities of the genus *Streptomyces* in the utilization of a wide variety of polysaccharides relevant to biofuels production.

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

32

Identification of Novel Biomass Deconstructing Enzymes via Phylogenetic and Structural Analyses

Adam J. Book[1,2]* (ajbook@wisc.edu), Ragothaman M. Yennamalli,[3] Taichi E. Takasuka,[2,3] Gina R. Lewin,[1,2] Brian G. Fox,[2,3] George N. Phillips Jr.,[2,3] and **Cameron R. Currie**[1,2]

[1]Department of Bacteriology, [2]DOE Great Lakes Bioenergy Research Center, and [3]Department of Biochemistry, University of Wisconsin–Madison

Project Goals: Develop more accurate functional annotation models for cellulase enzyme families based on phylogenetic, structural, and enzymatic analyses.

The tremendous amount of -omic data (genomes, metagenomes, transcriptomes, and proteomes) has provided a great resource for the discovery of highly efficient or novel enzymes for biomass deconstruction. Work in the Currie Lab alone has provided >50 bacterial genomes and >10 metagenomes from lignocellulose-rich environments. The standard annotation processes for carbohydrate active enzymes (CAZymes) gives good general classification; however, high quality functional predictions are elusive because each CAZy superfamily contains multiple enzymatic functional groups (*e.g.* GH5 contains endoglucanases, mannanases and more). The overall goal of this project is to develop more accurate functional annotation models for cellulase based on phylogenetic, structural, and enzymatic analyses. To achieve this we have constructed phylogenetic trees for two major cellulase CAZy superfamilies, the highly populated GH5 superfamily, and the relatively uncharacterized CBM33 superfamily (oxygenase). When these trees are layered with protein structures (solved and modeled), gene expression data, and biochemically tested enzymatic information, predicted functional annotations can be assigned with much higher precision. These improved annotations will then be used to create prediction algorithms so that new sequences can be quickly and accurately classified into functional groups.

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

33

Strategies to Enable the Assembly of Soil Metagenomes

Adina Howe* (howead@msu.edu), Jason Pell, Arend Hintze, James M. Tiedje, and **C. Titus Brown**

Michigan State University, East Lansing

http://www.glbrc.org

Project Goals: A goal of the Great Lakes Bioenergy Research Center is to develop sustainable biofuel practices by optimizing soil, microbe, and plant interactions. Soil because of its great diversity has been identified as a "Grand Challenge" for exploring microbial communities through metagenomic sequencing. Our goal has been to develop effective strategies to analyze and assemble very large and complex metagenomic datasets from soil, thus creating opportunities to explore the biological basis and ecological services of soil microbial communities.

The development of next-generation short-read sequencing technologies has allowed us to sequence soil microbial communities to unprecedented depths. We now have extremely large soil metagenomes, which because of their numbers and short read lengths cannot be analyzed with traditional genomic tools. A de novo metagenomic assembly approach offers several solutions. It significantly reduces the data size by collapsing numerous short reads into fewer contigs while providing longer sequences containing multiple genes and operons. Furthermore, it does not rely on the availability of reference genomes. De novo assembly, however, relies on the ability to store the information on the connectivity between all sequences in a dataset within an assembly graph. Thus, soil assembly is challenged by extremely high sequence diversity, uneven sequencing coverage, sequencing errors and biases, and the availability of large computational resources. We have developed novel approaches to enable soil metagenome assemblies through data reduction, scalable assembly graph representations, and removal of sequencing errors and biases.

Our initial step is to reduce the size of the metagenomic dataset by normalizing the average coverage of the soil metagenome using an approach we term "digital normalization." We eliminate redundant, high-coverage short-reads within a dataset using a single-pass, constant memory algorithm. This normalization reduces the metagenome dataset size so that it has an increasingly even distribution of read-coverage. Comparing assemblies before and after digital normalization for an *E. coli* genome (50x coverage normalized to 5x), we found that assemblies were >99% similar despite eliminating 90% of the reads. Similar results were observed for a subset of a soil metagenome.

Another strategy for data reduction is to partition disconnected components of the assembly graph. To do this, we have implemented a probabilistic representation of the assembly graph using bloom filters. Bloom filters are memory-constant data structures that can be used to store and traverse the assembly graph. As we evaluated the use of bloom filters to represent large assembly graphs, we found that despite the presence of false positive nodes and edges in the bloom filter representation, it is effective in dividing its disconnected components.

Using bloom filters, we were able to partition several metagenomic datasets into millions of disconnected assembly subgraphs. Among these subgraphs, we consistently found a single, dominant partition consisting of 5 to 76% of metagenomic reads. Characterizing the sequences and connectivity within this dominant partition, we identified position-specific biases within sequenced reads suggesting the presence of spurious connectivity within metagenomes. Using a systematic traversal algorithm, we could identify and remove highly connecting sequences from this partition. We found that the filtering of these sequences not only removes potential sequencing artifacts but also improves assemblies (as demonstrated in simulated datasets) and breaks apart the largest partition allowing for scalable assembly. Applying this partitioning approach to a soil metagenome (30 million reads), we decreased assembly memory requirements by 8-fold.

In conclusion, we have developed approaches that can be applied to the assembly of the growing amounts of soil metagenomic sequencing data. Our approach results in numerous smaller datasets which can be analyzed and/or assembled independently (with separate parameters) and in parallel and subsequently combined into a final assembly for a metagenome.. Furthermore, many of our methods can be extended and applied to other sequence analyses (i.e. transcriptomes).

Sequencing was funded and completed by the Joint Genome Institute, a DOE Office of Science User Facility. Other work was funded by the Great Lakes Bioenergy Research Center (DOE BER Office of Science); the National Institute of Food and Agriculture, Amazon AWS Education; and the National Science Foundation.

34

Developing a Toolbox for Soil Metagenomes

Adina Howe[1]***** (howead@msu.edu), Erick Cardenas, Jordan Fish, Jiarong Guo, Aaron Garoutte, Jason Pell,[1] Arend Hintze,[1] James Cole,[1] Patrick Chain,[3,4] Susannah Tringe,[3] Janet Jansson,[2,3] C. Titus Brown,[1] and **James M. Tiedje**[1]

[1]Michigan State University, East Lansing; [2]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [3]DOE Joint Genome Institute, Walnut Creek, Calif.; and [4]Los Alamos National Laboratory, Los Alamos, N.M.

http://www.glbrc.org

Project Goals: Our research supports a goal of the Great Lakes Bioenergy Research Center which is to develop sustainable biofuel practices by optimizing soil, microbe, and plant interactions. We have developed tools for soil metagenome analysis which enables us to explore the genetic and population components of soil and rhizosphere communities, with special emphasis on ecological services. Our tools focus on (1) scalable de novo metagenomic assembly, (2) gene-targeted assembly, (3) comparing 16S rRNA gene phylogenetic profiles from gene-targeted and metagenomic sequencing, and (4) detecting the presence of known soil genomes within soil metagenomes.

Metagenomic de novo assembly with digital normalization and Khmer

Assembly of metagenomic datasets is limited to the diversity of the metagenome and available computational memory. We have developed approaches for scaling de novo metagenomic assembly through a data reduction process we call "digital normalization" and a "divide and conquer" strategy of breaking apart large metagenome assembly graphs. We initially reduce the size of the metagenomic dataset by eliminating redundant, high-coverage short-reads within a dataset using a single-pass, constant memory algorithm. This normalization reduces the metagenome dataset size so that it has an increasingly even distribution of read-coverage. Next, we partition disconnected components of the assembly graph using a probabilistic representation data structure called a Bloom filter. These approaches, combined, enable the scalable assembly of very large soil metagenomic datasets. We present the assemblies of two soil metagenomes assembled within less than 66 Gb of memory.

HMMgs: gene-targeted assembly from metagenomic datasets

HMMgs is a tool for assembling protein-coding sequences of targeted genes from large metagenomic datasets. Rather than doing a global de novo assembly, HMMgs enables the assembly of only contigs likely to code for a specific gene of interest. We use Hidden Markov Models (HMMs) to guide the local assembly. Starting assembly points are chosen by identifying specific sequences present in the reads which are similar to reference sequences for the gene of interest.

* Presenting author

Advantages to this approach are that it requires significantly fewer resources than global assembly and tends to have higher specificity than read-based approaches. We demonstrate and present the successful assembly of *nifH* and *rplB* genes from three Illumina GAII soil metagenomes.

Phylogenetic profile comparisons between pyrosequencing and shotgun sequencing

Pyrosequencing of the SSU rRNA genes amplified directly from environmental samples is now the standard procedure for surveying microbial community structure. In contrast, metagenomic shotgun sequencing does not depend on gene-targeted primers and amplification and thus is not affected by primer bias or chimeras, respectively. Furthermore, shotgun sequences provide opportunities to detect other genes in the ribosomal RNA (rrn) operon, including large subunit genes. Short read lengths and efficiency of annotating enormous number of reads, however, limit the use of metagenomic reads. We compared the phylogenetic profiles of the SSU rRNA genes in pyrotag and shotgun sequencing data from the same soil samples. Pyrosequencing reads were processed using the Ribosomal Database Project Pyrosequencing Pipeline (RDP) and Mothur. SSU rRNA gene fragments in shotgun reads were identified by alignment to known SSU rRNA genes and subsequently classified by the RDP classifier. We found that short reads aligned to the primer regions targeted by our pyrotag data shared high sequence similarity but also detected bases that were not conserved in a few small clades. These sequences were used for evaluating and improving SSU rRNA gene primers for future gene-targeted sequencing efforts. At the phylum-level, community profiles were similar for pyrotag and shotgun data, suggesting that our method for identifying SSU rRNA gene fragments in shotgun data is effective. Our approach also revealed some taxa missed by pyrosequencing primers and supports the use of shotgun metagenomic sequencing for detecting 16S rRNA genes.

Analysis of genomes detected by mapping raw reads to known genomes

To rapidly screen the soil metagenomes to access information about the microbial community and to validate de novo assemblies, we have aligned raw shotgun reads to known reference genomes. A reference genome collection was constructed from 492 complete genomes covering 19 different phyla was retrieved from the NCBI genome database based on metadata in the Gold Database on August 19th, 2011. Microorganisms were manually selected for originating from soil or a similar habitat. Alignment of reads to reference genomes was performed using Bowtie. The coverage of each genome was then calculated for each gene and for the total mapped genome. Out of the 2.7 billion reads from soil growing biofuel crops, 33 million of were successfully mapped to our soil reference genome database (1.28%). Mapped reads ranged from 0.09% to 4.75%, and were more prevalent in rhizosphere samples. The most frequently detected organisms were mostly *Proteobacteria* of which many are known rhizosphere dwellers. Community structure based on detected genomes was similar but not identical to that inferred from 16S rRNA gene pyrosequencing. This is likely due to bias of the databases toward Proteobacteria and human-related organisms. In general average genome coverage was a good predictor of the percentage of mapped reads that were successfully aligned to the contigs.

Sequencing was funded and completed by the Joint Genome Institute, a DOE Office of Science User Facility. Other work was funded by the Great Lakes Bioenergy Research Center (DOE BER Office of Science); the National Institute of Food and Agriculture, Amazon AWS Education; and the National Science Foundation.

35

Microbial Communities in Agricultural Soils have the Potential to Increase Atmospheric Concentrations of Greenhouse Gases

Tracy K. Teal[1]* (tkteal@msu.edu), Vicente Gomez-Alvarez,[2] and **Thomas M. Schmidt**[1,3]

[1]Michigan State University, East Lansing; [2]Enivronmental Protection Agency ORD-NERL, Cincinnati, Ohio; and [3]Kellogg Biological Station, Hickory Corners, Mich.

Project Goals: The conversion of land to agricultural use has resulted in an increased production of carbon dioxide and nitrous oxide from soils and a decreased consumption of methane– all changes that exacerbate the problem of rising concentrations of greenhouse gases in the atmosphere. As the flux of nitrous oxide and methane are almost entirely microbially-mediated, this shift is due to changes in microbial community structure and function. We aim to survey microbial communities under different land management strategies to determine how communities differ, with a particular focus on traits linked to N_2O production and CH_4 consumption. Towards this aim we are studying soil from the Kellogg Biological Station (KBS) Long Term Ecological Research (LTER) study site, focused on field crop ecosystems, as well as the GLBRC network where potential biofuel crops are assessed for their sustainability and productivity. We are combining metadata with metagenomic data and targeted-gene surveys to address specific questions about the relationships between the genetic structure of microbial communities in soil and their production and consumption of greenhouse gases.

The expansion of land under agricultural management has significantly increased atmospheric forcing from greenhouse gases—carbon dioxide, methane and nitrous oxide. Agricultural soils now contribute approximately half the anthropogenically derived N_2O in the atmosphere[1], and soils that are typically a sink for methane consume little to no methane when converted[2]. Microbial communities are almost entirely responsible for catalyzing nitrous oxide flux, reducing nitrate to nitrous oxide and nitrogen in a process known as denitrification[3]. Although present at low abundances, methanotrophs in soil oxidize methane, with higher diversity communities consuming up to 8 g CH_4-C per hectare per day[4]. In an effort to assess changes in the taxonomic composition and metabolic potential of microbial communities, metagenomic and targeted genomic surveys across a gradient of land uses at the KBS LTER and Michigan GLBRC sites were conducted. We have found that microbial communities take approximately 20 years to recover, both taxonomically and functionally, from agricultural management, concomitant with soil biogeochemistry, and that denitrifier communities are particularly affected by agricultural practices. Under agriculture, the denitrifier community increases from approximately 10% of the community to 33%. Additionally the community composition of denitrifiers changes to an increased proportion of ammonia oxidizing bacteria that lack the capacity to reduce N_2O to N_2. It is likely that fertilization is having a primary affect on this shift[5] and further studies on a switchgrass fertilization gradient are being conducted. Our data suggests that both the abundance of denitrifiers in a community, and their structure, determine the rate of nitrous oxide production in soils and that an understanding of denitrifier communities could lead to solutions for mediation or more accurate models of terrestrial nitrous oxide fluxes.

As methanotrophs are rarer community members, targeted gene approaches for *pmoA* were used and revealed that among agricultural management, perennial crops—swithgrass and prairie—maintain a higher diversity of methanotrophs suggesting higher methane consumption at these sites.

As we strive to develop biofuel crops, the sustainability of these crops and their effect on ecosystem services is an essential component. Our work suggests that land management has important implications for soil microbial communities and the greenhouse gas fluxes they catalyze. An enhanced knowledge of the effects of agriculture on microbial community composition that drives N_2O production and CH_4 consumption is the first step towards managing or restoring microbial biodiversity in soil to mitigate the production of this potent greenhouse gas.

References

1. Crutzen, P., Mosier, A. and Smith, K. N 2 O release from agro-biofuel production negates global warming reduction by replacing fossil fuels. *Atmospheric …* (2008).
2. Robertson, G. and Paul, E. Greenhouse gases in intensive agriculture: Contributions of individual gases to the radiative forcing of the atmosphere. *Science* (2000).
3. Moreno-Vivián, C., Cabello, P., Martínez-Luque, M., Blasco, R. and Castillo, F. Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. *J Bacteriol* 181, 6573–6584 (1999).
4. Levine, U.Y., Teal, T.K., Robertson, G.P. and Schmidt, T.M. Agriculture's impact on microbial diversity and associated fluxes of carbon dioxide and methane. *ISME J* 5, 1683–1691 (2011).
5. Ramirez, K.S., Lauber, C.L., Knight, R., Bradford, M.A. and Fierer, N. Consistent effects of nitrogen fertilization on soil bacterial communities in contrasting systems. *Ecology* 91, 3463–70; discussion 3503–14 (2010).

This research was supported by the U.S. National Science Foundation (MCB-0731913 and the LTER program), and the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494 and DOE OBP Office of Energy Efficiency and Renewable Energy DE-AC05-76RL01830). TKT is supported by an NSF postdoctoral fellowship in Biological Informatics.

36

Microarray Design for Linking Agroecosystem and Gene Function in Soil Microbial Communities

David Duncan[1,2]* (dsduncan@wisc.edu), Richard LeDuc,[1] James Tiedje,[2,3] and **Randall Jackson**[1,2]

[1]University of Wisconsin–Madison; [2]DOE Great Lakes Bioenergy Research Center, Madison, Wis.; and [3]Michigan State University, East Lansing

http://www.glbrc.org/
http://agronomy.wisc.edu/jackson

Project Goals: (1) Identify microbial genetic annotations whose abundances contrast among metagenomes sequenced from functionally different agricultural and grassland soil environments; (2) Design a microarray platform to quantify the abundances of genes that differentiate among environments and known microbial functional markers linked to relevant biogeochemical cycling processes; (3) Evaluate the explanatory power of the relative abundances of these sequences over greenhouse gas and N-cycling fluxes in bioenergy cropping systems toward the end of identifying predictive and diagnostic relationships that can be used to inform land management decisions, identifying links between genes and environmental function, and exploring to what extent knowledge of soil microbial community composition improves our capacity to model biogeochemical cycles. This poster presents our progress on the first and second goals and the approach to be taken with the third.

Among the expectations placed on bioenergy production systems is that they sustainably provide ecosystem services equal to or beyond those obtained from traditional agriculture. Research at the GLBRC and elsewhere provides evidence that these goals can be furthered by perennial, high diversity cropping systems whose low management intensity and spatial extent poses a challenge to efficient and effective monitoring. In order to provide land managers with timely information on the health and functioning of their agroecosystem, we need both an improved understanding of the fundamental drivers of key ecosystem functions and novel techniques for assessing agroecosystem health.

For biogeochemical cycling, the soil microbial community presents a promising target as both a major driver of observed agroecosystem processes and a sensitive integrator of climatic, edaphic, and agronomic effects. However, the size and complexity of the soil microbial community pose a serious technical challenge to meaningfully operationalizing composition metrics. Shotgun metagenomic sequencing has greatly advanced our understanding of the genetic composition microbial communities, particularly since the advent of Next-Gen sequencing and improvements in the computational approaches to handle such data. Despite these advances, metagenomic sequencing remains a nontrivial technical challenge, making it ill-suited for widespread diagnostic use or studies with high biological replication. Narrower approaches to community characterization, such as targeted sequencing and microarrays, provide more tractable datasets, but are limited in their scope to a predefined universe of genes and sequences. In this project, we are combining the open-ended exploratory benefits of shotgun metagenomic sequencing with the throughput and turnaround time of a microarray.

While generalized microbial community functional gene platforms exist, for this project we are designing a custom array populated primarily by sequences from the bulk soil and rhizosphere metagenomes of potential bioenergy crops grown in the Upper Midwest generated by the GLBRC. This will prevent the dilution of relevant sequences with sequences drawn from exotic environments and unlikely to be found in soil systems. In addition, we employ these metagenomes in a novel approach to identify candidate genes for empirical functional association. We identify a set of gene annotations with contrasting abundances among different metagenome sources (e.g. between rhizosphere and the bulk soil), then assay the abundance of these gene groups across a range of systems, potentially linking the genes to a system-scale process of interest. This helps mitigate the dependence of microarrays on *a priori* target selection without need to resort to a blind search. By selecting genes with systematically contrasting abundance in soil environments that have different properties, we expect to identify a gene set enriched for sequences that either drive ecosystem-level functions, or are highly responsive to the drivers of such functions.

In the upcoming growing season, we will use these newly-defined arrays to characterize the microbial community composition associated with potential bioenergy crops grown in the GLBRC Bioenergy Cropping Systems Trial. We will combine this compositional information with trace gas and N-cycling measurements generated through the GLBRC's sustainability research into biogeochemical cycling to identify connections between aspects of community composition and observable ecosystem properties. Going forward, we will revise the contents of the array to reflect our changing understanding of the composition of these soil communities and of their operational units, working toward the development of a platform for diagnosing agroecosystem health and functioning and toward a better understanding of the role of microbial community composition in agroecosystem biogeochemistry.

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.

37

A Genomic Comparison of Two Gram Positive Aerobes Reveals Different Strategies for Carbohydrate Degradation

David Mead[1]* (dmead@lucigen.com), Julie Boyum,[1] Lynne Goodwin,[2] and Philip Brumm[3]

[1]Lucigen Corp. and DOE Great Lakes Bioenergy Research Center, University of Wisconsin–Madison; [2]Los Alamos National Laboratory, Los Alamos N.M.; and [3]C5•6 Technologies 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.

New and improved biomass-degrading enzymes are a major requirement to achieving a viable cellulosic fuels business. *Cellulomonas fimi* ATCC 484 (Celf) and *Streptomyces flavogriseus* ATCC 33331 (Sfla) are high GC actinobacteria reported as cellulolytic organisms. Preliminary results indicate that both microorganisms are capable of hydrolyzing crystalline cellulose. Genomic DNA was prepared for sequencing, assembly and annotation by the DOE Joint Genome Institute. Genomic analysis indicates that both of these soil microbes utilize a secretory model for producing most of their carbohydrate active enzymes (CAZymes); there is no evidence for cellulosomes and little evidence for membrane bound CAZymes. Sfla has almost twice as many genes (6858) compared to Celf (3864) but roughly the same number of CAZymes (220 vs 173). The number and type of glycosyl hydrolases (GH), carbohydrate esterases (CE), pectate lyases (PL), and carbohydrate binding modules (CBM) and the structural families they belonged to were compared. Numerically, the most abundant CBM families are 2, 13, and 32 for Sfla compared to family 2 and 13 for Celf. CMB family 32 binds preferentially to galactosyl saccharides, family 2 binds cellulose, and family 13 binds xylan. Celf has 5 genes and Sfla has 15 genes containing annotated CBMs attached to unknown domains. The Sfla genome contains 5 CBM33 members, compared to only one in Celf. CBM33 proteins may be functionally similar to the GH61 present in fungi, suggesting a more oxidative route to cellulose degradation in Sfla. The Sfla genome contains twelve GH23 and two GH25 genes, compared to two GH23 and zero GH25 genes in Celf; GH23 and GH25 genes code for lysozyme-type enzymes. Sfla possesses eight GH18 family members compared to only one in Celf; GH18 genes typically code for chitinase or cellulase activities. Celf contains slightly more GH13 family members (18 vs. 13); GH13 genes code for amylases. The potency and utility of Celf and Sfla cellulosic enzymes are being studied by GLBRC research scientists.

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

38

Comparative Multiomic Analysis of an *Escherichia coli* Ethanologen During Fermentation of Lignocellulosic Hydrolysates and Synthetic Hydrolysate Reveals Effects of Multiple Inhibitors and Stress Responses on Ethanol Production

Yaoping Zhang[1]* (yzhang@glbrc.wisc.edu), Michael S. Schwalbach,[1] David H. Keating,[1] Mary Tremaine,[1] William Bothfeld,[1] Jeffrey A. Grass,[1] Cameron Cotten,[1] Jennifer L. Reed,[1] Leonardo da Costa Sousa,[2] Mingjie Jin,[2] Venkatesh Balan,[2] James Ellinger,[1] Bruce Dale,[2] Patricia J. Kiley,[1] and **Robert Landick**[1]

Great Lakes Bioenergy Research Center, [1]University of Wisconsin–Madison, Madison and [2]Michigan State University, East Lansing

http://www.glbrc.org/improved-catalysts

Project Goals: The Great Lakes Bioenergy Research Center (GLBRC) aims to identify and overcome key barriers to the sustainable conversion of lignocellulosic biomass to biofuels. Toward that goal, we seek to understand the effects of lignocellulosic hydrolysates on microbial physiology and gene regulation. We compare *Escherichia coli* K-12 ethanologenic strains grown in hydrolysates prepared from corn stover pretreated by ammonia fiber expansion (AFEX-CS hydrolysate or ACSH) to strains grown in a synthetic hydrolysate (SynH) that mimics ACSH and contains similar concentrations of sugars, amino acids, and other small molecules, but lacks the lignotoxins found in authentic ACSH. Comparative analysis of multiomic data allows us to understand the regulatory and metabolic changes associated with growth in ASCH, and to identify the effects of single or multiple inhibitory compound(s) on *E. coli* growth, xylose utilization, and ethanol yield.

A comparison of model ethanologenic *E. coli* strains in ACSH and SynH was done by multiomic analysis to understand the effects of multiple inhibitors and stress responses in hydrolysate on strain performance in the conversion of lignocellulose to biofuel. Growth and gene-expression profiling revealed complicated patterns of metabolic physiology and cellular stress responses throughout an exponential growth phase, a transition phase, and the metabolically active stationary phase that was remarkably similar in ACSH and SynH. Transcriptomic data indicated that genes associated with numerous stress responses were highly expressed in ACSH and SynH, including pathways involved with mitigation of osmotic stress. A number of stress-related genes were uniquely activated in ACSH relative to SynH, including genes encoding efflux pumps associated with export of aromatic hydroxylates, heavy metals and small toxic molecules. Although the glucose in the ASCH and SynH was consumed completely during the fermentation, xylose was utilized more efficiently in SynH than ASCH. This result suggests that xylose utilization is inhibited in ACSH by compounds not present in SynH.

To understand the stresses associated with growth in ACSH, and their effects on xylose utilization and gene expression, we are currently testing a second generation SynH containing potential inhibitors, such as phenolic compounds (lignotoxins), acetate, and higher osmolarity, as well as osmoprotectants such as betaine found in ACSH. These compounds will be investigated individually and collectively to assess their effects on *E. coli* growth, xylose utilization, and ethanol yield. The comparative analysis of these multiomic data will allow for the identification and removal of the bottlenecks associated with the conversion of lignocellulosic biomass to biofuel.

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

39

Transplantation of *Cellvibrio japonicus* Biomass Deconstruction System into *Escherichia coli*

Rembrandt J.F. Haft[1,2]* (rhaft@wisc.edu), Jeffrey G. Gardner,[1] David H. Keating,[1] and **Robert Landick**[1,2,3]

[1]DOE Great Lakes Bioenergy Research Center, Departments of [2]Biochemistry and [3]Bacteriology, University of Wisconsin–Madison

Project Goals: The Great Lakes Bioenergy Research Center (GLBRC) aims to identify and overcome key barriers to the sustainable conversion of lignocellulosic biomass to biofuels. An organism capable of both digesting plant biomass and synthesizing biofuels could be a useful tool for manufacturing next-generation biofuels. One goal is to construct microbial strains that can both release of sugars from cellulose and hemicellulose and convert them to ethanol or other biofuels. To this end, we are working to imbue *Escherichia coli* with the lignocellulolytic capabilities of the plant-degrading γ-proteobacterium *Cellvibrio japonicus*. We are identifying the core set of glycosyl hydrolases necessary for lignocellulose breakdown by *C. japonicus*, and expressing these enzymes in *E. coli* along with the genes encoding the Type II secretion system necessary for their secretion.

A key barrier to economically sustainable lignocellulosic biofuels is the difficulty in releasing sugars from lignocellulose. One approach to overcoming this barrier is to engineer the expression of genes encoding glycosyl hydrolases capable of biomass deconstruction into a biofuel- producing microbe. *E. coli* offers an excellent platform in which to engineer various metabolic pathways to biofuels, but lacks a native ability to produce and secrete relevant glycosyl hydrolases. Our studies have shown that the evolutionarily related γ-proteobacterium *Cellvibrio japonicus* can deconstruct the plant cell walls of bioenergy-relevant substrates such as corn stover and switchgrass. Examination of the genome of

C. japonicus has identified 154 candidate glycosyl hydrolase genes potentially involved in lignocellulose degradation. However, the role of most of these genes in plant cell wall degradation has not been defined.

To identify the core set of lignocellulases necessary for cell wall deconstruction, we used global gene expression profiling to investigate how *C. japonicus* degrades lignocellulose. Although *C. japonicus* encodes a large number of candidate glycosyl hydrolases, fewer than half were found to be expressed during growth in the presence of corn stover. Transplantation of a subset of these genes into *E. coli* allowed *E. coli* to degrade cellulose and xylan, and to grow on cellulosic carbon sources such as carboxymethyl cellulose and Avicel. Collectively, these results demonstrate the potential of this approach to identify lignocellulase genes for engineering of CBP organisms, and to contribute to our overall knowledge of microbial cell wall degradation.

C. japonicus secretes much of its glycosyl hydrolase activity using a Type II secretion system (1). We transplanted the *gsp* operon encoding the *C. japonicus* Type II secretion system into *E. coli* to determine whether it could allow delivery of *C. japonicus* enzymes expressed in *E. coli* to extracellular substrates. We found that the *gsp* genes were expressed by *E. coli*, and that co-expression with certain enzymes lead to general leakage of periplasmic proteins. However, the *C. japonicus gsp* operon alone, while expressed, did not generate a functional Type II secretion system. This provides an ideal system with which we can now screen for components of *C. japonicus* necessary for functional transplantation of its Type II secretion system.

Reference

1. Gardner JG, Keating DH (2010). Appl Environ Microbiol 76 (15):5079-5087.

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

40

Improved Enzyme Mixtures for Biomass Deconstruction

Dina Jabbour* (jabbour@msu.edu), John S. Scott-Craig, Goutami Banerjee, and **Jonathan D. Walton**

DOE-Great Lakes Bioenergy Research Center, Michigan State University, E. Lansing, Mich.

Project Goals: To perform fundamental research on cell wall degrading enzymes that will lead to the development of more efficient mixtures for the conversion of biomass to fermentable sugars.

Currently available commercial enzymes for biomass depolymerization are complex and poorly defined mixtures of activities and proportions. Many of the enzymes, secreted by fungi such as *Trichoderma reesei*, are not necessary for industrial biomass conversion, whereas other critical enzymes are absent or present at sub-optimal proportions. Our long term goal is to construct enzyme mixtures of optimized composition. To do so, we are building synthetic mixtures starting with pure, individual enzymes. Mixtures are optimized using statistical Design of Experiment and robotic liquid handling in an integrated system called GENPLAT. The resulting mixtures are further improved by the addition of novel enzymes that are lacking from commercial preparations such as Accellerase 1000, CTek2, and Spezyme CP. To date, we have constructed optimized cocktails containing up to 15 components. Some contribute to glucose (Glc) yield, others to xylose (Xyl) yield, and some enhance yield of both sugars from various pretreated biomass materials (Banerjee et al., 2010a). The required enzymes and their optimized proportions depend on the pretreatment conditions and biomass composition (Banerjee et al., 2010b).

We are also identifying novel biomass deconstruction enzymes and are testing their contribution to Glc and Xyl yield when combined with synthetic mixtures and with commercial enzymes. Secreted α-xylosidase (AX) is predicted to be important for release of Glc and Xyl from plant cell walls, especially from dicots, which have high levels of xyloglucan. Secreted AX's are rare in nature. We identified and purifed a secreted AX from *Aspergillus niger*. AX activity and even the encoding gene are absent from *T. reesei*. AX in combination with βglucosidase depolymerizes pea xyloglucan to free Xyl and Glc (Scott-Craig et al., 2012).

Another novel enzyme is α-fucosidase, which is predicted to be essential for the deconstruction of fucosylated xyloglucan, found in many dicot plants. As the terminal sugar on xyloglucan sidechains, fucose must be removed before the other enzymes (β-galactosidase, αxylosidase, xyloglucanase, and β-glucosidase) can act.

The use of mixed plant species as biomass feedstock has potential economic and environmental advantages over monocultures. The "Sustainability" area of GLBRC is actively investigating mixed prairie as biomass feedstock. Mixed prairie and old fields contain a high percentage of herbaceous dicots (known as forbs). The cell wall composition of forbs is quite different from grasses, and little is known about what pretreatments and enzymes will be effective on forbs. In consultation with the Sustainability Area, our lab has developed a model set of forbs for fundamental studies on the factors that limit their digestibility and hence potential as biofuel "crops". The plants being studied are goldenrod (*Solidago canadensis*), Queen Anne's lace (*Daucus carota*), lambsquarters (*Chenopodium album*), and milkweed (*Asclepias syriaca*). These species grow over a wide geographical range, they are taxonomically diverse, they are often very common in old fields and prairies, and although weedy, they are not noxious. We are testing pretreatments and optimized enzyme cocktails for release of Glc and Xyl from these model forbs. We hypothesize that the recalcitrance of forbs is due to some combination of the ineffectiveness of current thermochemical pretreatments, release of enzyme inhibitors by the pretreatments, or absence of certain essential enzyme activities in current commercial enzyme mixtures.

References

1. Banerjee G, S Car, JS Scott-Craig, MS Borrusch, M Bongers, JD Walton (2010a) Synthetic multi-component enzyme mixtures for deconstruction of lignocellulosic biomass. Bioresour Technol 101:9097-9105.
2. Banerjee G, S Car, JS Scott-Craig, MS Borrusch, JD Walton (2010b) Rapid optimization of enzyme mixtures for deconstruction of diverse pretreatment/biomass feedstock combinations. Biotechnol Biofuels 3:22.
3. Scott-Craig JS, MS Borrusch, G Banerjee, CM Harvey, JD Walton (2012) Biochemical and molecular characterization of secreted α-xylosidase from *Aspergillus niger.* J Biol Chem, in press.

This work was funded by the U.S. Department of Energy Great Lakes Bioenergy Research Center (Great Lakes Bioenergy Research Center) (DOE Office of Science BER DE-FC02-07ER64494) and by grant DE-FG02-91ER200021 to the MSU-Plant Research Laboratory from the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences and Biosciences.

41

Three-Dimensional Structures of Enzymes Involved in Cellulose Degradation

Kate E. Helmich,[1] Goutami Banerjee,[2] Christopher M. Bianchetti,[1] Nathaniel L. Elsen,[1] John S. Scott-Craig,[2] Robert W. Smith,[1] Thomas J. Rutoski,[1] Brian G. Fox,[1] Jonathan D. Walton,[2] and **George N. Phillips Jr.**[1]* (phillips@biochem.wisc.edu)

[1]Departments of Biochemistry, University of Wisconsin, Madison and [2]Department of Energy Plant Research Laboratory, Michigan State University, East Lansing

A thorough structural understanding of the enzymes involved in cellulose degradation is essential to reduce the enzyme loading needed to generate fermentable sugars. Structural characterization of model cellulases and glycoside hydrolases, along with their substrates, not only contributes to the basic understanding of how these enzymes assist in cellulose degradation, and also serve as a guide in the design of modified or novel enzymes for use in industrial biomass conversion. Currently work is underway to determine and analyze the structures of various cellulose degrading enzymes relevant in bioenergy research.

42

Phenotypic and Genomic Characterization of Natural Isolates of *Saccharomyces cerevisiae* for Growth Tolerance in Lignocellulosic Hydrolysates

Trey K. Sato[1]* (tksato@glbrc.wisc.edu), Jason Shao,[1] Dana J. Wohlbach,[1,3] Li Hinchman,[1] James McCurdy,[1] Enhai Xie,[1] Adam Halstead,[1] Wendy Schackwitz,[4] Christa Pennacchio,[4] Bruce E. Dale,[2,3] Blake A. Simmons,[5] David B. Hodge,[2,3] Venkatesh Balan,[2,3] Kyria Boundy-Mills,[6] **Yury Bukhman**,[1] and **Audrey P. Gasch**[1,3]

[1]DOE Great Lakes Bioenergy Research Center, University of Wisconsin, Madison; [2]DOE Great Lakes Bioenergy Research Center, Michigan State University, Lansing; [3]Department of Genetics, University of Wisconsin, Madison; [3]Department of Chemical Engineering and Materials Science, Michigan State University, East Lansing; [4]DOE Joint Genome Institute, Walnut Creek, Calif.; [5]DOE Joint BioEnergy Institute, Emeryville, Calif.; and [6]University of California, Davis

http://www.glbrc.org/

Project Goals: While cellulosic biofuels have tremendous potential for relieving the global energy demand, a number of hurdles prevent the efficient bioconversion of lignocellulose into ethanol and other biofuels. One well-known example is the cellular stress that is imposed from side products generated from biomass pretreatment, which impact fermentation yield and productivity. While a combination of directed engineering and evolution can improve stress tolerance of microbial strains, this approach can be time consuming. An alternative approach is to utilize a microbial strain with endogenous stress tolerant properties, which could then be engineered and evolved for improved biofuel-producing properties. At the Great Lakes Bioenergy Research Center, we have adopted this approach by identifying environmental isolates of the ethanologenic yeast, *Saccharomyces cerevisiae*, with growth tolerance to variety of lignocellulosic hydrolysates. Upon publishing these results, the complete set of phenotypic data will be made publically available to facilitate strain selection for specific applications in lignocellulosic biofuel production.

Here, we report the results of our phenotypic analysis of approximately 200 natural and industrial isolates of *S. cerevisiae* across lignocellulosic hydrolysates prepared from a variety of biomass pretreated by Ammonia Fiber Expansion (AFEX™), alkaline hydrogen peroxide, ionic liquid, or dilute acid, or lab media containing relevant stress-inducing compounds, such as ethanol, acetic, p-Coumaric and ferulic acids. In addition, we developed an automated software tool that can rapidly analyze growth curves and report growth properties for strain comparison. Using this approach, we identified 3 natural *S. cerevisiae* isolates that maintain rapid growth in multiple stress conditions. In collaboration with the Joint Genome Institute, we have utilized NextGen sequencing and developed mapping tools to resequence the genomes of these diploid strains. Comparative analyses of these genome sequences revealed interesting structural features, including regions of homozygosity and novel gene sequences. Further phenotypic and sequence analyses may uncover genetic properties that confer strain-specific stress tolerance.

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

43

Co-Fermentation of Glucose, Xylose, and Cellobiose by the Beetle-Associated Yeast, *Spathaspora passalidarum*

Tanya M. Long,[1] Yi-Kai Su,[2] Alan Higbee,[1] Jennifer Headman,[1] Laura B. Willis,[1,3] and **T.W. Jeffries**[1,2,3]* (twjeffri@wisc.edu)

[1]Great Lakes Bioenergy Research Center, University of Wisconsin–Madison; [2]Biological Systems Engineering, University of Wisconsin–Madison; and [3]Institute for Microbial and Biochemical Technology, Forest Products Laboratory, Madison, Wis.

http://genome.jgi.doe.gov/Spapa3/Spapa3.home.html

Project Goals: The goal of this project is to uncover basic regulatory and genetic mechanisms that will enable the efficient conversion of lignocellulosic feedstocks to higher value fuels and chemicals.

Fermentation of cellulosic and hemicellulosic sugars from herbaceous or woody biomass could resolve food vs. fuel conflicts inherent in the bioconversion of grains. However, the inability to co-ferment these sugars is a major challenge to the economical use of lignocellulose as a feedstock for renewable fuels. Simultaneous co-fermentation of mixed sugars is problematic for almost all microbes—including yeasts that ferment xylose and cellobiose natively—because glucose represses utilization of other sugars. Recent studies have shown that *Saccharomyces cerevisiae* can be engineered to ferment xylose in the presence of cellobiose (1), but competition by glucose for xylose transport still presents a problem.

Surprisingly, the ascomyceteous, beetle associated yeast, *Spathaspora passalidarum* (2, 3), which ferments xylose and cellobiose natively, can also co-ferment these two sugars in the presence of glucose when appropriate conditions are employed. This could be highly advantageous in simultaneous saccharification and fermentation (SSF) processes. Under appropriate conditions, *S. passalidarum,* will simultaneously assimilate glucose and xylose aerobically; it will simultaneously co-ferment glucose, cellobiose and xylose under oxygen limitation and it has a specific ethanol production rate on xylose more than 3 times faster than the corresponding rate on glucose. Moreover, adapted strains of *S. passalidarum* can co-ferment glucose and xylose from acid and enzymatic hydrolysates containing significant amounts of acetic acid. Metabolome analysis of *S. passalidarum* before onset and during the fermentations of glucose and xylose showed that the concentration of glycolytic intermediates is significantly higher on xylose than on glucose.

We examined the co-fermentation of xylose/cellobiose mixtures, or glucose/xylose/cellobiose mixtures, in duplicate bioreactors under oxygen limitation. In the absence of glucose, xylose and cellobiose were metabolized at essentially similar rates until xylose was depleted. In the presence of glucose, co-utilization of cellobiose and xylose were delayed, but all three sugars were co-utilized at similar rates and sugars were consumed by 68 h. The maximum ethanol production rate from xylose and cellobiose was 1.07 g/l·h and from all three sugars was 0.73 g/l·h. Ethanol yields during the phase of maximum production rate were 0.43 g/g and 0.42 g/g for xylose/cellobiose and glucose/xylose/cellobiose mixtures respectively.

Two strains of *S. passalidarum*, AF8 and E7, showed co-utilization of glucose and xylose in AFEX corn stover and maple hydrolysates. The AF8 strain was able to ferment 93% of raw AFEX hydrolysate containing 10% of monosaccharide and could tolerate 1.6-fold more acetic acid than the parent. The maximum ethanol productivity of AF8 was about 2.5 times higher than the parental adapted strain AF2. The E7 strain was able to produce about 38 g/l ethanol from maple hydrolysate medium containing 65 g/l of xylose and 35 g/l of glucose in 59 h in 2-l bioreactors. The ethanol yields from AFEX corn stover and maple hydrolysate were 0.33 and 0.38, respectively. Our results showed that adaption can improve strains performance and adapted *S. passalidarum* can still co-ferment glucose and xylose in hydrolysates.

Understanding metabolic flux during fermentation is an important challenge in identifying factors that limit efficient ethanol production. Metabolite profiling has long been a useful tool for identifying bottlenecks, however obtaining meaningful data in sufficient replicates can be arduous and the results are often limited in scope. Metabolomics is a rapidly developing field, and HPLC-MS/MS provides a powerful tool to efficiently identify and quantitate a large number of metabolites simultaneously. We have used this approach to simultaneously determine the concentrations of more than 40 intracellular metabolites during the cultivation of *S. passalidarum* on glucose, or xylose under aerobic and oxygen limiting conditions. Metabolites were analyzed during early growth phase and oxygen limited fermentation in order to better define the differences between the growth and ethanol production. The metabolite levels exhibited during fermentation on glucose and xylose reflect the higher fermentation rates observed with xylose and suggest a regulatory mechanism.

These features combined with metabolomic analytical tools make *S. passalidarum* very attractive for SSF applications and for studying regulatory mechanisms enabling bioconversion of lignocellulosic materials by yeasts.

References

1. Ha SJ, *et al.* (2011) Engineered *Saccharomyces cerevisiae* capable of simultaneous cellobiose and xylose fermentation. *Proceedings of the National Academy of Sciences of the United States of America* 108(2):504-509.
2. Nguyen NH, Suh SO, Marshall CJ, and Blackwell M (2006) Morphological and ecological similarities: wood-boring beetles associated with novel xylose-fermenting yeasts, *Spathaspora passalidarum* gen. sp nov and *Candida jeffriesii* sp nov. *Mycological Research* 110:1232-1241.
3. Wohlbach DJ, *et al.* (2011) Comparative genomics of xylose-fermenting fungi for enhanced biofuel production. *Proceed-*

* Presenting author

ings of the National Academy of Sciences of the United States of America 108(32):13212-13217.

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

44

Metabolic Engineering of Bacteria for Sustainable Production of Fatty Acid Derived Products

Rebecca M. Lennen,[1,2] J. Tyler Youngquist,[1,2] Daniel Mendez-Perez,[1] Matthew B. Begemann,[1] Spencer W. Hoover,[1,2] Daniel E. Agnew,[1] and **Brian F. Pfleger**[1,2]* (pfleger@engr.wisc.edu)

[1]Department of Chemical and Biological Engineering, University of Wisconsin–Madison and [2]Great Lakes Bioenergy Research Center, Madison, Wis.

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

Project Goals:

- **Discover and characterize novel biochemical pathways for producing small molecules useful as fuels and commodity chemicals**
- **Apply metabolically engineering strategies to increase flux from heterotrophic carbon sources to desired fuels and chemicals**
- **Apply systems biology and functional genomics methods to understand the impact of engineered pathways on cell viability and product tolerance**

Finding a sustainable alternative for today's petrochemical industry is a major challenge facing chemical engineers and society at large. To be sustainable, routes for converting solar energy into organic compounds for use as both fuels and chemical building blocks must be identified, understood, and engineered. Advances in synthetic biology and other biological engineering disciplines have expanded the scope of what can be produced in a living organism. As in other engineering disciplines, synthetic biologists want to apply a general understanding of biology to construct complex systems from well-characterized parts. Once novel synthetic biological systems (e.g. enzymes for biofuel synthesis) are constructed, they must be engineered to function inside living cells without negatively impacting the host's physiology. In most cases first generation systems fail to meet this goal. My group uses systems biology tools to identify metabolic, regulatory, and/or physiological barriers which often can be overcome with metabolic engineering strategies. Here, I present work to develop strains of bacteria for producing hydrocarbon compounds from sustainable feedstocks. In this poster, I will describe published work which identified and characterized enzymes capable of producing alpha-olefins and work describing engineering efforts to produce fatty acids and fatty acid derived hydrocarbons in *Escherichia coli*. Our work has combined functional genomics analysis, synthetic biology construction techniques, bioinformatics, and metabolic modeling to metabolically engineer superior microorganisms. We have tested our strains in media formulated with biomass-derived sugars and are developing photosynthetic microorganisms to by-pass the biomass middle man.

Publications

1. J.T. Youngquist et al. Kinetic modeling of free fatty acid production in *Escherichia coli* based on continuous cultivation of a plasmid free strain. *Biotechnol Bioeng* (accepted)
2. S.W. Hoover et al. Isolation of improved free fatty acid overproducing strains of *Escherichia coli* via Nile red based high-throughput screening. *Environmental Progress and Sustainable Energy* 2011 Oct 17. [Epub ahead of print]
3. R.M. Lennen et al. Membrane Stresses Induced by Endogenous Free Fatty Acid Overproduction in *Escherichia coli*. *Appl Environ Microb*. 2011 Nov;77(22):8114-28
4. D. Mendez-Perez et al. A gene encoding a modular synthase is involved in α-olefin biosynthesis in *Synechococcus sp.* PCC7002. *Appl Environ Microb*. Jun;77(12):4264-7 (2011).
5. S.W. Hoover et al. Bacterial production of free fatty acids from freshwater, macroalgal cellulose. *Appl Microbiol Biot*. Jul;91(2):435-46 (2011).
6. Lennen, R.M. et al. A process for microbial hydrocarbon synthesis: Overproduction of fatty acids in *Escherichia coli* and catalytic conversion to alkanes. *Biotechnol Bioeng*. 106:2, 193-202 (2010).

This work was funded by the DOE Great Lakes Bioenergy Research Center (GLBRC; DOE Office of Science BER DE-FC02-07ER64494), by the Air Force Office of Scientific Research Young Investigator Program (FA9550-11-1-0038), and Start-up funds from the University of Wisconsin-Madison.

45

Constraint-Based Analysis of Microbial Metabolism and Regulation for Improving Biofuel Production

Cameron Cotten[1,2]* (ccotten@wisc.edu), Joonhoon Kim,[1,2] Christos Maravelias,[1,2] and **Jennifer L. Reed**[1,2]

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

Project Goals: Computational models of biological systems can be used to explain observed behaviors, predict un-measurable quantities, and predict cellular behavior arising from environmental and/or genetic perturbations. Models can be useful in engineering biofuel production strains, where some of the challenges are finding bottlenecks in metabolic pathways and suggesting the appropriate perturbations to force a microorganism to produce more of a compound of interest. Our research efforts over this past year have focused on developing and improving computational tools for designing strains, refining models, integrating omics datasets, and estimating kinetic parameters from experimental data. We have developed

a number of methods and applied them to *Escherichia coli* and *Saccharomyces cerevisiae*.

Metabolic engineering seeks to improve cellular production of valuable biochemicals, such as biofuels. Computational tools are becoming increasingly available to design microbial strains using *in silico* models of metabolism that predict the re-distribution of metabolic fluxes after genetic perturbations. In this poster, we describe results from four recent projects. First, we analyzed fermentation results generated by our collaborators at GLBRC, including biomass and extracellular metabolite concentration data, and made predictions about metabolic states during fermentative growth in AFEX-treated corn stover hydrolysate by *Saccharomyces cerevisiae* and *Escherichia coli*. Second, we developed a constraint-based kinetic model by integrating multi-omics datasets and used it to identify potential reactions limiting flux through central metabolic pathways. Third, we developed a new approach (RELATCH) that shows significantly improved predictions of intracellular flux distributions in response to genetic and environmental perturbations. Fourth, we applied some of our recently developed methods (OptORF, BiMOMA, SimOptStrain) to propose metabolic engineering strategies to improve chemical production using constraint-based models. The genetic strategies identified can include deletion of transcription factors and metabolic genes, as well as over-expression of native and non-native metabolic genes. Using these approaches we have identified strategies for improving ethanol production in *Escherichia coli* and *Saccharomyces cerevisiae*. These approaches can be applied to improve production of a wide variety of compounds in other biological systems by proposing beneficial metabolic and/or regulatory changes.

Publications

1. Dipak Barua, Joonhoon Kim, Jennifer L. Reed, "An Automated Phenotype-Driven Approach (*Gene*Force) for Refining Metabolic and Regulatory Models," PLoS Computational Biology, 2010, vol. 6, issue 10.
2. Joonhoon Kim and Jennifer L. Reed, "OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains," BMC Systems Biology, 2010, vol 4, issue 53.
3. Joonhoon Kim, Jennifer L. Reed, Christos T Maravelias, "Large-Scale Bi-Level Strain Design Approaches and Mixed-Integer Programming Solution Methods," PLoS ONE, 6(9): e24162 2011.
4. Cameron Cotten and Jennifer L. Reed, "Analysis of Multiomics Datasets to Generate Kinetic Parameters for Constraint-Based Metabolic Models," Submitted.
5. Michael S. Schwalbach, David H. Keating, Mary Tremaine, Wesley D. Marner, Yaoping Zhang, William Bothfeld, Alan Higbee, Jeffrey A. Grass, Cameron Cotten, Jennifer L. Reed, Leonardo da Costa Sousa, Mingjie Jin, Venkatesh Balan, James Ellinger, Bruce Dale, Patricia J. Kiley, and Robert Landick, "Complex physiology and compound stress responses during fermentation of alkaline-pretreated corn stover hydrolysate by an *Escherichia coli* ethanologen." Under review.

46

Enhancing Energy Density in Crop Biomass by Redirecting Photosynthate into Triacylglycerols in Vegetative Tissues

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

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

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 feed stocks for biofuel. Enhancing the energy yield of plant biomass can be achieved by accumulating energy-dense compounds such as triacylglycerol (TAGs).

The plant oils group of GLBRC is testing if carbon partitioning can be redirected from starch to oil in plant biomass. We have engineered *Arabidopsis* plants to overproduce the transcription factor WRI1 that controls oil accumulation in seeds, and we have reduced starch biosynthesis by RNA interference (RNAi) of ADP-glucose pyrophosphorylase (AGPase) expression. The resulting transgenic lines accumulated less carbohydrate and produced up to 1% oil per DW in the vegetative tissues. The relative contribution of TAG compared to starch to the overall energy density increased in the AGPRNAi-WRI1 double transgenic line. In addition, these transgenic *Arabidopsis* lines resulted in the accumulation of 10% oil per DW when grown on a medium supplemented with 3% sugar. Heterologous expressions of a type 2 diacylglycerol acyltransferase (DGTT2) from *Chlamydomonas reinhardtii* in *Arabidopsis* transgenic lines also led to the accumulation of TAG in seedlings and mature leaves. The abundance of oil droplets and very long chain fatty acid (VLCFA) in the mature plants was confirmed by ESI-MS. Gene stacking of 35S-DGTT2 and AGPRNAI-WRI1 increased oil content close to 2% per DW in the *Arabidopsis* seedlings. These transgenic lines accumulated up to 14% oil per DW when medium was supplemented with 3% sugar. Transgenic rutabaga lines expressing the AGPRNAI-WRI1 (double gene) construct (T_1 transgenic rutabaga plants) accumulated up to 5% oil per DW in soil grown leaves as measured by ESI-MS.

In addition, we used biodiversity and EST sequencing to discover a novel acyltransferase gene EaDAcT (*Euonymys alatus d*iacyl*glycerol a*cetyl*t*ransferase) from burning bush that produces acetyl-glycerols. These novel oils are low viscosity and therefore can be used directly in some diesel engines. Expression of EaDAcT in *Arabidopsis* seeds resulted in the accumulation of acTAGs, up to 65 mol % of total TAG in the seed oil. The development of novel strategies to address compartmentalization of oil metabolism by metabolic flux analysis is in progress.

47

Improving the Sustainability of Bioenergy Crops Through Arbuscular Mycorrhization

Arijit Mukherjee[1]* (amukherjee@wisc.edu), Rajandeep Sekhon,[2] Shawn Kaeppler,[2] and **Jean-Michel Ané**[2]

[1]Great Lakes Bioenergy Research Center and [2]Department of Agronomy, University of Wisconsin, Madison

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 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 4000 lines of mutagenized population of B73 and identified six 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 (GSE); 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 the GSE-induced early responses in maize at the transcript level and their regulation by ethylene, microarray experiments were performed. Data analysis indicates that these signals induce gene expression in maize and ethylene inhibits this gene expression.

48

Unifying Physics and Biology to Simulate Microbial Denitrification and Nitrous Oxide Flux with the EPIC Model

R.C. Izaurralde[1,2]* (cesar.izaurralde@pnnl.gov), W.B. McGill,[3] J.R. Williams,[4] R.P. Link,[1] D.H. Manowitz,[1,2] and D.E. Schwab[5]

[1]Joint Global Change Research Institute, Pacific Northwest National Laboratory and University of Maryland, College Park; [2]Great Lakes Bioenergy Research Center; [3]University of Northern British Columbia, Prince George, BC, Canada; [4]Texas Agri-Life Extension Service, Texas A&M University, Temple; and [5]University of Natural Resources and Applied Life Sciences, Vienna, Austria

http://www.glbrc.org

Project Goals: The goal of the DOE Great Lakes Bioenergy Research Center is to perform the basic research that generates technology to produce, sustainably, cellulosic biomass and to convert it to ethanol and other advanced biofuels.

Microbial denitrification occurs when nitrate in anaerobic soils and aquatic environments is reduced to form nitrous oxide (N_2O) and dinitrogen gases, which eventually escape to the atmosphere. Nitrous oxide is a potent greenhouse gas and also contributes to reduce the protective layer of ozone in the stratosphere. Atmospheric concentrations of N_2O have been on the rise since the beginning of the industrial revolution due to large-scale manipulations of the N cycle in managed ecosystems; especially through the use of synthetic nitrogenous fertilizer. Process-based simulation models—together with observations—can help design solutions to reduce N_2O emissions from managed ecosystems; especially under a large-scale deployment of biofuel crops and land-use change.[1] Here we describe a process-based submodel of microbial denitrification incorporated in the terrestrial ecosystem model EPIC (Environmental Policy Integrated Climate) and linked to a coupled carbon-nitrogen submodel. Each day during a simulation, EPIC calculates heterotrophic respiration based on carbon-nitrogen pool transformations and adjusted by environmental controls (e.g., temperature, water, mineral nitrogen). EPIC then calculates whether there is enough oxygen to accept the electrons generated by both microbial and root respiration. If oxygen is insufficient, then other nitrogen species (nitrate, nitrite, and nitrous oxide) act as electron acceptors following a competitive inhibition scheme. A ratio of electron accepted / electrons generated is used to correct microbial respiration. An hourly implementation of the gas transport equation is used to move oxygen, carbon dioxide, and nitrous oxide across the gaseous phase of the soil layers and across the soil-atmosphere interface. Examples of model performance will be presented and discussed.

References

1. Zhang, X., R.C. Izaurralde, D. Manowitz, T.O. West, W.M. Post, A.M. Thomson, V.P. Bandaru, J. Nichols, and J.R. Williams. 2010. An integrative modeling framework to evaluate the productivity and sustainability of biofuel crop production systems. Global Change Biol. – Bioenergy 2:258–277.
2. Izaurralde, R.C., W.B. McGill, and J.R. Williams. Development and application of the EPIC model for carbon cycle, greenhouse-gas mitigation, and biofuel studies. 2012. In A. Franzluebbers, R. Follett, and M. Liebig (eds.) Managing Agricultural Greenhouse Gases: Coordinated agricultural research through GRACEnet to address our changing climate, Elsevier (in press).

49

Lipid Production in *Rhodobacter sphaeroides*: From Photosynthetic Membranes to Biofuels

Kimberly C. Lemmer[1,2]* (kchristensen@glbrc.wisc.edu) and **Timothy J. Donohue**[1,2]

[1]Great Lakes Bioenergy Research Center and [2]Bacteriology Department, University of Wisconsin–Madison

Project Goals: We are using *Rhodobacter sphaeroides* as a model system to understand the cellular mechanisms regulating lipid synthesis, and to exploit these pathways for production of biofuel-relevant products. This facultative bacterium has the natural ability to increase phospholipid membrane under photosynthetic conditions. We are using this physiologic shift to investigate the underlying increase in fatty acid biosynthesis and its regulation. Once characterized, these regulatory mechanisms will be used to increase fatty acid production in this and other microbial systems for the production of commercially relevant lipid products. These studies will provide critical insight into the regulation and requirements of cellular lipid biosynthesis and accumulation; knowledge that will be critical for the engineering of this and/or other microorganisms for large-scale biofuel production.

Rhodobacter sphaeroides is a facultative bacterium that can grow via aerobic or anaerobic respiration, photosynthesis, or fermentation. This microbe is an attractive model organism to study lipid biosynthesis, as it has the unique ability to increase lipid production under photosynthetic conditions. Under low oxygen and anaerobic conditions *R. sphaeroides* develops invaginations of the cytoplasmic membrane, called intracytoplasmic membranes (ICM), which increase its membrane surface area and allow for synthesis of the photosynthetic apparatus that harvests light energy. We are using this physiologic increase in phospholipid membrane under anaerobic conditions to investigate the underlying increase in fatty acid biosynthesis and its regulatory mechanisms.

Here we show that *R. sphaeroides* increases its total fatty acid content by 3-4 times per cell under low oxygen and anaerobic growth conditions. We have found that this anaerobic induction of lipid synthesis is dependent on multiple regulatory pathways that collaborate to induce photosynthetic capabilities of the bacterium, yet have not identified one that alone is capable of inducing increased lipid production. To identify other gene products that may induce anaerobic lipid production, we have performed a genome-wide screen, using the lipophilic dye Nile Red, for mutations that increase lipid production under aerobic growth conditions. Several such lesions have been identified, and we are currently investigating how these gene products impinge on membrane lipid accumulation. Once characterized, these regulatory mechanisms will be used to increase fatty acid production in this and other microbial systems for the production of commercially relevant lipid products. These studies will provide critical insight into the regulation and requirements of bacterial lipid biosynthesis and accumulation; knowledge that will direct the genetic engineering of microbes for increased yield of fatty acid-derived biofuels to help alleviate petroleum-dependence.

The Great Lakes Bioenergy Research 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.

50

Using Genomic Investigations of the Photosynthetic Bacterium *Rhodobacter sphaeroides* to Determine the Origin of the Increased H_2 Production Capacity of a Mutant Strain

Wayne S. Kontur[1,3]* (wskontur@wisc.edu), Wendy Shackwitz,[4] Joel Martin,[4] Shweta Deshpande,[4] Hope Tice,[4] Kurt LaButti,[4] Natalia Ivanova,[4] **Daniel R. Noguera**[2,3] (noguera@engr.wisc.edu), and **Timothy J. Donohue**[1,3] (tdonohue@bact.wisc.edu)

[1]Department of Bacteriology and [2]Department of Civil and Environmental Engineering, University of Wisconsin–Madison; [3]DOE Great Lakes Bioenergy Research Center, Madison, Wis.; and [4]DOE Joint Genome Institute, Walnut Creek, Calif.

Project Goals: Our main goal is to use the purple nonsulfur photosynthetic bacterium *Rhodobacter sphaeroides* as a model for studying cellular electron flow in photosynthetic bacteria. Building off of our understanding of the distribution of electrons into various cellular pathways, we ultimately aim to develop *R. sphaeroides* and other photosynthetic bacteria into platforms for producing biofuels, including hydrogen gas (H_2). Here, we use microbiological and genomic analyses to investigate and characterize a high H_2-producing derivative (strain Ga) of wild-type strain *R. sphaeroides* 2.4.1.

Rhodobacter sphaeroides is a metabolically diverse purple nonsulfur photosynthetic bacterium. During anaerobic photoheterotrophic growth (with light as an energy source and an organic substrate as a carbon and electron source), *R. sphaeroides* can produce hydrogen gas (H_2) via its nitrogenase enzyme. The wild-type strain 2.4.1 is capable of channeling up to 40% of the electrons it consumes from organic substrates into H_2, depending on the organic substrate.

We seek to identify cellular pathways that compete with H_2 production for electrons, either by recycling 'excess' electrons that are not utilized in primary biosynthetic pathways, or in the production of cellular components that are not necessary to the cell in an industrial H_2 production setting. Examples of these types of cellular pathways are CO_2 fixation via the Calvin Cycle, which the cell uses to recycle electrons even in the presence of a fixed carbon source, and the production of polyhydroxybutyrate (PHB), a carbon and electron storage polymer. Mutants in which these pathways have been inactivated show increased H_2 production capacities.

We find that strain Ga, an uncharacterized mutagenized derivative of wild-type 2.4.1, channels electrons into H_2 more efficiently than 2.4.1 (~56% of the electrons from succinate are converted into H_2 by Ga, versus ~35% of the electrons from succinate for 2.4.1). Ga grows more slowly and reaches lower final cell densities than 2.4.1, which suggests that some of the additional electrons channeled into H_2 by Ga may be coming at the expense of cellular biomass production.

In collaboration with the DOE's Joint Genome Institute, we have sequenced the genome of strain Ga. We find that 18 genes contain an insertion or deletion resulting in a frame shift between 2.4.1 and Ga. There are 56 additional genes that contain a single nucleotide polymorphism between 2.4.1 and Ga. We also find that a large (>38 kb) region, containing a number of viral associated genes, is missing from the Ga genome.

A comparison of the 2.4.1 and Ga genomic sequences has helped to identify the genomic origins of other differences between 2.4.1 and Ga not related to H_2 production. For example, mutations that result in inactive forms of methoxyneurosporene dehydrogenase (CrtD, an enzyme involved in carotenoid biosynthesis) and of fructokinase help explain, respectively, differences in pigmentation between the strains, and the deficiency of Ga in utilizing fructose as an organic substrate. The other genetic differences between the strains are being investigated to determine the origin(s) of Ga's heightened H_2 production capacity.

BioEnergy Science Center (BESC)

51

BioEnergy Science Center Education and Outreach—Overview

S. Kral,[1,4] J. Griffin,[1] H. Schulson,[1] S. Fowler,[3,4] B. Davison,[3,4]* (davisonbh@ornl.gov), **J. Westpheling**[2,4]* (janwest@uga.edu), W. Robinson,[1]* and **Paul Gilna**[4] (BESC PI)

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

Our Center has taken a novel approach to education and outreach in that our education efforts begin with 5th graders. This is 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. 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. They 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. These lessons were piloted in schools in Georgia and Tennessee and were made available to schools nationwide in the spring of 2010. We have piloted a series of "science night" programs offered to students and the general public through local schools, museums and community centers and have reached more than 35,000 students, teachers and parents. In addition, we have developed educational programs in the form of games that teach strategies for energy use. Students are allowed to design their own cars and select types of fuel to travel to familiar destinations. The games, exhibited in kiosks to be placed in schools, museums and other educational venues, incorporate lessons to explore fuel efficiency, fuel availability (for example electric and E85 cars) as well as environmental impact. A prototype kiosk has been built and was exhibited at a recent ASTC meeting attracting interest

* Presenting author

from the Director of Education at the Smithsonian where we hope to have an exhibit next year. A prototype of one of the travel games will be available at the meeting.

The lesson plan is available on the BioEnergy Science Center website at: http://bioenergycenter.org/besc/education/teachertools.cfm.

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.

52

BESC Knowledgebase, Tools, and Laboratory Information Management System

Mustafa Syed[1]* (syedmh@ornl.gov), Susan Holladay,[1] Tatiana Karpinets,[1] Morey Parang,[2] Guru Kora,[1] Doug Hyatt,[2] Sheryl Martin,[1] Leslie Galloway,[1] Byung Park,[1] Ying Xu,[3] **Ed Uberbacher**,[1] Brian Davison,[1] and **Paul Gilna**[1] (BESC PI)

[1]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]University of Tennessee, Knoxville; and [3]University of Georgia, Athens

http://besckb.ornl.gov
http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is a multi-institutional partnership undertaking large experimental campaigns to understand and reduce recalcitrance of biomass for degradation by enzymes and organisms, and converting plant biomass into biofuels in a single step. The BESC Knowledgebase and associated tools is a discovery platform for bioenergy research. It consists of a collection of metadata, data, and computational tools for data analysis, integration, comparison and visualization for plants and microbes in the center. BESC Knowledgebase (KB) and BESC Laboratory Information Management System (LIMS) enable bioenergy researchers to perform systemic research.

A unified way of sharing experimental workflows, standardized protocols, tracking samples, and capturing metadata/data across more than a dozen BESC institutions is critical to the proper functioning of BESC both from the standpoint of ensuring meaningful and reproducible results, as well as tracking the generation of intellectual property. The BESC LIMS has been used to track metadata from data collection to analysis for over 80,000 samples from more than sixty campaigns undertaken by BESC. For each of these campaigns details such as sample generation, sample splitting and shipping, sample processing, protocols, controls, replicates, and results have been captured. The LIMS web portal is available for BESC members that provides reports for material transfer, shipping, sample history and provenance, and analytical results. The LIMS is also linked to the BESC Knowledgebase, which contains the results of experiments interpreted in the context of biological systems.

The BESC KB infrastructure consists of separate public and private platforms cohesively integrated through leveraging a loosely coupled middleware layer which facilitates information exchange and access control. While the majority of the data and tools are available for public access via its public portal, a limited set of information and tools is only accessible by the BESC researchers, via its private portal through user authentication and access control. Some omics data, tiling array data, biochemical characterization data for biomass samples, and candidate genes for analysis through the transformation and biomass characterization are private. These results are integrated with reference genomic, metabolic and other omics data obtained from public resources which form core KB.

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 consists 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. For example, our *Populus* database alone contains over 500,000 versions of gene models each of which are associated with one or more annotation data including functional predication, domains, GO terms, KEGG pathways, ortholog and in-paralog data. We have also collected rich set of omics data from external resources like NCBI, GEO and EMBL ArrayExpress related to 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 *Populus*.

The microbial core KB for microbes consists of 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. Pair-wise predictions of orthologs and inparalogs for the genomes generated using bi-directional BLAST hits and Inparanoid software, BLAST hits against NCBI NR database. Trans membrane protein predictions using TMHMM and signaling proteins predictions using SignalP. We also have omics data on effect of alcohol on cell, fermentation time course study, growth on model substrates found in lignocelluloses, growth on simple and complex sugars, biomass deconstruction, and many other experiments from several *Clostridium* and other bioenergy relevant microbes. Recent addition to microbial core includes a microbial resequencing database (ReSeqDB); we have made available resequencing data for an ethanol-adapted *Clostridium thermocellum* strain. We provide not only high confidence differences, but also processed results from several different tools (Genome Resequencing Toolkit) and comprehensive annotations of those sequence changes, position of sequence change with respect to gene and intergenic region, changed protein sequence, and probable change in gene function or regulation. SNPs indels are also visualized and made available from the genome browser. The goal of the resequencing projects is to discover genomic modifications underlying the specific phenotype of the mutant. The Genome Resequencing Toolkit reveals mutational changes in the mutant strains to understand the biological effects of each change on the mutant phenotype.

The KB is tightly integrated with computational tools for genome annotation, comparative genome analysis, data integration, data mining, analysis and visualization. Genome annotations tools such as CAT for annotating protein sequences with CAZy family, DOOR for predicting operons, BLAST, ClustalW, Muscle, and tools to build phylogenetic trees. Comparative analysis tools to compare multiple genomes in terms of pathway, enzymes, domain architecture, or sequence similarity. Browse interfaces, such as organism card, and gene card, allow the user to easily retrieve detailed information about a specific gene or organism. Search interfaces provide mechanisms to search data objects using keyword, protein, enzyme and pathways with custom filters. Integrated views for genes and organisms (gene cards/organism cards) which provide a rich set of information, omics data, annotations, and tool results. Visualization tools such as Gbrowse – genome browser, CMAP (Comparative Map Viewer) can be used to view and compare maps (genetic, physical) across multiple genomes. Layers of heterogeneous data and annotations can be overlaid on GBrowse or CMAP browsers for browsing and comparing across maps. Tools to search omics data in resources such as GEO, and ArrayExpress integrate experimental data into the knowledgebase OnDemand, making it possible to draw interactive heat maps, scatterplot, overlay omics data on pathways, build networks and identify regulatory motifs.

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.

53

Computational Biology to Target Plant Pathways and Experimental Data

Ying Xu[1,2]* (xyn@bmb.uga.edu), Eberhard Voit,[1,2] Yanbin Yin,[1,2] Wen-Chi Chou,[1,2] Shan Wang,[1,2] Xizeng Mao,[1,2] Michael Udvardi,[1,2] Maor Bar-Peled,[1,2] Jerry Tuskan,[2] Brian Davison,[2] and **Paul Gilna**[2] (BESC PI)

[1]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). 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 is working with experimental biologists on a number of collaborative projects to study plant cell wall recalcitrance. Four examples are highlighted.

To facilitate gene discovery and genomic research in switchgrass the assembly of ~11 million switchgrass, ESTs were sequenced by 454 technologies and ~700k Sanger ESTs into a "Switchgrass gene index." The data set contains 139,200 assembled UniGenes, based on which an Affymetrix cDNA chip was designed and manufactured. A systematic annotation of the "gene index" set was performed by searching against CDD, Panther and CAZy databases.

A novel mathematical modeling approach was developed to investigate the lignin biosynthesis pathways by analyzing different omics datasets in wild type and transgenic plants. The modeling effort led to two novel postulates regarding the control of the lignin biosynthetic pathway, which were partially validated by subsequent laboratory experiments. (Lee, et al., 2011)

A systematic bi-clustering co-expression analysis using *Arabidopsis* microarray data (with >1,300 experimental conditions) and ~800 annotated cell wall-related (CWR) proteins was completed. This analysis identified ~2,000 new proteins co-expressed with annotated CWR genes, forming 217 co-expression modules. Conserved regulatory motifs were also predicted for genes from the same co-expression modules. Based on this result, three BESC partners have suggested over 50 genes to be verified within by our coordinated plant genetic transformation.

The dbCAN database and webserver (http://csbl.bmb.uga.edu/dbCAN/index.php) is now automated for CAZy annotation. Member proteins and literature of each CAZy family are analyzed and manually curated to extract a signature domain to represent the family. A hidden markov model (HMM) was build based on the alignment of all the signature domains in the member proteins for the HMM to specifically represent the corresponding CAZy family. With the HMMs, dbCAN provides the annotation service so

* Presenting author

users can upload their proteins for signature domain-based CAZy annotation.

Reference

1. Lee, Y., et al., PLoS Computational Biology, 7(5), 2011.

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.

54

Enhanced Pathway Visualization in Pathway-Genome Database (PGDB) Captures Subcellular Localization of Metabolites and Enzymes: The Nucleotide-Sugar Biosynthetic Pathways of *Populus trichocarpa*

Ambarish Nag,[1,2] Tatiana Karpinets,[2,3] Christopher H. Chang,[1,2] Maor Bar-Peled,[2,4]* (mao@ccrc.uga.edu), **Debra Mohnen**,[2,4] and **Paul Gilna**[2] (BESC PI)

[1]National Renewable Energy Laboratory, Golden, Colo.; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]University of Tennessee, Knoxville; 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). 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 genomic tools for detailed studies into poorly understood cell wall biosynthesis pathways.

Understanding how cellular metabolism works and is regulated requires that the underlying biochemical pathways be adequately represented and integrated with large metabolomic data sets to establish a robust network model. Genetically engineering energy crops to be less recalcitrant to saccharification requires detailed knowledge of plant polysaccharide structures and a thorough understanding of the metabolic pathways involved in forming and regulating cell wall synthesis. Nucleotide-sugars are building blocks for synthesis of cell wall polysaccharides. The biosynthesis of nucleotide-sugars is catalyzed by a multitude of enzymes that reside in different subcellular organelles, and precise representation of these pathways requires accurate capture of this biological compartmentalization. The lack of simple localization cues in genomic sequence data and annotations however leads to missing compartmentalization information for eukaryotes in automatically generated databases, such as the Pathway-Genome Databases (PGDBs) of the SRI Pathway Tools software that drives much online biochemical knowledge representation today. In this report, we provide an informal mechanism using the existing Pathway Tools framework to integrate protein and metabolite sub-cellular localization data with the existing representation of the nucleotide-sugar metabolic pathways in a prototype PGDB for *Populus trichocarpa*. The enhanced pathway representations have been successfully used to map SNP abundance data to individual nucleotide-sugar biosynthetic genes in the PGDB. The manually curated pathway representations are more conducive to the construction of a computational platform that will allow the simulation of natural and engineered nucleotide-sugar precursor fluxes into specific recalcitrant polysaccharide(s).

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.

55

A Workflow for Archiving and Integrating Diverse Data for Structural Characterization of Biomass

William S. York[1,2] (will@ccrc.uga.edu), Malcolm O'Neill,[1,2] Trina Saffold,[1,2] Mehdi Allahyari,[1] Breeanna Urbanowicz[1,2]* (breeanna@uga.edu), Maria Peña,[1,2] and **Paul Gilna**[2] (BESC PI)

[1]The 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 number of natural and modified plant samples as well as developing genomics tools for detailed studies into poorly understood cell wall biosynthesis pathways. 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.

Terrestrial biomass is a highly complex material comprised primarily of secondary plant cell walls. The molecular structures of the biopolymers that constitute these cell walls vary depending on a large number of factors, including the plant species and the genetic makeup of the plant as well as growth, harvest, drying and storage conditions. The observable differences include variation in cell wall composition (i.e., relative proportions of various biopolymers), the molecular structures of these biopolymers and the overall infrastructure of resulting composite. These differences, in turn, lead to differential susceptibilities of the individual cell wall components to chemical and/or enzyme-catalyzed solubilization and saccharification. Although complete structural and ultrastructural characterization of the cell walls contained in different biomass samples is beyond our current technological capabilities, considerable structural information can be obtained using a wide range of spectroscopic, chemical and immunochemical methods. Each of these methods provides unique information, and interpretation of the data generated by one method almost always requires information provided by another method. Therefore, we have designed a workflow to collect diverse types of complementary analytical data from each biomass sample under study and archive this data in a dedicated database that facilitates integration and analysis of the global data set. This will provide an experimental basis for interpretation of each data set and the identification of correlations (e.g. covariance) among the different data sets. For example, changes in the molecular structure or amount of a specific polysaccharide might be correlated to changes in the capacity of a specific antibody to bind to the biomass, suggesting that antibody binding depends on the specific structure that changed. Correlation of such data to measured variations in biomass recalcitrance should lead to testable hypotheses regarding the molecular basis for recalcitrance and the development of genetic or agronomic approaches to modulate recalcitrance.

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.

56

^{13}C Labeling and NMR Analysis: Critical Tools in the Development of Next Generation Biofuel Platforms

Marcus Foston,[1] Reichel Samuels,[1] Rui Katahira,[2] Erica Gjersing,[2] Mark Davis,[2] Heather L. McKenzie,[3] Charles E. Wyman,[3] **Arthur J. Ragauskas**[1]* (Arthur.ragauskas@chemistry.gatech.edu), and **Paul Gilna**[4] (BESC PI)

[1]School of Chemistry and Biochemistry, Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta; [2]BioEnergy Science Center, National Renewable Energy Laboratory, Golden, Colo.; [3]Department of Chemical and Environmental Engineering, Center for Environmental Research and Technology, University of California, Riverside; 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). 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 economics of next generation biorefineries and biofuel, biomaterial production is extremely dependent on the overall yields and energy/material cost associated with biomass deconstruction. Therefore, it is important to gain a detailed understanding of the physico-chemical structure of the cell wall, mechanisms of efficiently overcoming biomass recalcitrance and methods of biomass deconstruction/conversion. A major component of our recent research program has been the development of a NMR toolkit utilizing partial or uniform ^{13}C isotopic enrichment. The elegance of this system resides in the fact any analysis of conversion processes or biomass-derived material will also benefit from this labeling. Untreated and pretreated corn stover stems along with the resulting continuous flow pretreatment reactor effluent were subjected to various ^{13}C 1D and 2D solution and solid-state NMR techniques to demonstrate the usefulness of ^{13}C labeling in biofuel research including high-throughput screening, *in-situ* destruction studies, metabolite analysis and cell wall 3D ultrastructural characterization.

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.

57

High- and Medium-Throughput Characterization of Biomass to Understand Causes of Recalcitrance in Plant Cell Walls

Angela Ziebell[1]* (Angela.Ziebell@nrel.gov), Robert Sykes,[1] Steve Decker,[2] Erica Gjersing,[1] Crissa Doeppke,[1] Geoff Turner,[2] Cody Law,[2] Melissa Carlile,[1] **Mark Davis**,[1] and **Paul Gilna**[3] (BESC PI)

[1]National BioEnergy Center, National Renewable Energy Laboratory, Golden, Colo.; [2]BioSciences Center, National Renewable Energy Laboratory, Golden, Colo.; and [3]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://www.bioenergycenter.org

* Presenting author

Project Goals: 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. BESC characterizes samples in order to identify low recalcitrance lines of switchgrass and *Populus*. This collaboration with plant biologists provides understanding how and why plants low recalcitrance lines are easier to process.

The high-throughput (HTP) pipeline has analyzed tens of thousands of samples using the HTP pipeline which characterizes cell wall chemistry and sugar release on pretreatment and enzymatic saccharification (Sykes et al. 2008; Selig et al. 2011). In the last year, we have added two new capabilities, starch removal and rapid NMR analysis, and worked with plant biologists to interpret their data.

The characterization group works together with plant biologists to understand the chemical modifications occurring when plants are genetically modified and why they might cause changes in recalcitrance to pretreatment and saccharification. The characterization group also studies natural populations to understand how natural populations vary and what we might learn from natural diversity. Below (Figure 1) is an example of how the HTP data is used to study natural populations.

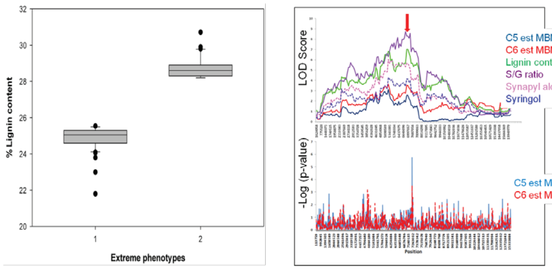


Figure 1: Data from a study of 800 naturally occurring *Populus* from the Pacific Northwest. Left extreme lignin phenotypes, right genetic mapping finds results from both the cell wall chemistry and the sugar release assay coincide.

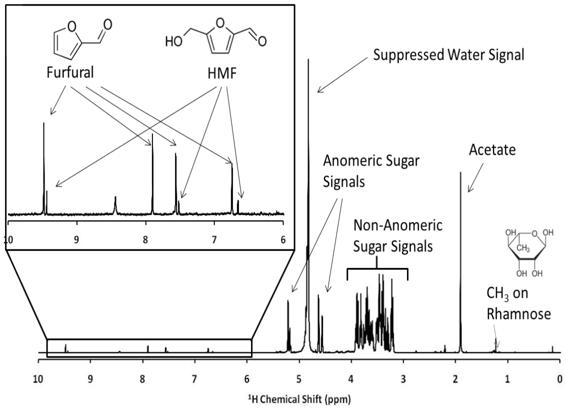


Figure 2: Example NMR spectrum from HTP sugar analysis.

Plant cell wall sugar content is being determined by a HTP NMR (Nuclear Magnetic Resonance) adaptation to NREL's traditional wet chemistry analysis. This technique is also being used to assess the tendency of a plant to produce inhibitors e.g. furfural and HMF. The method uses proton NMR (Figure 2) and Partial Least Squares analysis in conjunction with a set of standards which have the full wet chemistry method. The current NMR facility houses a HTP NMR sample changer which can house >500 samples at a time and sample analysis time take 1-2 minutes. The original full wet chemistry method has also been down scaled to 1/300th by researchers at the University of California at Riverside and is available for high-throughput analysis (DeMartini et al. 2011).

References

1. DeMartini, J. D. et al. (2011). "Small-scale and automatable high-throughput compositional analysis of biomass." *Biotechnol. Bioeng.* 108(2): 306-312.
2. Selig, M. et al. (2011). "High-throughput determination of glucan and xylan fractions in lignocelluloses." *Biotechnol. Lett*: 1-7.
3. Sykes, R. et al. (2008). "Within tree variability of lignin composition in *Populus*." *Wood Sci. Technol.* 42: 649-661.

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.

58

High Resolution Characterization of Biomass

Ali Passian* (passianan@ornl.gov), Laurene Tetard, **Brian Davison**, and **Paul Gilna** (BESC PI)

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.

BESC has developed a method to provide chemical information about biomass at the scale of tens of nanometers. When employing absorption spectroscopy as a means for obtaining chemical information from samples in bulk form, one is faced with an inherent limitation on the spatial resolution. This lack of resolution, which is caused by the classical diffraction limitation of the excitation and detection in conjunction with the delocalized nature of the nonradiative decay, has hindered studies that require spectral information on sub-wavelength scale or more importantly simultaneous spectral and topographic information. An important example of such studies is the understanding of the plant cell wall organization at the nanoscale. In order to overcome recalcitrance so as to reach higher levels of efficiency for biofuel production that would make lignocellulosic biomass a viable option for the next generation of sustainable energy solutions, it is essential to study the cell spatio-chemically. In an effort to achieve this and thus extend the capabilities of the variety of existing analytical tools available to characterize biomass, we introduced Mode Synthesizing Atomic Force Microscopy (MSAFM), and further proposed to augment it with spectroscopic capabilities. The resulting system aims to nano mechanically reduce the effects of both the diffraction limitation and the nonlocalized energy landscape of the sample on the resolution with which chemical and morphological information can be obtained. Preliminary results show that it is possible to probe the various spatio-chemical properties of biomass and furthermore reach subsurface information. The extension of MSAFM into a hybrid photonic-MSAFM (hp-MSAFM) and the capitalization of the rich underlying dynamics (such as the recently discovered virtual resonance) can provide chemical information of biomass. The results include localized spectroscopic measurements performed with the MSAFM probe, using a quantum cascade laser (QCL) and broadband sources (FTIR) of infrared radiation. We will demonstrate that MSAFM can also be used to provide chemical mapping of biomass at a given wavelength, and will present our results on *Populus* stem cross section samples subjected to a sequence of chemical treatments (from Soxhlet extraction to acid treatment).

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.

59

How Do Bacteria Eat Biomass? Multi-scale Imaging and Modeling of Cells, Cell Walls, and Cellulosomes

Michael F. Crowley[1,2]* (michael.crowley@nrel.gov), Bryon S. Donohoe,[1,2] Peter N. Ciesielski,[1,2] Yannick J. Bomble,[1,2] Loukas Petridis,[2] Jeremy C. Smith,[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.

Some bacteria use a complex of macromolecular structures that are bound to both the bacterial cell and cell walls of biomass and contain the molecular machinery for digesting biomass and utilizing the products, the Enzyme-Microbe-Substrate Interface (EMS interface). An understanding of the structure and functioning of the EMS interface enhances our ability to tune its components as well as the biomass target. We use our insights to suggest modifications of biomass structure (genetic modifications of plants) and pretreatment processes for optimal conversion of biomass to fuel precursors and modifications of cellulosomal composition and structure for enhanced interaction and degradation of the modified cell walls. We present the Electron Tomography of the EMS interface that produces three-dimensional volume renderings of objects at the 3-5 nanometer scale, appropriate for the macromolecular structures found in the EMS interface and we present the Molecular Modeling of EMS interface components such as cellulosomes and enzymes and of cell wall components such as cellulose, hemicellulose, and lignin. The modeling contributes to interpretation of tomographic images, testing hypotheses of mechanisms in the EMS interface, and proposing new hypotheses in the same way that the tomography can validate theoretical findings of modeling. Recent successes in our research have brought together the length scales of electron tomography and molecular modeling making new insights into microbial interactions with biomass possible.

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.

60

Real-Time Imaging of Plant Cell Wall Structure at Nanometer Scale, with Respect to Cellulase Accessibility and Degradation Kinetics

Yu-San Liu,[1,2] Yining Zeng,[1,2] John O. Baker,[1] **Shi-You Ding**[1,2]* (Shi.you.Ding@nrel.gov), and **Paul Gilna**[2] (BESC PI)

[1]Biosciences 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 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.

Cost-effective production of lignocellulosic biofuels is contingent upon deeper understanding of the structure and chemistry of plant cell walls, as well as of the molecular basis of the conversion processes such as chemical pretreatment and enzymatic saccharification. Label-free and real-time imaging approaches, such as coherent Raman scattering and single molecule tracking, are employed for the first time to investigate plant cell wall structure at nanometer scale, spatial distribution of chemical constituents of the cell wall, and kinetics of degradation by fungal "free" cellulase and bacterial cellulosome systems *in situ*. Specific objectives of these studies are to determine (1) structures and spatial arrangements of microfibrils in plant cell walls, and how they affect accessibility to cellulases and ultimate digestibility, (2) differences in the relationships of cellulose and lignin in different types of plant cell walls, and the ways in which these differences affect the efficiency of pretreatment and enzyme hydrolysis, and (3) the locations at which cellulases bind to the cell wall, in relation to the manner in which individual microfibrils are hydrolyzed. Preliminary (published and unpublished) results will be presented and discussed (1-6).

References (recent publications supported by this project)

1. Liu Y.-S., et al., Real-time imaging contrasts modes of biomass degradation by fungal cellulase and bacterial cellulosome. In preparation.
2. Foston M., et al. 2011, Chemical, ultrastructural and supramolecular analysis of tension wood in *Populus* as a model substrate for reduced recalcitrance. *Energy Environ. Sci.* DOI: 10.1039/c1ee02073k.
3. Yang B., et al. 2011, Enzymatic Hydrolysis of Cellulosic Biomass: A Review. *Biofuels* 2:421-450.
4. Liu Y.-S., et al. 2011, Cellobiohydrolase Hydrolyzes Crystalline Cellulose on Hydrophobic Faces. *J. Bio. Chem.* 286:11195-11201.
5. Dagel D.J., et al., 2011, DOPI and PALM Imaging of Single Carbohydrate Binding Modules Bound to Cellulose Nanocrystals. SPIE Proc. *Proc. SPIE* 7905, 79050P. doi:10.1117/12.875285.
6. Dagel D. J., et al. 2010, *In situ* Imaging of Single Carbohydrate-Binding Modules on Cellulose Microfibrils. *J. Phys. Chem. B.* 115:635-41.

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.

61

A Novel APAP1 Glycoconjugate Indicates a New Role for Pectin in Plant Cell Walls

Li Tan[1,3]* (tan@ccrc.uga.edu), Melani Atamodjo,[1,3] Stefan Eberhard,[1,3] Sivakumar Pattathil,[1,3] Xiang Zhu,[1,3] Michael G. Hahn,[1,3] Alan Darvill,[1,3] Marcia Kieliszewski,[2] **Debra Mohnen**,[1,3] and **Paul Gilna**[3] (BESC PI)

[1]The University of Georgia, Athens; [2]Ohio University, 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 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 genomic tools for detailed studies into poorly understood cell wall biosynthesis pathways.

Plant cell walls are a renewable resource that consists of roughly 10% protein and 90% of an interacting carbohydrate network of cellulose, hemicellulose, and pectin. We report the detailed structural characterization of an *Arabidopsis* glycoconjugate, purified from *Arabidopsis* cell culture medium and presented in plant cell walls. N-terminal sequencing, chemical, enzymatic and glycome profiling analyses revealed that the glycoconjugate contains covalently-attached protein and glycan domains previously attributed to independent wall components. Multiple *apap1* T-DNA insertion mutants exhibit modified cell wall and growth properties. This glycoconjugate structure, named APAP1, has broad implications for the synthesis, structure and function of wall components, especially for roles of pectins in plant walls. APAP1 structure supports *gaut* mutant studies which showed that insertion mutation of 13 of the 15 *Arabidopsis GAUT* genes caused significant changes of major wall monosaccharides such as xylose, glucose, mannose, fucose, arabinose, galactose, and rhamnose. Recent results showing that pectins are synthesized in both primary and secondary walls indicate a role for pectins in both herbaceous and woody plant biomass species. The combined results suggest that pectins not only serve as a negatively charged matrix in the wall, but also function as a cross-linker that holds different polysaccharides, especially hemicelluloses, in the wall. The ramification of the results for reducing plant biomass recalcitrance will also be discussed.

Funding from National Science Foundation NSF-MCB 0646109, DOE center grant DOE DE-FG02-09ER20097 and BioEnergy Science Center

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.

62

Genome-Scale Discovery of Cell Wall Biosynthesis Genes in *Populus*

Steve P. DiFazio[1,2]* (spdifazio@mail.wvu.edu), Wellington Muchero,[2] G.T. Slavov,[1] J. Martin,[3] W. Schackwitz,[3] E. Rodgers-Melnick,[1] C.P. Pennacchio,[3] U. Hellsten,[3] L. Pennacchio,[3] L.E. Gunter,[2] P. Ranjan,[2,4] D. Rokhsar,[3] **G.A. Tuskan**,[2,4] and **Paul Gilna**[2] (BESC PI)

[1]West Virginia University, Morgantown, W.Va.; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [3]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, 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). 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 number of natural and modified plant samples as well as developing genomic tools for detailed studies into poorly understood cell wall biosynthesis pathways.

The primary goal of the *Populus* Activity in BESC has been the identification of genes controlling cell wall formation which ultimately positively impact sugar release, i.e., overcoming recalcitrance. The quantitative genomics portion of the project has focused on Quantitative Trait Locus (QTL) analysis in two large interspecific families, and association genetics to mine natural variation in *Populus trichocarpa*. Specifically, we established two QTL populations, one in West Virginia and one in eastern Oregon and have created a saturated genetic map containing >6000 genetic markers. This map was used to identify regions of the *Populus* genome that control sugar release. Six such regions were found and, in combination with the transcript profiling and association genetics results, six genes within these regions have been verified as improving sugar release. In addition, in the association mapping study we collected 1,100 genotypes from across the native range of *Populus trichocarpa*, established clonal replicates of each genotype in three common gardens in the Pacific Northwest and subjected two-year-old samples from the Corvallis, OR site to the high throughput compositional and recalcitrance phenotyping pipeline established at NREL. Simultaneously, we created a 36,000 single nucleotide polymorphisms (SNP) genotyping Infinium chip based on resequencing data generated by JGI from 15 alternate *P. trichocarpa* genotypes. This SNP array was used to interrogate all 1,100 genotypes found in the common gardens. Association genetics statistical approaches were used to identify specific SNP within specific genetic loci that controlled sugar release and other relevant cell wall phenotypes. From this analysis we identified 46 genes and their amino acid substitutions that are controlling the phenotypes measured in this population. The average increase in sugar yield associated with each SNP is approximately 26% above the wild type control. These genes have been initiated within the *Populus* transformation pipeline managed by ArborGen. We are continuing to phenotype the QTL and association populations for wood chemistry and sugar release, as well as a wide array of traits that will impact productivity in addition to recalcitrance, thereby paving the way for follow-on studies and commercialization during the next phases of the project.

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.

63

Comparative Glycomics Provide Insights into Cell Wall Components that Affect Biomass Recalcitrance

Michael G. Hahn[1,2]* (hahn@ccrc.uga.edu), Sivakumar Pattathil,[1,2] Shishir P.S. Chundawat,[3] Jaclyn D. DeMartini,[1,4] Jeffrey Miller,[1,2] Virginia Brown,[1,2] Sindhu Kandemkavil,[1,2] Bruce E. Dale,[3] **Charles Wyman**,[1,4] and **Paul Gilna**[1] (BESC PI)

[1]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge Tenn.; [2]The University of Georgia, Athens; [3]Great Lakes Bioenergy Research Center, East Lansing, Mich.; 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

* Presenting author

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

Plant biomass, which is considered a primary feedstock for sustainable biofuel production, is largely composed of plant cell walls, whose principal components are polysaccharides (e.g., cellulose, hemicellulose, pectin) and lignin. Plant cell walls are thus complex heteropolymeric composites that have evolved to be resistant to deconstruction. In addition, cell walls are different in their structure and composition in different plants. Achieving efficient, economic viable and ecologically sustainable biofuel production will require an in-depth understanding of cell wall structure/dynamics in the various plant species being considered as biomass feedstocks. We have developed a moderate throughput approach, called Glycome Profiling that can be used for comparative glycomic analyses of various biomasses to determine the glycan composition/make up of cell walls and also to identify cell wall components that affect biomass recalcitrance. Glycome Profiling takes advantage of the availability of a large and diverse collection of cell wall glycan-directed monoclonal antibodies that can monitor most major plant polysaccharides. We report here on the use of Glycome Profiling to monitor changes in cell wall composition and extractability that result from hydrothermal and ammonia fiber expansion (AFEX) pre-treatments of biomass from diverse plants. These studies suggest that xylans play a critical role in governing recalcitrance in grasses, while lignin plays a major role in poplar biomass recalcitrance. Furthermore, we have demonstrated that the two pre-treatments have very different effects on the biomasses examined. Hydrothermal pre-treatments rapidly cleave carbohydrate-lignin associations and destroy arabinogalactan structures in the wall. AFEX, on the other hand, results in changes to overall wall structure that result in the more facile release of sub-populations of hemicelluloses (xyloglucans and xylans) and pectic arabinogalactans from the pre-treated walls, with less destruction of wall glycan epitopes. The effects of both pre-treatments vary depending on plant source of the biomass. Experiments are underway to correlate the observed cell wall changes to improve sugar release from the treated 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.

64

Comparative Proteomics Unveils Functional Signatures of Cellulose Formation (*Populus*) and Deconstruction (Cellulolytic Microorganisms) at a Cellular Level

Richard J. Giannone* (giannonerj@ornl.gov), Paul Abraham, Rachel Adams, Andrew Dykstra, **Gerald A. Tuskan**, Robert L. Hettich, and **Paul Gilna** (BESC PI)

BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://bioenergycenter.org

Project Goals: 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 availability of complete genome sequences for both plant model systems (such as *Populus*) as well as cellulolytic bacteria has provided a means to further unveil the functional intricacies of *in situ* biomass degradation at an unprecedented level. Our current work not only develops and demonstrates advanced "shotgun" mass spectrometry (MS) techniques that dive even deeper into the proteomes of relevant bioenergy-related organisms, but also furthers our understanding of the molecular machinery employed by both plants and microorganisms as they either establish (plant) or degrade (microbe) a variety of simple and complex biomass feedstocks. To accomplish these goals, we have established a robust, high-throughput approach for the comprehensive quantitative analysis of deep LC-MS measurements utilizing both label-free and stable isotope-labeling strategies. In particular, we have examined the enzymatic features of the cellulolytic apparatus both shared between and unique to several different cellulose-degrading bacteria (i.e., *C. thermocellum*, *C. bescii* and *C. obsidiansis)* over the course of fermentative growth on simple (e.g., Avicel, cellobiose) and complex (e.g., *Populus*, switchgrass) feedstocks alike. For these microbial systems, we have characterized cellulolytic machinery for both cell-attached (i.e., cellulosome) and cell-free enzymatic systems, with a particular focus on multi-domain glycosidases, extracellular solute binding proteins, and uncharacterized proteins, all of which remain vitally important to sugar release and utilization from biomass.

More recently, we have demonstrated a robust quantitative methodology utilizing selected reaction monitoring (SRM) on a triple quadrupole MS to pursue absolute quantification of targeted proteins. Employing SRM as a quantification tool enabled the accurate stoichiometric determination of key structural and enzymatic constituents of *C. thermocellum*'s cellulosome. These measurements enhance our understanding of the link between absolute cellulosome abundance and overall enzymatic activity—a result which could help direct future molecular engineering endeavors. This MS technique has several advantages over ELISA-based quantification methods including high-precision, ease of use, high-throughput, as well as the ability to quantify proteins in samples ranging in proteomic complexity (purified cellulosomes to whole-cell lysates).

Though much work thus far has investigated biomass deconstruction, recent efforts in our lab have focused on uncovering the mechanisms of biomass formation. In this regard, we have developed an experimental approach that combines an enhanced cell lysis and proteome solubilization protocol with state-of-the-art LTQ-Velos MS technology to achieve the deepest proteome coverage of *Populus* to date. In response to the enhanced genomic complexity inherent to eukaryotic species, especially plants, we have optimized our post-data acquisition informatics to resolve ambiguous protein identifications resulting from the substantive genetic redundancy exhibited by *Populus* spp. Clustering redundant proteins prior to mapping peptides to the proteome allowed us to more thoroughly characterize *Populus* organ-specific protein information, focusing on comparing and contrasting the proteomes of leaf (young and old), stem, and root samples. Taken together, these MS-based proteome approaches provide remarkable insight into the fundamental mechanisms of biomass formation by plants as well degradation by cellulolytic microbes as a step to enhance biofuel 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.

65

Development of a High-Throughput Genetic Transformation System for Switchgrass (*Panicum virgatum* L.)

Zeng-Yu Wang[1,2]* (zywang@noble.org), Chunxiang Fu,[1,2] Hiroshi Hisano,[1,2] Frank Hardin,[1,2] Xirong Xiao,[1,2] **Richard A. Dixon**,[1,2] and **Paul Gilna**[2] (BESC PI)

[1]The Samuel Roberts Noble Foundation, Ardmore, Okla. 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 number of natural and modified plant samples as well as developing genomic tools for detailed studies into poorly understood cell wall biosynthesis pathways.

Switchgrass (*Panicum virgatum*), a native C4 perennial grass throughout North America, is an excellent candidate for the production of cellulosic biofuels. Genetic improvement of switchgrass through biotechnological approaches is expected to play a crucial role in modifying quality or quantity of biomass suitable for biofuel production. Development of genetic tools is essential for effective improvement of existing switchgrass cultivars. Switchgrass, like many other grasses, is generally considered difficult to genetically manipulate at the cellular level. The low transformation efficiency has been recognized as a bottleneck in genetic manipulation and functional test of transgenes in grasses. The establishment of a well defined, rapid and highly efficient genetic transformation system is an important prerequisite for genetic engineering of this species. We have successfully solved this bottleneck problem in switchgrass and established a high throughput system for the production of large numbers of transgenic plants.

By identification of highly tissue culture responsive genotypes and by optimization of transformation parameters, we have developed a highly efficient genetic transformation system for the widely used switchgrass cultivar Alamo. Embryogenic calli were induced from immature inflorescences or from seeds. Overexpression vectors and RNAi vectors were constructed and transferred into *Agrobacterium tumefaciens* strain EHA105 or AGL1. The hygromycin phosphotransferase (*hph*) gene was used as a selectable marker. After co-cultivation with *Agrobacterium,* the infected calli were transferred onto selection medium containing the antibiotic hygromycin. Resistant calli obtained after about six weeks of selection were transferred to regeneration medium. Regenerated green shoots were transferred to rooting medium, and the rooted plantlets were later transferred to the greenhouse. The timeline from callus infection to rooted plants was about 4 months. Regenerated plants were screened by PCR analysis. Stable integration of the transgenes into the plant genome was confirmed by Southern blot analysis. The transformation efficiency (number of independent transgenic plants/number of calli used for infection) reached more than 90%. Moreover, the system is consistent and highly reproducible. We easily produced more than 800 independent transgenic plants in 6 months. Our high throughput transformation system offers a solid basis for functional analysis of large numbers of genes in switchgrass.

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.

66

Cell Wall Characteristics and Ethanol Fermentation Studies of PvMYB4-Over-Expressing Switchgrass Plants

Hui Shen[1,4]* (hshen@noble.org), Charleson R. Poovaiah,[2,4] Angela Ziebell,[3,4] Erica Gjersing,[3,4] Robert Sykes,[3,4] Fang Chen,[1,4] Jonathan R. Mielenz,[4] Timothy J. Tschaplinski,[4] Mark Davis,[3,4] Neal Stewart Jr.,[2,4] **Richard A. Dixon**,[1,4] and **Paul Gilna**[4] (BESC PI)

[1]The Samuel Roberts Noble Foundation, Ardmore, Okla.; [2]University of Tennessee, Knoxville; [3]National Renewable Energy Laboratory, Golden, Colo.; 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.

Overcoming cell-wall recalcitrance for cellulosic ethanol production is the identifying focus of the DOE BioEnergy Science Center. Our previous studies showed that lignin content and wall-bound phenolic ratios had significant impacts on cell wall saccharification efficiency in switchgrass biomass. The aromatic monolignol monomers and wall-bound phenolics share many common biosynthetic reactions, which are genetically controlled by sets of positive and negative transcriptional regulators. Over-expression of a newly characterized transcriptional repressor, *PvMYB4*, dramatically increases, by around 300%, the sugar release efficiency from cell wall residues. To gain more detailed understanding of cell wall recalcitrance and potential impacts on cellulosic ethanol production of utilizing these transgenic lines, we conducted a series of studies utilizing the BESC feedstock characterization pipeline. The results show that, although PvMYB4 over-expressing lines give about three fold enhanced ethanol yields without pretreatment, the total sugar levels of the biomass are the same. Generation of potential inhibitors of fermentation by the thermophile, *C. thermocellum*, during consolidated bioprocessing (CBP) was evaluated under conditions of hot-water pretreatment. Solid-state NMR was used to investigate the changes in the characteristics of the plant cell walls between the transgenic and control plants. Gel permeation chromatography of isolated ball-milled lignin was used to determine the average lignin molecular weight. Together, the data indicate that the aromatic acids embedded in the cell-wall, the lignin content and the lignin polymer size all likely have significant impacts on saccharification efficiency and cellulosic ethanol production. This case study of PvMYB4-overexpressing switchgrass lines demonstrates one successful strategy to overcoming cell wall recalcitrance 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.

67

Consolidated Bioprocessing Conversion of Genetically Modified Switchgrass

Kelsey Yee,[1] Choo Hamilton,[1] Miguel Rodriguez Jr.,[1] Scott Hamilton-Brehm,[1] Chunxiang Fu,[2] Zeng-Yu Wang,[2] **Jonathan R. Mielenz**[1]* (mielenzjr@ornl.gov), and **Paul Gilna**[1] (BESC PI)

[1]Bioenergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn. and [2]The Samuel Roberts Noble Foundation, Ardmore, Okla.

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. BESC research in biomass deconstruction and conversion targets CBP by studying model organisms and thermophilic anaerobes to understand novel strategies. This grand challenge calls for an integrated research approach, which is illustrated by the project described below.

Consolidated bioprocessing (CBP) conversion comprises biomass fermentative microorganisms that produce some or all of the needed biomass hydrolytic enzymes, mitigating the cost of added enzymes. It has been shown previously that down regulation of the caffeic acid O-methyltransferase (COMT) gene in switchgrass (*Panicum virgatum*) reduced lignin content, reduced S/G ratio, increased sugar release, and improved the yeast-based SSF conversion of the modified switchgrass line compared to the wild-type switchgrass (Fu, et al. 2011). As expected, after either dilute acid or hot water pretreatment of switchgrass provided by the Noble Foundation, yeast-based SSF conversion showed the COMT transgenic switchgrass yielded at least 25% more ethanol than the wild-type regardless of the pretreatment process. To follow-up this work, we examined the fermentation of both wild-type and COMT transgenic switchgrass lines using anaerobic bacteria including *Thermoanaerobacterium saccharolyticum* and very active consolidated bioprocessing (CBP) microorganisms *Clostridium thermocellum*, *Caldicellulosiruptor obsidiansis*, and *Caldicellulosiruptor bescii.* Comparison of the wild-type switchgrass to a modified COMT plant line after either dilute acid or hot water pretreatment showed that the transgenic switchgrass yielded over 20% greater total fermentation products on a (g/g) substrate basis when fermented by *Thermoanaerobacterium saccharolyticum* after addition of enzymes ala the SSF mode. Biomass fermentation by *C. thermocellum*, *C. obsidiansis*, or *C. bescii* requires no added industrial enzymes but switchgrass fermentation showed differential inhibition with wild-type and COMT transgenic switchgrass. Specifically, wild-type and COMT transgenic switchgrass were pretreated with dilute acid (0.5% H_2SO_4, 180°C, 7.5 min.) followed by water washing of the solids. Bioconversion of the solids using *C. thermocellum* showed partial inhibition of fermentation with efficient hydrolysis liberating free sugars, but only partial conversion of these sugars to fermentation products. The COMT transgenic feedstock showing greater inhibition compared to the wild-type switchgrass. However, full fermentation capabilities for *C. thermocellum* were restored after an extra hot water extraction of soluble components from the wild-type and transgenic switchgrass and the COMT line produced about 25% more total products than from the wild-type switchgrass, as seen with yeast and *Thermoanaerobacterium saccharolyticum* SSF conversion. However, using the same biomass sources processed identically with extensive water extraction, both *Caldicellulosiruptor* species failed to ferment the transgenic switchgrass (ca. 7-10% of expected) while successfully fermenting the wild-type switchgrass. This suggests *Caldicellulosiruptor* can be used as an indicator for fermentation inhibition with different biomass sources.

Reference

1. Fu et al., PNAS 108:3803-3808 (2011)

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.

68

Down-Regulation of the Caffeic Acid *O*-methyltransferase Gene in Switchgrass Reveals a Novel Monolignol Analog

Timothy J. Tschaplinski[1,5]* (tschaplinstj@ornl.gov), Robert F. Standaert,[1,4,5] Nancy Engle,[1,5] Madhavi Z. Martin,[1,5] Amandeep K. Sangha,[1,4,5] Jerry M. Parks,[1,5] Jeremy C. Smith,[1,4,5] Reichel Samuel,[2,5] Yunqiao Pu,[2,5] Art J. Ragauskas,[2,5] Choo Y. Hamilton,[1,4,5] Chunxiang Fu,[3,5] Zeng-Yu Wang,[3,5] Brian H. Davison,[1,5] Richard A. Dixon,[3,5] Jonathan R. Mielenz,[1,5] and **Paul Gilna**[1,5] (BESC PI)

[1]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Georgia Institute of Technology, Atlanta; [3]The Samuel Roberts Noble Foundation, Ardmore, Okla.; [4]University of Tennessee, Knoxville; 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). 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.

Down-regulation of the caffeic acid 3-O-methyltransferase (COMT) gene in the lignin biosynthetic pathway of switchgrass (*Panicum virgatum*) resulted in cell walls of transgenic plants releasing more constituent sugars after pretreatment by dilute acid and treatment with glycosyl hydrolases from an added enzyme preparation and from *Clostridium thermocellum*. Fermentation of both wild-type and transgenic switchgrass after milder hot water pretreatment with no water washing showed that only the transgenic switchgrass inhibited *C. thermocellum*. Gas chromatography-mass spectrometry-based metabolomics were undertaken on cell wall aqueous extracts to determine the nature of the microbial inhibitors. Metabolomic analyses of the transgenic biomass revealed the presence of a novel monolignol-like metabolite, identified as *trans*-3, 4-dimethoxy-5-hydroxycinnamyl alcohol (*iso*-sinapyl alcohol). Down-regulated COMT SWG accumulated *iso*-sinapyl alcohol, its glucoside *iso*-syringin, and putative lignan conjugates. *iso*-sinapyl alcohol is likely synthesized from its acid and aldehyde precursors that are also only evident in COMT-deficient plants.
As hypothesized, COMT-deficient plants have lowered concentrations of sinapyl alcohol and increased concentrations of phenolic acid and aldehyde inhibitors of microbial fermentation. Quantum chemical calculations were used to predict the most likely homodimeric lignans generated from dehydration reactions, but these products were not evident in plant samples. Such analyses indicated fewer and different conjugation sites for iso-sinapyl alcohol than for sinapyl alcohol, but dehydrogenation polymerization assays indicated that *iso*-sinapyl alcohol does not affect the size of the lignin polymer. Although there was no indication that *iso*-sinapyl alcohol was integrated into the cell wall, diversion of substrates from sinapyl alcohol to free *iso*-sinapyl alcohol, its glucoside, and associated upstream lignin pathway changes, including increased phenolic aldehydes and acids, are associated with more facile cell wall deconstruction, and to the observed inhibitory effect on microbial growth. The release of wall constituents by mild pretreatment hydrolysis may be an informative approach that integrates wall-polymerized and associated metabolic responses that occurred over time rather than at a single point in time.

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.

* Presenting author

69

Flowthrough Pretreatment to Characterize Biomass Deconstruction

Heather L. McKenzie,[1,4] **Charles E. Wyman**[1,4]* (cewyman@engr.ucr.edu), Marcus B. Foston,[2,4] Jung Seokwon,[2,4] Arthur Ragauskas,[2,4] Nancy L. Engle,[3,4] Joshua F. Emory,[3,4] Bruce A. Tomkins,[3,4] Timothy Tschaplinski,[3,4] Gerald Tuskan,[4] Gary J. Van Berkel,[3,4] and **Paul Gilna**[4] (BESC PI)

[1]University of California, Riverside; [2]Georgia Institute of Technology, Atlanta; [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.; 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). 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 conversion of lignocellulosic biomass to ethanol requires recovery of sugars contained in hemicellulose and cellulose with high yields. High temperature water-only pretreatment can hydrolyze and solubilize hemicellulose and prepare cellulose left in the solids for enzymatic hydrolysis to glucose. As hemicellulose and lignin are thought to be key barriers to enzymatic hydrolysis, their alteration and/or removal are frequently cited as important pretreatment goals. However, there is limited knowledge about hemicellulose and lignin behavior during pretreatment. Lignin may be removed by depolymerization, melting, or solubilization. The hydrolysis of xylan, a key component in hemicellulose, has been modeled as a first order homogeneous reaction but the effectiveness of such models is limited. It has been difficult to follow lignin and xylan removal in batch pretreatment reactors because the final cooling may cause precipitation from the liquid phase. In comparison, biomass pretreatment by flowing water through a fixed biomass bed in a flowthrough mode removes these products from the reactor prior to final cooling, facilitating tracking its history. Both flowthrough and batch reactors are being applied to study the effects of pretreatment on birchwood xylan, lignin isolated from poplar, and poplar itself over a range of times and temperatures. Ultra high pressure liquid chromatography, gas chromatography mass spectrometry, and high pressure liquid chromatography are used to measure xylooligomers, phenols, and the total sugar content of hydrolyzate samples. The composition of lignin before and after pretreatment is characterized using gel permeation chromatography and heteronuclear single quantum coherence NMR. Based on the presence of phenol monomers in the liquid phase, loss of characteristic lignin bonds, side chains, and functional groups, and higher than expected number average and weight average molecular weight of polymers in the pretreated lignin solid, it seems likely that lignin is removed from and redeposited to the solid phase through a reactive mechanism. The substantial differences in lignin and xylan removal from native poplar and model substrates suggest that lignin-carbohydrate interactions enhance lignin removal while limiting the release of xylan.

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.

70

Plant Biomass-Degrading Loci Play a Role as Determinants for Lignocellulose Degradation From the Extremely Thermophilic Genus, *Caldicellulosiruptor*

Sara E. Blumer-Schuette[1,4]* (seblumer@ncsu.edu), Richard J. Giannone,[2,4] Jeffrey V. Zurawski,[1,4] Inci Ozdemir,[1,4] Dhaval B. Mistry,[1] Qin Ma,[3,4] Yanbin Yin,[3,4] Ying Xu,[3,4] Farris L. Poole II,[3,4] Irina Kataeva,[3,4] Mike W.W. Adams,[3,4] Scott D. Hamilton-Brehm,[2,4] James G. Elkins,[2,4] Frank W. Larimer,[2] Miriam L. Land,[2] Loren Hauser,[2] Robert W. Cottingham,[2,4] Robert L. Hettich,[2,4] **Robert M. Kelly**,[1,4] and **Paul Gilna**[4] (BESC PI)

[1]North Carolina State University, Raleigh; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]University of Georgia, Athens; 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). 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.

Consolidated bioprocessing (CBP) of second-generation biofuels from lignocellulose will require that the microbe(s) be able to deconstruct complex polysaccharides from plant biomass. The extremely thermophilic, non-cellulosomal genus *Caldicellulosiruptor* have come under renewed interest due to their plant biomass-degrading abilities at high temperatures (T_{opt}, 70~80°C). In order to fully identify the genetic diversity of the genus, eight *Caldicellulosiruptor* species, ranging from weakly to strongly cellulolytic were selected as a representative samples for whole genome sequencing (WGS). Comparative genomics using these eight genome sequences determined that the pan-genome for the genus *Caldicellulosiruptor* is not yet saturated. The *Caldicellulosiruptor* pan-genome (4,009 ORFs, based on eight species) encodes 92 glycoside hydrolases (GHs) representing 43 GH families, but only 25 GHs from 17 families are included in the core genome (1,543 ORFs). Three of these core CAZy-related genes are both multi-modular and extracellular and include a bi-functional cellobiohydrolase/endo-xylanase, an a-amylase, and an endo-xylanase. Non-core ancestral glucan degrading and xylan degrading loci were identified based on protein homology and chromosomal location. Comparative analysis revealed that the key determinant for cellulolytic activity was not the number of GHs but rather the presence of a genomic locus encoding one or more novel multi-domain cellulases, coupled to a specific type of carbohydrate binding modules (CBM3). Weakly cellulolytic species have completely or partially lost this locus, specifically genes encoding for GH48 catalytic domains coupled with a CBM3 module. Furthermore, the cellulolytic *Caldicellulosiruptor* species produced adhesins (substrate-binding portions of type IV pili), encoded upstream of CBM3 gene clusters in the same locus, that were found to bind to crystalline cellulose. Overall, the use of comparative genomics has identified novel multi-modular CAZy-related enzymes, highlighted the importance of previously known and characterized cellulase loci and identified mechanisms for substrate attachment. The genus *Caldicellulosiruptor*, by virtue of their plant biomass-degrading abilities make an attractive model system for CBP organisms, and with advances in genetic systems, can become engineered CBP biocatalysts as well.

This work was supported by a grant from the BioEnergy Science Center (BESC), a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.
The 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.

71

Identifying Potential Detoxification Mechanisms for Furan and Aromatic Aldehydes in Thermophilic, Anaerobic Bacteria Using 2D LC-MS/MS Proteomics

James G. Elkins* (elkinsjg@ornl.gov), Scott D. Hamilton-Brehm, Richard J. Giannone, Nancy L. Engle, Steven D. Brown, Timothy J. Tschaplinski, Robert L. Hettich, and **Paul Gilna** (BESC PI)

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.

Chemical and physical pretreatment of lignocellulosic biomass improves substrate reactivity but also releases microbial inhibitors such as furan aldehydes, low molecular weight fatty acids, and phenolic compounds. Candidate organisms for consolidated bioprocessing (CBP), including *Clostridium thermocellum* and *Caldicellulosiruptor sp.*, will likely require a combination of engineering and evolved adaptation for robust growth and fermentation in the presence of pretreatment inhibitors. Resistance mechanisms to primary inhibitors such as furfural have been described for *Saccharomyces cerevisiae* and ethanologenic *Escherichia coli* but similar responses in cellulolytic thermophiles are poorly characterized. Saccharolytic thermophiles from the genera *Thermoanaerobacter* and *Caloramator*, including a novel isolate from our laboratory, readily grow in the presence of biomass acid-hydrolyzates and demonstrate conversion of furan and aromatic aldehydes to less toxic alcohols. The enzymes responsible for detoxification are unknown but once identified, could potentially be expressed in other organisms relevant to CBP. To identify protein targets for further characterization, the proteomes of two furfural-resistant strains, *Thermoanaerobacter pseudethanolicus* 39E and *Caloramator proteoclasticus* ALDO1, were measured via multidimensional liquid-chromatography mass spectrometry after growing in the presence or absence of 15 mM furfural. While the two anaerobic thermophiles showed similarities in their respective physiological response to the inhibitor, distinct differences were also apparent. Both organisms displayed increases in ABC transporter-related proteins for sugar uptake by two orders of magnitude. *T. pseudethanolicus* 39E upregulated a number of enzymes involved in purine, pyrimidine, and amino acid biosynthesis while *C. proteoclasticus* ALDO1 showed increases in proteins encoded by a large operon responsible for lipid production. Several enzymes involved in redox reactions were also identified in both strains upon exposure to furfural and their role in aldehyde reduction will be further investigated.

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.

72

Methylation by a Unique α-class N4-Cytosine Methyltransfersase is Required for DNA Transformation of *Caldicellulosiruptor bescii* DSM6725: Use for Construction of Mutants That Affect Biomass Utilization

Daehwan Chung[1,2]* (chung31@uga.edu), Min Cha,[1] Joel Farkas,[1,2] Estefania Olivar,[1,2] **Janet Westpheling**,[1,2] and **Paul Gilna**[2] (BESC PI)

[1]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 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.

Thermophilic organisms offer special advantages for the conversion of lignocellulosic biomass to biofuels and bioproducts. The use of these complex substrates often requires pretreatment, involving exposure to acid or base at high temperature and the addition of hydrolytic enzymes that partially digest the plant cell walls. Enzymatic pretreatment is expensive and often prohibitive for the production of low value commodity products. Members of the Gram-positive bacterial genus *Caldicellulosiruptor* are anaerobic thermophiles with optimum growth temperatures between 65 °C to 78 °C and are the most thermophilic cellulolytic organisms known. *C. bescii* is capable of using both unpretreated switchgrass and *Populus* for growth. The ability to genetically manipulate these organisms is a prerequisite for engineering them for use in conversion of these complex substrates to fuels and products of interest. Here we report the first example of DNA transformation of a member of this genus, *C. bescii*. We show that restriction of DNA is a major barrier to transformation and that methylation of heterologous DNA with a unique α-class N4-Cytosine methyltransfersase is required for DNA transformation. We have used this genetic system to generate deletions of genes predicted to be involved in biomass utilization and identified a cluster of genes that encode pectinase enzymes that play an important role in biomass utilization.

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.

73

Improving Ethanol Production in *Clostridium thermocellum*

Daniel G. Olson[1,2]* (daniel.g.olson@dartmouth.edu), Adam Guss,[2] Yu Deng,[1,2] Jon Lo,[1,2] Aaron Argyros,[2,3] Chris Herring,[2,3] A. Joe Shaw,[2,3] Douwe van der Veen,[1,2] Ranjita Biswas,[2] **Lee R. Lynd**,[1,2,3] and **Paul Gilna**[2] (BESC PI)

[1]Dartmouth College, Hanover, N.H.; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.; 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 has been considered as a candidate for biofuel production due to its ability to rapidly solubilize crystalline cellulose and produce ethanol. Low ethanol yield and titer are currently the primary obstacles preventing *C. thermocellum* from further consideration as a consolidated bioprocessing organism. Recent development of a system for genetic modification of this organism has resulted in the creation of more than 30 mutants, some of which exhibit dramatic improvements in ethanol yield. Analysis of these strains using a metabolic flux analysis framework has allowed us to understand the role of electron carrier flexibility in determining ethanol yield.

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.

74

Thermophilic Isobutanol Production in *Geobacillus thermoglucosidasius*

Paul P. Lin,[1,2] Jennifer Takasumi,[1,2] Elizabeth Felnagle,[1,2] **James C. Liao**[1,2]* (liaoj@seas.ucla.edu), and **Paul Gilna**[1,2] (BESC PI)

[1]University of California, Los Angeles 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.

Consolidated bioprocessing using thermophilic microorganisms has a potential to provide the lowest cost for biological conversion from lignocellulosic biomass to biofuels. However, potential problems of thermophilic production of biofuels include the enhanced toxicity of the product and the volatility of intermediates at high temperatures. To test the feasibility of thermophilic production of higher-chain alcohols, we used *Geobacillus thermoglucosidasius* as a test organism for isobutanol and n-butanol production. We engineered key enzymes for thermostability, identified promoters, and conducted preliminary expression optimization. We successfully achieved isobutanol production in *G. thermoglucosidasius.*

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.

75

Engineering Improved Cellulases for CBP Microbes

Qi Xu,[1,3] Mo Chen,[2,3] Michael G. Resch,[1,3] Markus Alahuhta,[1,3] Roman Brunecky,[1,3] Vladimir V. Lunin,[1,3] John W. Brady,[2,3] Michael F. Crowley,[1,3] Yannick J. Bomble[1,3]* (yannick.bomble@nrel.gov), **Michael E. Himmel**,[1,3] and **Paul Gilna**[3] (BESC PI)

[1]National Renewable Energy Laboratory, Golden, Colo.; [2]Cornell University, Ithaca, N.Y.; 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.

Several recent studies have shown that certain cellulases in *Clostridium thermocellum* and *C. bescii* are some of the most efficient cellulases for biomass degradation. Being able to understand and improve these cellulases, and ultimately create a more efficient microbe to overcome biomass recalcitrance would be a huge step in improving the CBP process. Here we show how we use basic understanding, rational protein design and engineering to improve cellulases in *C. thermocellum* and *C. bescii*. We demonstrate how computer simulations and design can bring new understanding and help in the enzyme engineering process.

More specifically, new evidence in the function of X1 domains in CbhA from *C. thermocellum* has motivated the creation of a more efficient chimera using domain swapping. This improved chimera exhibits twice the activity of the wild type CbhA. Additionally, new understanding in the mechanisms and strengths of multi-catalytic domain cellulases, such as CelA (*C. bescii*) has led to more efficient minisomes able to convert 55% of cellulose at 60°C. Finally, we report that the energy required for product expulsion in the GH48 from CelA, one of the most active cellulases ever reported, is lower than in other GH48 cellulases. This result could partially explain the superior nature of this cellulase over other GH48-borne cellulases and inspire mutational strategies in other GH48 cellulases.

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

* Presenting author

Biofuels Research
Analytical Technologies, Engineering, and Production

76
Phylogenetic Distribution of Potential Cellulases in Bacteria

Renaud Berlemont* (rberlemo@uci.edu) and **Adam Martiny**

Earth System Science and Ecology and Evolutionary Biology, University of California, Irvine

http://allison.bio.uci.edu/DOE_Microbes.html

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.

In most terrestrial ecosystems, the depolymerization of plant cell wall is the rate limiting step in the turnover of organic material. The composition of plant detritus is known to depend mainly on enzymes produced by microorganisms. This raises the question: which phylogenetic lineages of microorganisms can degrade plant cell wall material, including cellulose?

To address this question, we compared the distribution of Glycoside Hydrolases (GH) potentially related to the cellulose degradation among 3744 bacteria.

Some phylogenetic groups are especially rich in GHs whereas some are very poor. For example, in bacteria from the Bacteroidetes phylum ~40 GHs (from the families 1, 3 5, 6, 8, 9, 10, 12, 16, 45, 48) are described (per genome) but in the Chlorobi phylum less than 5 of these protein-encoding genes are observed, per sequenced genome. The others bacteria that lack these GHs belong to the Aquificae, Chlamydiae, Chrysiogenetes, Cyanobacteria, Deferribacter, Elusomicrobia, Fusobacteria, Gemmatimonadetes, Synergistetes, Tenericutes and Nitrospirae phyla.

In addition, when present in genomes, not all the GH families are equally observed. Indeed, enzymes from the family 6, 8, 9, 45, and 48 are 1.5 to 3 times less abundant than enzymes from the family 1, 3 and 5. The over-representation of these enzymes is related to their numerous activities (e.g. GH1: β-glucosidase; β-galactosidase; β-mannosidase; β-glucuronidase; exo-β-1,4-glucanase; etc.) whereas the enzymes with low abundance display only few activities (e.g. GH6: endoglucanase and cellobiohydrolase).

For each phylum, we defined an abundance profile (for GH1, 3, 5, 6, 8, 9, 10, 12, 16, 45 and 48). These profiles were compared and clustered in order to discriminate the potential cellulose degraders from the other organisms. This clustering highlights the adaptation of microorganisms to different "life-styles". For example, autotrophic bacteria (e.g. Cyanobacteria) and/or intracellular pathogens (e.g. Tenericutes) generally lack many enzymes involved in polymers degradation, outside the cell (e.g. GH3, 5, 6, 8).

The association of 'Carbohydrate Binding Modules' (CBM) to catalytic domains (GH) is assumed to potentiate the catalytic activity by proximity effects. The association of GHs, from the above mentioned GH families, with CBMs from the family 2, 3 and 4 was thus investigated. Surprisingly, the CBM-GH profiles differ significantly from the profiles of their 'free enzymes' relatives. The association of CBM to GHs is a signature for enzymes involved in the catabolism of cellulose whereas the physiological function of 'free' enzymes is still pending (anabolism and/or catabolism).

Early Career Program

77
Engineering Robust Hosts for Microbial Biofuel Production

Mary Harrison and **Mary J. Dunlop*** (mjdunlop@uvm.edu)

School of Engineering, University of Vermont, Burlington

Project Goals: The goal of this project is to develop tools for improving microbial tolerance of biofuel production conditions. We aim to develop specific strategies for improving tolerance to enhance microbial synthesis of next-generation biofuels. The work is organized around three objectives: (1) Identify novel biofuel tolerance mechanisms from a targeted set of microorganisms. We are focusing on microorganisms that survive in hydrocarbon-rich environments as a source of tolerance genes, looking specifically for strategies that prevent biofuel from compromising membrane integrity and mechanisms for exporting biofuel-like compounds from the cell. (2) Engineer a synthetic feedback loop that responds to biofuel production. To optimize biofuel production yields, cells must balance several competing sources of stress. We are designing and constructing a novel feedback loop that senses biofuel production and turns on export pumps in response. (3) Finally, we are integrating multiple tolerance strategies in a biofuel production strain. In addition to

having the potential to greatly enhance biofuel yields, this work will advance understanding of how multiple stress tolerance mechanisms interact within a cell.

Microorganisms can be used to synthesize fuel from renewable materials. Microbial biofuel synthesis is a cost effective and environmentally sustainable way of producing replacements for gasoline, diesel, and jet fuel from lignocellulosic biomass. In a typical production process, biomass is deconstructed into sugars that are metabolized by a microbe engineered to convert sugar into biofuel. In this work, we focus on fuel synthesis—the final stage of biofuel production—and develop engineering tools for increasing the robustness of a biofuel production host.

A major challenge when using microorganisms to produce bulk chemicals like biofuels is that the production targets are often toxic to cells. Biofuel-like compounds are known to reduce cell viability through damage to the cell membrane and interference with essential physiological processes. Thus, cells must trade off biofuel production and survival, reducing potential yields. The majority of microbial biofuel research to-date has focused on engineering metabolic pathways for biofuel production. It is essential that we also engineer strains for biofuel tolerance. This is especially important for the large-scale production environments that will be required for generating cost effective biofuels.

Recent work on ethanol has already demonstrated that engineering strains for end-product tolerance can greatly improve overall yield. Here, we focus on engineering tolerance to advanced biofuels, taking a targeted approach toward improving biofuel tolerance and yield. Our goal is to develop a list of rational targets, rather than employing strategies like adaptation and random mutagenesis, to identify mechanisms that can be easily moved into production strains.

Recent work by the PI has indicated that microorganisms that survive in oil-rich environments are a valuable source of tolerance mechanisms [1-3]. We are testing genes from microbes that have been isolated from environments near natural oil seeps and in the vicinity of oil spills to see if they improve biofuel tolerance when expressed in *Escherichia coli*. Whole genome sequencing efforts have included several of these hydrocarbon-tolerant microbes. We are using this sequence data as a starting point for identifying tolerance genes to heterologously express in a biofuel production host.

References

1. Dunlop, M.J., et al., *Regulatory activity revealed by dynamic correlations in gene expression noise.* Nature Genetics, 2008. 40(12): p. 1493-1498.
2. Dunlop, M.J., et al., *Engineering microbial biofuel tolerance and export using efflux pumps.* Mol Syst Biol, 2011.
3. Dunlop, M.J., J.D. Keasling, and A. Mukhopadhyay, *A model for improving microbial biofuel production using a synthetic feedback loop.* Systems and Synthetic Biology, 2010. 4(2): p. 95-104.

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

78

Engineering Ethylmalonyl-CoA Pathway in *Methylobacterium extorquens* AM1 for Butanol Production: Identification of a Regulator Activating the Expression of a Key Gene

Bo Hu[1]* (bohu@uw.edu), **David Beck**,[1] **Robert Synovec**,[2] and **Mary Lidstrom**[1,3]

[1]Department of Chemical Engineering, [2]Department of Chemistry, and [3]Department of Microbiology, University of Washington, Seattle

https://depts.washington.edu/mllab/

Project Goals: The main goal of this project is to characterize control points for assimilatory metabolism in a facultative methylotroph, *Methylobacterium extorquens* AM1. Methylotrophs have potential for conversion of reduced one-carbon compounds to value-added chemicals, but a more detailed understanding of assimilatory metabolism is necessary for that potential to be reached. This project will generate systems-level datasets to inform optimization approaches for a new platform for biofuels synthesis, tools for systems approaches via multi-level datasets, and starting strains for biofuels strain development using methanol + CO_2 as biofeedstocks.

Our initial work focused on potential regulatory genes for the ethylmalonyl-CoA (EMC) pathway, the part of assimilatory metabolism that involves crotonyl-CoA as an intermediate. We have identified CcrR, a TetR-type activator, which has been shown to regulate the expression of crotonyl-CoA reductase/carboxylase, a key enzyme of EMC pathway. The *ccrR* mutant is impaired in its ability to grow on C1 and C2 compounds, correlating with the reduced activity of crotonyl-CoA reductase/carboxylase. The *ccr* gene was found to be cotranscribed with an upstream gene (*katA)* and the promoter strength was found to reduce as much as 50% in the absence of CcrR compared to that in wild type *M. extorquens* AM1. Gel retardation assays with purified His-tagged CcrR showed that Ccr binds to the promoter-regulatory region of *ccr*. A palindromic sequence upstream of *katA* at position -334 to -321 with respect to the translational start site was found to be the binding site and mutations in this region eliminated gel retardation with CcrR. These results show that CcrR stimulates expression of the *katA-ccr* promoter on the order of two-fold but is not required for this expression. The identification of a specific activator protein regulating expression of one of the genes of the EMC pathway is a first step in understanding how the pathway as a whole is regulated, and this work also generates information for manipulating flux to end products that use crotonyl-CoA as precursor, such as butanol.

As a proof of principle example for manipulation of carbon flux through crotonyl-CoA, we have initiated studies to introduce a two-step pathway into *M. extorquens* AM1 to convert crotonyl-CoA into butanol. The initial step would be reduction of crotonyl-CoA to butyryl-CoA by

* Presenting author

a crotonyl-CoA specific trans-enoyl-CoA reductase from *Treponema denticola* (Ter), coupled with another reduction step catalyzed by alcohol dehydrogenase from *Clostridium acetobutylicum* ATCC8244 (AdhE2) to produce butanol. Genes encoding Ter and AdhE2 were synthesized with codon usage optimized using a codon usage table derived from highly expressed genes of *M. extorquens* AM1 grown on methanol. Each gene was subcloned into *M. extorquens* AM1-adaptable expression vectors for expression. Enzyme assays showed that both Ter and AdhE2 activities were detected in recombinant *M. extorquens* AM1 strains containing plasmids with *ter* or *adhE2*. The immediate future plan is to integrate both genes into a single operon with various expression capabilities to develop a strain with high butanol productivity.

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

79

Integration of Carbon, Nitrogen, and Oxygen Metabolism in *Escherichia coli*

Yi-Fan Xu[1,2]* (yifanxu@Princeton.EDU), Christopher D. Doucette,[1,3] Junyoung O. Park,[1,4] Nathaniel Roquet,[1] Daniel Amador-Noguez,[1] Marshall L. Reaves,[1,3] Xiao-jiang Feng,[2] Herschel Rabitz,[2] David J. Schwab,[1,3] Ned S. Wingreen,[1,3] and **Joshua D. Rabinowitz**[1,2]

[1]Lewis-Sigler Institute for Integrative Genomics, [2]Department of Chemistry, [3]Department of Molecular Biology, and [4]Department of Chemical and Biological Engineering, Princeton University, Princeton, N.J.

Project Goals: A key challenge for living systems is balancing utilization of different elemental nutrients, such as carbon, nitrogen, and oxygen. Although the regulation of specific nutrient assimilation systems has been extensively studied, how these systems coordinate with each other remains poorly understood. Here we aim to obtain a quantitative understanding of such coordination in *Escherichia coli* with a focus on fast-acting regulatory mechanisms. To this end we are (1) quantifying metabolic responses to nutrient perturbations using metabolomics, (2) building differential equation models that bridge multiple nutrient systems, and (3) discovering regulatory principles from this unified combination of metabolomics and modeling.

Results:

Biomass production requires the integration of multiple external nutrients, whose availability is subject to environmental fluctuations. As growth can be limited by the scarcity of any one nutrient, the rate at which each nutrient is assimilated must be sensitive not only to its own availability, but also to that of other nutrients. Remarkably, across diverse nutrient conditions, *E. coli* grow nearly optimally: they balance effectively the conversion of carbon into energy versus biomass, and excrete only small amounts of waste. What type of regulatory architecture might (i) enable homeostasis of intracellular metabolite concentrations and (ii) produce metabolic fluxes that nearly optimize growth? The most basic homeostatic regulatory mechanism is feedback inhibition. Through a series of reductionist differential equation models, with fluxes expressed in Michaelis-Menten form, we found that *feedback inhibition alone is sufficient to meet both of the above goals*. While such a result was expected for a linear biosynthetic pathway, it was striking that feedback inhibition can in theory also effectively regulate metabolic cycles, the portioning of carbon between energy production and biomass, and the integration of multiple nutrients (e.g., nitrogen and carbon).

We then considered how actual feedback regulatory architectures compare with those predicted to be effective based on our conceptual analysis. *E. coli's* ammonia assimilation cycle, involving the enzymes glutamine synthetase (GS) and glutamate synthase (GOGAT), is among the best-studied metabolic cycles from the perspective of regulation. Despite this, existing literature focuses on a single feedback (from glutamine on GS), whereas our analysis indicated the need also for a feedback from glutamate on GOGAT. Through detailed metabolomic analysis and modeling of the cycle, including analysis of various metabolic and regulatory mutants, we showed that a feedback on GOGAT is indeed required, and that surprisingly it is provided by aspartate, whose concentration mirrors that of glutamate.

To investigate the link between carbon and nitrogen metabolism, we measured changes in the glucose uptake rate in response to changing nitrogen availability. We found that glucose uptake in *E. coli* is subject to rapid regulation by nitrogen. Remarkably, nitrogen up-shift doubles glycolytic flux in 2 min without substantial changes in the concentration of any glycolytic intermediate. What type of feedback might enable such a response? The most elegant possibility involved the carbonaceous substrate of nitrogen assimilation α-*ketoglutarate directly inhibiting the phosphotransferase system,* which simultaneously uptakes glucose and drains the bottom of the glycolysis pathway. Such a feedback loop, never previously proposed, was verified biochemically and shown genetically to be both *necessary and sufficient for nitrogen-based regulation of sugar uptake*.

Much of central carbon metabolic flux is altered by changing oxygen levels. Nevertheless, there are parallels to the case of nitrogen metabolism: α-ketoglutarate, the product of carbon metabolism that directly feeds nitrogen metabolism, controls carbon metabolic response to nitrogen availability; similarly, we believe that NADH, the product of carbon metabolism that immediately feeds into oxygen metabolism, controls carbon metabolic response to oxygen availability. NADH levels impact carbon metabolism in multiple ways, including shutting off the right side of the tricarboxylic acid (TCA) cycle, reversing the left side of the TCA cycle, and inhibiting glycolysis. This is basically feedback inhibition writ large.

Finally, we considered when feedback inhibition might be insufficient as a metabolic regulatory strategy. Our analyses suggested its sufficiency for all steady-state conditions.

---

* Presenting author

Therefore, we considered the possible importance of other regulatory motifs in oscillating conditions. A proposed feed-forward circuit involves activation of phosphoenolpyruvate carboxylase (the anapleurotic enzyme in *E. coli* that provides new 4 carbon units into the TCA cycle) by fructose-1,6-bisphosphate. We verified genetically the existence of this regulation, found that the regulation is ultrasensitive, and showed that ultrasensitive activation of this enzyme by fructose-1,6-bisphosphate markedly enhances *E. coli* growth specifically during oscillating glucose conditions.

Taken together, these findings support a general mechanism for nutrient integration: *Limitation for a nutrient other than carbon leads to build-up of the most closely related product of carbon metabolism, which in turn feedback inhibits further carbon metabolism.* Our efforts to translate this principle into a predictive, quantitative model of central carbon, nitrogen, and oxygen metabolism are ongoing. Eventual success will require intensive re-examination of the regulation of many enzymes, with the ultimate payoff being enhanced understanding and rational engineering of bacterial metabolism.

80

Metabolomics of Clostridial Biofuel Production

Daniel Amador-Noguez,[1] Ludmilla Aristilde,[1] Caleb Bastian,[2] Herschel Rabitz,[2] Leonid Kruglyak,[1] and **Joshua D. Rabinowitz**,[1,2]* (joshr@genomics.princeton.edu)

[1]Lewis-Sigler Institute for Integrative Genomics and [2]Department of Chemistry, Princeton University, Princeton, N.J.

Project Goals: The production of biofuels from cellulosic biomass holds promise as a source of renewable clean energy. Members of the genus *Clostridium* collectively have the ideal set of the metabolic capabilities for biofuel production from cellulosic biomass: *C. acetobutylicum* rapidly ferments glucose to biofuels (butanol, hydrogen) and *C. cellulolyticum* effectively degrades cellulose. Here we aim to integrate metabolomics, genomics and genetic engineering to dramatically advance understanding of metabolism in *C. acetobutylicum* and *C. cellulolyticum*. In so doing, we will lay basic science groundwork for engineering of an organism that cost-effectively converts cellulose into solvents and/or hydrogen gas.

Results and Plans:
The fermentation carried out by *C. acetobutylicum* is characterized by production of acids during exponential growth ("acidogenesis"), followed by production of solvents as growth slows down ("solventogenesis"). We previously used metabolomics and quantitative flux modeling to map the metabolic changes associated with this acidogenic-solventogenic transition. This effort revealed a marked reorganization of central metabolism that involved both large changes in intracellular metabolite concentrations and metabolic fluxes. Solventogenesis was characterized by a dramatic down regulation of fluxes through pyruvate carboxylase, the reductive TCA cycle, and amino acid biosynthesis. These flux changes favor build-up of pyruvate and reducing power, the substrates of solventogenesis. Recently, we have knocked out pyruvate carboxylase, and find that this facilitates transition into solventogenesis, improves acid re-assimilation, and slightly increases final solvent yields.

Another key aspect in biofuel production from cellulose is hemicellulosic sugar catabolism. We previously mapped the pathways of glucose metabolism in *C. acetobutylicum* by following the dynamics of ^{13}C-labeled glucose assimilation. Here we are conducting similar isotope tracer experiments to explore the simultaneous utilization of glucose and hemicelluloses sugars (xylose, arabinose, mannose, and galactose). Our initial experiments show that while galactose is not assimilated in the presence of glucose, the organism simultaneously and non-discriminantly catabolizes both glucose and mannose. Similarly, both xylose and arabinose are assimilated in the presence of glucose not only into the pentose phosphate pathway, but also lower glycolysis. There is no assimilation into upper glycolysis, however. This suggests that, in the presence of pentose sugars, there is minimal activity of the transaldolase reaction, which yields fructose-6-phosphate. Follow-up experiments with different types of ^{13}C-labeled glucose and xylose are underway to confirm this hypothesis. In addition, we are working to model quantitatively the associated fluxes.

Finally we aim to understand *C. cellulolyticum's* catabolic limitations. While an impressive cellulose degrader, *C. cellulolyticum* is a sluggish fermenter (20-fold slower than *C. acetobutylicum*). We will apply metabolomics and genomics to investigate the reasons for its slow sugar catabolism.

Funding: DOE DE-SC0006839.

81

Experimental Systems Biology Approaches for Clostridia-Based Bioenergy Production: The Metabolite Stress-Response System in Solventogenic Clostridia

E. Terry Papoutsakis[1,2]* (epaps@udel.edu), Blake C. Meyers,[2,3] Kelvin H. Lee,[1,2] Maciek R. Antoniewicz,[1] Costas D. Maranas,[4] H. Huang,[2] and **Cathy H. Wu**[2,5]

[1]Department of Chemical Engineering, University of Delaware, Newark; [2]Delaware Biotechnology Institute, Newark; [3]Department of Plant and Soil Sciences, University of Delaware, Newark; [4]Department of Chemical Engineering, Pennsylvania State University, University Park; and [5]Department of Biochemistry and Molecular and Cellular Biology, Georgetown University Medical Center, Washington, D.C.

http://research.bioinformatics.udel.edu/clostridia-stress-response

Project Goals: The objectives of this project are to engage enabling experimental systems-biology approaches to support the development of integrated, predictive models of the metabolic and regulatory networks underlying the metabolite stress response in solventogenic clostridia. Clostridia are Gram$^+$, obligate anaerobic, endospore-forming bacteria of major importance to fermentative biofuel production. Here, we focus on understanding and modeling the stress-response of *Clostridium acetobutylicum* to two important toxic metabolites: butanol and butyrate. Systems-level understanding is expected to lead to better strategies for industrial-strain development, as well as bioprocessing strategies taking advantage of the stress response to achieve superior bioprocessing outcomes.

Solventogenic and other clostridia are of major importance for developing technologies for biofuel production (3). A major and unique advantage is their ability to utilize a large variety of substrats (hexoses, pentoses, oligosaccharides, xylan, and starches). Among the two sequenced solventogenic clostridia, *C. acetobutylicum* is the only one that contains a full cellulosome (2) and may thus directly utilize cellulosic material for production of fuels and chemicals.

The toxic-metabolite stress response is a problem of major and general importance not only in clostridial biotechnologies but in all microbial systems of interest to bioenergy production (1). Achieving the goals of this project will solve a long-standing problem of both fundamental and practical importance. The significant broader impact is that the approaches and tools developed here can be applied to other cellular systems, aiming to achieve the ultimate goal of comprehensive multidimensional understanding of cells via Genome Scale Model (GSM) multidimensional platforms and models. The outcomes of this project aim to become an enabling paradigm for modeling complex programs of organisms and biological systems of importance to DOE's mission on energy and the environment.

To achieve these goals, we apply genomic tools for collecting extensive transcriptomic (based on both deep sequencing and microarray analyses) and targeted fluxomic and proteomic data, and couple omics data integration with building stress models and modeling platforms that can be linked, as an added modeling dimension, to a 2nd generation GSM of this organism. Successful completion of this project aims to the following deliverables: (i) experimental omics data of the metabolite stress response in *C. acetobutylicum* for deposition to public repositories and that can be mined for systems-level predictive modelling; (ii) an in-depth systems-level molecular understanding at multiple genome-scale levels of the metabolite stress response; (iii) two models and modeling frameworks for the metabolite stress responses linking them to the constructed 2nd generation GSM; and, (iv) computational and bioinformatic tools and frameworks that can be applied to modeling other cellular programs.

References

1. Nicolaou, S. A., S. M. Gaida, and E. T. Papoutsakis. 2010. A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation. Metab Eng 12:307-331.
2. Nölling, J., G. Breton, M. V. Omelchenko, K. S. Makarova, Q. Zeng, R. Gibson, H. M. Lee, J. Dubois, D. Qiu, J. Hitti, Y. I. Wolf, R. L. Tatusov, F. Sabathe, L. Doucette-Stamm, P. Soucaille, M. J. Daly, G. N. Bennett, E. V. Koonin, and D. R. Smith. 2001. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. J Bacteriol 183:4823-4838.
3. Tracy, B. P., S. W. Jones, A. G. Fast, D. C. Indurthi, and E. T. Papoutsakis. 2011. Clostridia: the importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. Curr Opin Biotechnol doi: 10.1016/j.copbio.2011.10.008.

82

Population Level Analysis of Mutations Underlying Improvements in Biofuel Production by *Clostridium phytofermentans*

Supratim Mukherjee, Stephen Godin, and **Jeffrey L. Blanchard*** (blanchard@microbio.umass.edu)

Department of Microbiology, University of Massachusetts, Amherst

http://www.bio.umass.edu/micro/blanchard/

Project Goals: Although cellulosic plant biomass is plentiful and cheap, the cost of degrading the cellulosic and hemicellulosic components of the plant cell wall currently limits the use of plant biomass as a competitive renewable alternative to gasoline and is a key challenge in developing a U.S. biomass-based industry for manufacturing biofuels from agricultural and forestry wastes. The goal of this project is to determine the mechanistic basis of improvements in *C. phytofermentans* strains evolved to produce biofuels more rapidly. Understanding the mechanistic details of individual mutations will be of particular interest to the larger scientific community for improving rates of biomass degradation and biofuel production.

Clostridium phytofermentans is a genetically tractable, anaerobic bacterium isolated from forest soil near the Quabbin Reservoir and the Harvard Forest Long Term Ecological Research site in Massachusetts, U.S.A. *C. phytofermentans* (1) can saccharify all major carbohydrate components of plant biomass; (2) has acquired many biofuel-related enzymes through horizontal gene transfer; and (3) produces ethanol as the primary product of plant carbohydrate fermentation. The combination of its broad nutritional versatility coupled with its high levels of ethanol production distinguishes *C. phytofermentans* from all other cultured microbes characterized to date. Using long-term experimental evolution we have selected for strains of *C. phytofermentans* that grow and produce ethanol faster on plant carbohydrates. Full genome re-sequencing through the DOE Community Sequencing Program has allowed us to identify mutations present at varying allele frequencies in cultures with increased rates of biomass degradation. The results reveal new strategies for evolving and engineering microorganisms for the production of biofuels from plant feedstocks.

This research was funded by an Isenberg Scholar Award to S. Mukherjee, a Howard Hughes Medical Institute Fellowship to S. Godin and a DOE Community Sequencing Program award to J. Blanchard.

83

Understanding Fundamental Aspects of Butanol Production by *Clostridium beijerinckii*

Caroline B. Milne,[1,2] Yi Wang,[1] Beth Papanek,[1] Yong-Su Jin,[1] Nathan D. Price,[1,2] and **Hans P. Blaschek**[1]* (blaschek@illinois.edu)

[1]University of Illinois, Urbana and [2]Institute for Systems Biology, Seattle, Wash.

http://bioenergy.illinois.edu

Project Goals: The solventogenic clostridia offer a sustainable approach to petroleum-based production of n-butanol, an important chemical feedstock and potential fuel. With the availability of the genome sequence for *Clostridium beijerinckii* 8052, we can now employ the tools of systems biology in order to gain increased insight into the metabolic and regulatory networks relevant to solvent production. Project goals include examination of: 1) the mutations underlying the *C. beijerinckii* BA101 butanol-overproduction phenotype, 2) the molecular basis for the global shift from acidogenesis to solventogenesis, 3) the genetic basis of butanol tolerance in *C. beijerinckii* and 4) RNA-seq technology for single-nucleotide resolution analysis of the transcriptome of this microorganism.

We have constructed the first genome-scale metabolic model (*i*CM925) for *C. beijerinckii*, containing 925 genes, 938 reactions, and 881 metabolites. To build the model we employed a semi-automated procedure that integrated genome annotation information from KEGG, BioCyc, and The SEED, and utilized computational algorithms with manual curation to improve model completeness. To validate *i*CM925, we conducted fermentation experiments using the NCIMB 8052 strain, and evaluated the ability of the model to simulate measured substrate uptake and product production rates. Experimentally observed fermentation profiles were found to lie within the solution space of the model; however, under an optimal growth objective, additional constraints were needed to reproduce the observed profiles—suggesting the existence of selective pressures other than optimal growth. Notably, a significantly enriched fraction of actively utilized reactions in simulations—constrained to reflect experimental rates—originated from the set of reactions that overlapped between all three annotation databases used ($P = 3.52 \times 10^{-9}$, Fisher's exact test). Inhibition of the hydrogenase reaction was found to have a strong effect on butanol formation—as experimentally observed.

We also conducted a single-nucleotide resolution analysis of the *C. beijerinckii* NCIMB 8052 transcriptome using high-throughput RNA-Seq technology. We identified the transcription start sites and operon structure throughout the genome. We confirmed the structure of important gene operons involved in metabolic pathways for acid and solvent production in *C. beijerinckii* 8052, including *pta-ack*, *ptb-buk*, *hbd-etfA-etfB-crt* (*bcs*) and *ald-ctfA-ctfB-adc* (*sol*) operons; we also defined important operons related to chemotaxis/motility, transcriptional regulation, stress response and fatty acids biosynthesis along with others. We discovered 20 previously non-annotated regions with significant transcriptional activities and 15 genes whose translation start codons were likely mis-annotated. As a consequence, the accuracy of existing genome annotation was significantly enhanced. Furthermore, we identified 78 putative silent genes and 177 putative housekeeping genes based on normalized transcription measurement with the sequence data. We also observed that more than 30% of pseudogenes had significant transcriptional activities during the fermentation process. Strong correlations exist between the expression values derived from RNA-Seq analysis and microarray data or qRT-PCR results.

References

1. Milne, C.B, et al. 2011. Metabolic network reconstruction and genome-scale model of butanol-producing strain *Clostridium beijerinckii* NCIMB 8052. BMC Systems Biology. 5:130.
2. Wang, Yi., et al. 2011. Single-nucleotide resolution analysis of the transcriptome structure of *Clostridium beijerinckii* NCIMB 8052 using RNA-Seq. BMC Genomics, 12:479.

This work was funded by the DOE Office of Science through FOA: DE_FOA-0000368 and an NSF CAREER grant.

Early Career Award

84

Applying the Biology of Brown Rot Fungi to Consolidated Bioprocessing—Early Career Program

Robert A. Blanchette,[1] Shona M. Duncan,[1] Timothy Filley,[2] Joel Jurgens,[1] Justin Kaffenberger[1]* (kaffe002@umn.edu), and **Jonathan S. Schilling**[1]

[1]University of Minnesota, Saint Paul and [2]Purdue University, West Lafayette, Ind.

http://schillinglab.cfans.umn.edu/projects-pubs

Project Goals: *Postia placenta* is a wood-degrading brown rot fungus with a sequenced and annotated genome[1]. Like other brown rot fungi, *P. placenta* is capable of an oxidative hydroxyl radical pretreatment that occurs concurrently with enzymatic saccharification of woody carbohydrates. This consolidation of otherwise incompatible reactions is fundamentally interesting and has great implication on the potential to consolidate harsh pretreat-

ments with saccharification in a single processing step. Therefore, our research goals are as follows:

1. **physically sample wood degraded by the brown rot fungus *P. placenta* in order to map coincident pretreatment and saccharification reactions and to correlate relevant lignocellulose chemistry,**
2. **image pH and porosity at the fungus-plant interface and layer this data with images showing cellulase ingress, and**
3. **map, along the active hyphal front, the co-occurring expression of iron reductases associated with pretreatment and of cellulases used in saccharification.**

Enzymatic bioconversion of lignocellulose plant tissues generally requires an initial pretreatment step, followed by saccharification and then fermentation or other downstream processing approaches. Consolidated bioprocessing (CBP) of lignocellulose combines enzymatic sugar release (saccharification) with fermentation, but pretreatments typically remain separate and costly. In nature, lignocellulose-degrading brown rot fungi consolidate pretreatment and saccharification, likely using spatial gradients to partition these incompatible reactions. Our goal is to characterize how this is achieved, in order to better understand the fungus and to potentially apply this approach in a mimicked consolidated approach.

The goal of this research is characterizing this relevant biological system, with objectives (stated above) to 1) physically sample wood degraded by the brown rot fungus *Postia placenta* to map reactions spatially and to correlate with cell wall modifications, 2) produce images of the environmental variables (pH and porosity) affecting cellulase ingress over time during brown rot, and 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 (Obj. 1), or appropriately resolved microscopy (Obj. 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 C13-labeled tetramethyl ammonium hydroxide thermochemolysis for specific brown rot lignin modifications. For microscopy, I am utilizing 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 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 that we currently approach as separate, distinct steps.

Reference

1. Martinez et al. (40+ authors) (2009) Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *PNAS* 106:1954-1959.

85

Development of Quantum Dot Probes for Studies of Synergy Between Components of the Wood-Degrading Fungal Enzymes

Haw Yang,[1] Kevin Welsher,[1] Chih-Hao Hsia,[1] Daniel Montiel,[1] Robin Rohwer[2]* (rrr24@psu.edu), **B. Tracy Nixon**[2]* (btn1@psu.edu), and **Ming Tien**[2]

[1]Princeton University, Princeton and [2]The Pennsylvania State University, University Park

Project Goals: The overarching goal of this research program is to understand the cellular and molecular mechanisms that underlie the fungus-enzyme-lignocellulose gestalt that will enable the efficient and economical production of liquid fuels from plant biomass. The current project aims to develop quantum dot-based tagging and imaging technologies to simultaneously monitor several lignocellulosic enzymes in real-time and in the natural fungal/lignocellulose environment. The specific goals are: 1) Synthesis of commercially unavailable multiplexed luminescent quantum dots optimized for the study of lignocellulosic degradation. 2) Development of bio-orthogonal tags and protocols that can be generalized to other wood-degrading organisms for labeling secreted enzymes under realistic enzyme-biomass conditions. 3) Advancement of single-molecule imaging technology and the development of assays for studying individual lignocellulosic enzymes.

To achieve these objectives a team with highly complementary expertise was assembled: Tien of the Pennsylvania State University (co-PI) is the co-discoverer of the lignin-degrading enzyme lignin peroxidase, and is a world expert in fungal lignin biodegradation. He has extensive experience and knowledge of the enzymology, genetics, and biogenesis of lignocellulosic enzymes, as well as the culturing and genetic manipulation of white-rot fungi. Nixon of The Pennsylvania State University (co-PI) is the co-discoverer of the ubiquitous two-component signaling system, a central signal transduction pathway in bacterial, fungal and plant cells involved in response to environmental stimuli. He has experience using genetic, molecular, biochemical and structural approaches to link two-component signal transduction to the gene regulation machinery. Yang of Princeton University (PI) is an expert in the development of single-molecule assays and optical probes, and in using them to solve biological and biochemical problems. One of the instruments developed in his group allows real-time tracking of individual nanoscale optical probes in three dimensions with microsecond (μs) time resolution and nanometer localization resolution and will be used as the primary tool for this work. This team is uniquely able to discover how fungal enzymes synergistically degrade lignocelluloses, and the tools developed to accomplish this will be broadly applicable to study other systems in their natural microenvironments.

Goal 1 is to synthesize multiplexed luminescent quantum dots specifically optimized for the study of lignocellulosic degradation which are otherwise unavailable. The design stage for synthesis of compact and water-soluble non-blinking quantum dots is finished. Doped quantum dots that have multiple emitters in a single dot were developed as an improvement to conventional quantum dots which need a thick shell to suppress blinking behavior. The blinking behavior is overcome in doped quantum dots because all of the emitters (dopants) are unlikely to stay in the dark state at the same time when the quantum dots have a high doping concentration. With the bio-orthogonal tagging technologies being developed, the compact and water-soluble non-blinking quantum dots can be used as markers to follow the translational motions of enzymes.

Goal 2 is to develop bio-orthogonal tags and protocols that can be generalized to other wood-degrading organisms for simultaneously labeling multiple secreted enzymes under realistic enzyme-biomass conditions. The designs for engineering *Trichoderma reesei* to express bio-orthogonally tagged variants of Cel7A, Cel6A and Cel7B are finished. Our strategy improves upon the work reported in the existing literature by providing forward and reverse selections for introducing any recombinant form of the cellulase genes of interest in a way that leaves absolutely no footprint other than the desired change. In accomplishing changes this way, our manipulations will be cleanly restricted to the desired changes in a targeted cellulase. This maximizes our ability in the future to examine the behavior of secreted enzymes in a wood matrix that is being normally degraded by the fungus. The strategy also streamlines the process of testing many recombinant genes by providing for quick construction of desired genes and introducing them back into *T. reesei* via a single homologous recombination event. Strain construction is underway, with preliminary results anticipated by the time of the meeting.

Goal 3 is to advance single-molecule imaging technology and develop assays for studying individual lignocellulosic enzymes. To conquer the multiple scales covered by the action of lignocellulosic enzymes, the previously reported real-time single-particle spectroscopy (RT-3D-SPTS) method has been combined with a scanning two-photon microscope to create a high-resolution, multiscale spectral imaging system. The RT-3D-SPTS is implemented in the near infrared (NIR, 650 nm - 800 nm) region of the electromagnetic spectrum with the intent to avoid intrinsic autofluorescence of biological samples, which is typically confined to shorter wavelengths. The RT-3D-SPTS system has spatial localization resolution down to 10 nm in the lateral dimensions and 20 nm in the axial dimensions, with temporal resolution of down to 10 microseconds. The system is also fitted with a z-locking mechanism that counteracts objective drift to ensure fidelity of position measurement during tracking. The scanning-two photon microscope covers the larger length scales (up to 200 microns) allowing concurrent sampling of the substrate environment while the quantum dot tagged enzyme is investigated spectroscopically by the RT-3D-SPTS system. Crosstalk between the two modules is minimized by spectral separation, reaching further into the NIR (<800 nm) for two-photon excitation and collecting images in the visible (400 nm – 600 nm). This allows for imaging of conventional dyes, fluorescent proteins or autofluorescence. The system is further outfitted for biological samples with an onboard temperature control chamber. This multiscale imaging system has already been used to track freely diffusing nanoparticles in the presence of fluorescently labeled single cells.

This material is based upon work supported by the Department of Energy under Award Number DE-SC0006838.

86

Biodegradative Oxidant Production by Fungi in Lignocellulose: In Situ Mapping and Gene Expression

Premsagar Korripally,[1,2] Peter Kitin,[1,2] Christopher G. Hunt,[2] Carl J. Houtman,[2] **John Ralph**,[1] and **Kenneth E. Hammel**[1,2]* (kehammel@wisc.edu)

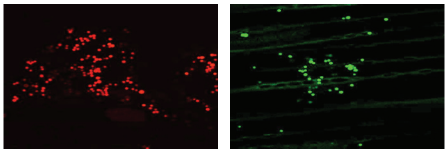
[1]University of Wisconsin, Madison and [2]U.S. Forest Products Laboratory, Madison, Wis.

Project Goals: Naturally occurring, biological mechanisms for lignocellulose deconstruction are of interest because they may inspire the development of environmentally friendly technologies for conversion of lignocellulose into useful fuels and chemicals. The only organisms that degrade lignocellulose efficiently are certain filamentous lignin-degrading fungi, especially white rot basidiomycetes such as *Phanerochaete chrysosporium*, whose genome was recently sequenced. It is known that lignin degradation is oxidative, but the specific mechanisms remain obscure–numerous enzymes and cofactors have been proposed to have a ligninolytic role. To address this problem, it would be useful to visualize fungal ligninolytic oxidants on biodegrading lignocellulose, to map their position relative to the fungus, and to correlate their production with the expression of fungal genes.

We have developed micrometer-scale silica beads that carry an oxidant-sensitive BODIPY dye, which fluoresces red in its native reduced state but green when oxidized. The beads were placed on thin sections of wood undergoing early decay by *P. chrysosporium* and the specimens were analyzed periodically by fluorescence microscopy. Concurrently, specimens were harvested for RNA extraction and transcript analysis by quantitative RT-PCR, using primers designed to amplify sequences encoded in the genome that have a proposed ligninolytic role. Three major findings have emerged from our

* Presenting author

study thus far: **(1)** In the first few days of biodegradation, there were marked gradients of oxidation around the fungal hyphae, i.e., beads close to the fungus were more oxidized than were more distant beads. This result is inconsistent with a ligninolytic mechanism that depends on small, freely diffusing oxidants. Instead, it indicates that enzymes associated with the fungal hyphae have a major oxidative role during the initial stage of biodegradation. **(2)** Bead oxidation was not continuous, but rather commenced abruptly several days after fungal inoculation onto the wood. This result indicates a rapid transition from nonligninolytic to ligninolytic metabolism by the fungus. **(3)** Concurrent with this oxidative burst, several fungal genes were highly up-regulated. Chief among these were several peroxidases that have been shown to cleave lignin in vitro. By contrast, a cellobiose dehydrogenase and two glycopeptides that have been proposed to produce ligninolytic hydroxyl radicals were not up-regulated. These results support a mechanism for incipient ligninolysis in which fungal peroxidases react directly with lignin to cleave it.



Fluorescence images of beads on spruce wood undergoing decay by *P. chrysosporium* at 40 h (left) and 96 h (right) after inoculation.

This work was supported by grant DE-FG02-11ER65271 from the Office of Science (BER), U.S. Department of Energy.

87

Multi-Species, Multi-Gene Co-Regulation: Finding DNA *Cis*-Regulatory Elements for Biofuel Pathways

Lee Ann McCue,[1] **Lee A. Newberg**,[2,3]* (lee.newberg@wadsworth.org), William A. Thompson,[4] and **Charles E. Lawrence**[4]

[1]Pacific Northwest National Laboratory, Richland, Wash.; [2]Wadsworth Center, New York State Department of Health, Albany, N.Y.; [3]Rensselaer Polytechnic Institute, Troy, N.Y.; and [4]Brown University, Providence, R.I.

Project Goals: Decreasing America's dependence on foreign energy sources and reducing the emission of greenhouse gases are important national priorities. We are undertaking research into the metabolic and regulatory networks responsible for biohydrogen and bioethanol production. In particular, among the hundreds of alpha-proteobacterial species with sequenced genomes are several species with metabolic capabilities of interest. The first long term goal of this research is to identify the ensemble of solutions that have been explored by the alpha-proteobacteria to regulate the metabolic processes key to biofuel production. The second long term goal of this project is to develop probabilistic models to represent these multiscale processes, through Bayesian statistical inference procedures and computational methods to identify the posterior distributions of these parameters, efficient point estimates of their values, and Bayesian confidence limits for these estimates.

This poster's message: Multiple sequence alignments often contain considerable uncertainty; rather than using only one alignment of multiple nucleotide sequences or leaving the sequences unaligned in searches for *cis*-regulatory elements, we are using a Bayesian approach that searches through the joint parameter space of *cis*-regulatory elements and multiple sequence alignments. This permits information about plausible multiple sequence alignments to aid in the locating of elements and also allows the plausible locations of elements to aid in the determination of the relative quality of multiple sequence alignments. This combination enables us to better locate *cis*-regulatory elements.

The details: Key to understanding gene regulation in the biofuel pathways is the locating of the *cis*-regulatory elements that are recognized by regulatory transcription factors. Each regulatory factor energetically prefers a short sequence of nucleotides, typically 10-30 nucleotides in bacteria. There is much tolerance for variation in the recognized sequence and thus the computational search for a regulator's elements is a search for a set of similar short sequences that are present significantly more often than would be expected by chance. The search is through the genomic regions near genes that are likely to be co-regulated and across species that are likely to exhibit similar pathways of interest.

Quantification of the significance of abundance and similarity among *cis*-regulatory elements depends upon the hypothesized evolutionary relationships of the elements. For example, when input sequences are nearly identical the similarity of putative *cis*-regulatory elements is a less compelling indication of functional conservation than when the same amount of similarity is present in more distantly related sequences or in evolutionarily unrelated sequences; in the latter cases it is easier to conclude that the similarity is due to selective pressures associated with functionality rather than mere evolutionary nearness. Almost always, multiple sequence alignment (or its variant, tree sequence alignment) is the way that the orthologous relationships among nucleotides are specified. However, alignments and regulatory elements are not independent; a supposition of an alignment affects the apparent quality of a set of proposed elements and, vice-versa. As a result, a motif finding algorithm that searches for *cis*-regulatory elements using a multiple alignment built using a tool that is unaware of elements, and those that search in unaligned sequences both make sacrifices in the trade-off between sensitivity and specificity.

With a Markov chain Monte Carlo, Bayesian approach, we are simultaneously searching the space of possible sequence alignments and possible regulatory elements to find combinations that work well together to enhance the identification of unreported *cis*-regulatory elements.

Here we present preliminary results regarding the effectiveness of this Bayesian approach as applied to biofuel pathways in alpha-proteobacteria.

This research is supported by the Office of Biological and Environmental Research in the DOE Office of Science under a multi-institutional grant entitled "Bayesian computational approaches for gene regulation studies of bioethanol and biohydrogen production." (Grant No. DE-FG02-09ER64757 to C.E.L., DE-FG02-09ER64756 to L.A.N., and Grant No. FWP 55426 to L.A.M.)

Early Career Program

88

Optimizing Plant-Microbial Systems for Energy—Mapping Feedstock Quality Genes in *Brachypodium distachyon*

Scott J. Lee and **Samuel P. Hazen*** (hazen@bio.umass.edu)

Biology Department, University of Massachusetts, Amherst

Project Goals: (1) Quantify relevant genetic diversity for feedstock quality and plant by microbial strain interaction effects. We will phenotype a large germplasm collection of the model grass species *Brachypodium distachyon* to determine the overall genetic diversity for the rate of ethanol production using a range of *Clostridium phytofermentans* strains. (2) Utilize advanced quantitative trait loci (QTL) mapping strategies to identify genetic loci that control biofuel feedstock quality. Genetic linkage mapping will be carried out in immortalized recombinant inbred lines to provide awareness of the number of QTL that influence the traits, and their degree of influence. (3) Identify and evaluate candidate genes for improved energy systems using a multi-tiered approach to tag several QTL. Ultimately, we are interested in identifying the genes that have an effect on plant feedstock quality.

Creation of a sustainable biofuels industry depends on new technologies that are able to release the energy stored in cellulose fibers at a cost that enables widespread use of biofuels as a competitive domestic alternative to fossil fuels. Currently, use of biomass is hampered by the economic costs associated with thermochemical treatment of biomass, enzyme production, and biomass saccharification. This limits not only the use of cellulosic ethanol, but also the advancement of third generation biofuels. The purpose of our work is to develop plant-microbe systems that will reduce the expense of these steps. This will be accomplished in part by elucidating the genetic mechanisms underlying the attributes of plants that lead to facile decomposition with an eye on the interactions between specific feedstock genotypes and microbial systems. *Clostridium phytofermentans* can directly produce ethanol from plant biomass. Using a high throughput format, we quantify ethanol of inoculated biomass as an estimate of plant feedstock quality for our genetic studies [1, 2]. To better understand genetic diversity for feedstock quality, we will concentrate effort on a genetically diverse panel of accessions of the model species *B. distachyon*. To advance from variance components to genetic mechanisms, we will take advantage of high throughput genotyping and the remarkable genome sequence resources available today to map QTLs. We will begin with the most developed populations and expand to explore others whose parents demonstrate not only striking phenotypic divergence but also possess differential interactions with the microbial strains. This will help bring the genetic landscape into focus, further elucidating the number of genes and their relative influence on plant feedstock-microbe interactions. A key advantage of *B. distachyon* as a model system is the relative ease of creating stable transgenic plants. With that, we will make a series of gain- and loss-of-function mutants to test the function of candidate genes derived from mapping experiments. The three project goals are designed to determine the plausibility of specific positive interactions between plant and microbial genotypes. Pairwise comparisons will be made, varying both plant and microbial genotypes. Similar to adapting crop varieties to different environments, these experiments will link the need for specific feedstock properties to biomass conversion processes. Leveraging the characterization of cellulose-degrading microbes and discovery of plant traits and loci associated with microbial digestibility, we seek to identify the components of a successful plant-microbe system.

References

1. Lee SJ, Warnick TA, Leschine SB, Hazen SP (*In press*) A high-throughput biological conversion assay for determining lignocellulosic quality. In J Normanly, ed, Methods in Molecular Biology: High Throughput Phenotyping in Plants. Springer.
2. Lee SJ, Warnick TA, Pattathil S, Alvelo-Maurosa JG, Serapiglia MJ, McCormick H, Brown V, Young NF, Schnell DJ, Smart LB, Hahn MG, Pedersen JF, Leschine SB, Hazen SP (*In review*) Biological conversion assay using *Clostridium phytofermentans* to estimate plant feedstock quality.

This work was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy grants DE-FG02-08ER64700DE and DE-SC0006641.

Early Career Program

89

A Systems Biology, Whole-Genome Association Analysis of the Molecular Regulation of Biomass Growth and Composition in *Populus deltoides*

Leandro Gomide Neves[1]* (gomide@ufl.edu) and **Matias Kirst**[1,2]

[1]Plant Molecular and Cellular Biology Graduate Program and [2]School of Forest Resources and Conservation, University of Florida, Gainesville

Project Goals: Poplars are the principal short rotation woody crop species for providing clean, renewable and sustainable fuels in North America. Poplars are fast growing, have a perennial growth habit and wide natural distri-

bution that supports its implementation in a broad range of environments. While poplars provide the benefits of ideal bioenergy crop, the genes that regulate biomass productivity and composition are largely unknown, despite their critical relevance for efficient conversion to biofuels. Association genetics has become the primary approach for identification of genes that regulate such complex traits because it permits capturing a broad allelic range at high-resolution. Poplars are particularly suited to unveil the molecular basis of biomass productivity and composition using an association genetics approach because of low domestication, large open-pollinated native populations, and high levels of genetic and phenotypic variation. However, with few exceptions, association genetic studies in poplar and other plant species have been hampered by limited gene and polymorphism coverage.

In this project we are genotyping a large genetically unstructured population of *Populus deltoides*, one of the most widely distributed short-rotation woody crop in North America, using a combination of sequence capture and high-throughput genotyping. In parallel, the population is being phenotyped for biomass productivity and quality traits, for future discovery of marker trait associations. The overall goal is to identify the polymorphisms that accounts for differences in biomass productivity and lignocellulose composition in this species, that can be utilized for genetic improvement through transgenics or marker assisted breeding. Our specific project goals and progress towards their completion is described below:

Initially, a sequence-capture platform based on ultra-long RNA oligonucleotides has been developed based on a reference sequence of *P. trichocarpa* and a partial sequence of *P. deltoides*. The oligonucleotide library has been designed to capture sequences from over 22,000 genes. In total we have designed 220,000 probes that target coding as well as regulatory (promoter and UTR) regions of selected genes, and optimized sequence-capture procedures to multiplex 5 genotypes in each hybridization.

The second objective of this project is to sequence-capture target coding and regulatory sequences in a *P. deltoides* association population of 500 individuals. This population is composed of first generation collections made in the south and southeastern U.S. in the last two decades, and is expected to represents the genetic diversity of *P. deltoides* that exists in breeding programs. This work is currently in progress. We have also propagated the populations for trails in greenhouse and field. Greenhouse trials have been completed during the summer of 2011, when biomass data from above and below ground plant compartments were collected, as well as weekly growth data. Two field trials, in Florida and South Carolina, are being installed in the winter of 2011-2012. Initial biomass data collected from the greenhouse trials will be presented.

During the next year we anticipate the completion of sequencing and polymorphisms detection, to begin the analysis of association between markers and traits. In addition, we are exploring the possibility of using the sequence capture data to uncover gene copy number variation (CNV) in the population of *P. deltoides*. If the presence of variation in copy number becomes evident, we will assess possible associations with trait phenotypes.

90

Next Generation Protein Interactomes for Plant Systems Biology and Biomass Feedstocks Research

Mary Galli, Shelly Wanamaker, and **Joseph R. Ecker*** (ecker@salk.edu)

Salk Institute for Biological Studies, La Jolla, Calif.

Project Goals: Proteins execute most biological processes, but rarely do so alone. Therefore, large-scale efforts to interrogate the composition of the macromolecular complexes in which these proteins operate and their dynamic behavior are essential for a genome-wide understanding of cellular networks. To facilitate the identification of such interactions on a proteome-wide scale in plants, our project aims to create a high throughput protein-protein interaction system in which interacting pairs can be captured in the form of a nucleic acid readout and identified using next-generation sequencing.

Large-scale protein interaction maps are currently available for several model organisms including yeast, worm, fly, and human. In addition, we have recently reported the first large-scale plant interactome containing ~6200 novel interactions for nearly 3000 *Arabidopsis* proteins. While this effort resulted in the generation of the largest plant interaction dataset to date, we estimate it covers only ~2% of all protein interactions that exist in *Arabidopsis*. This observation is consistent with previous mapping efforts in other model organisms, which also suffered from low coverage. Therefore new technologies with the capability to significantly expand assay coverage are critical to enable comprehensive interactome mapping. This is especially pertinent for plants (in particular non-model plant systems with biomass-production potential) which contain a significantly higher number of protein coding genes relative to other eukaryotes.

We aim to develop an experimental approach that harnesses the unparalleled depth of next-generation sequencing and allows for the efficient and cost-effective interrogation of all pairwise interactions for all proteins encoded in a genome. To accomplish this goal, we have proposed to adapt the yeast 2-hybrid (Y2H) system, currently widely used in both small and large-scale protein interaction studies, to a format that is amenable with a high-throughput sequencing based output. Specifically, our strategy exploits the ability of CRE recombinase from the bacteriophage P1 to catalyze the physical linkage of two pieces of lox-containing DNA in vivo.
We have introduced non-revertible lox sites into a set of Gateway-compatible Y2H vectors containing bait and prey ORF-fusion cassettes. Upon transformation into an induc-

ible CRE-containing yeast strain, an interaction between bait and prey fusion proteins results in the reconstitution of a functional yeast GAL4 transcription factor which then drives expression of a growth-based reporter gene while simultaneously driving expression of CRE recombinase. The expression of CRE triggers the physical linkage of the bait and prey plasmids that encode the interacting proteins. Because the interacting pairs are locked together, large numbers of colonies that survive selection can be pooled *en masse* and processed as a single sample using next-generation sequencing technologies to identify the interacting pairs. A simple shearing strategy of the pooled, fused interacting DNAs followed by ligation with Illumina adaptors simplifies the sequencing library preparation process. Our current Illumina HiSeq platform yields over 600 Gigabases of DNA sequencing output per run and thus provides sufficient coverage for the identification of millions of interaction pairs.

Following initial proof of concept experiments using a preexisting set of *Arabidopsis* positive and random reference set clones, we plan to re-query a subset of the *Arabidopsis* Interactome pairwise space. This will allow us to obtain a comparative estimate of the increase in assay coverage and quality of our next-generation interactome approach. Ultimately, such a system has the potential to move beyond traditional time-consuming pairwise tests using well-defined ORF collections and accommodate library vs. library screens using ‘cDNA shatter libraries’. Flexible formats such as this will pave the way for the generation of interactome maps in organisms, such as potential biomass fuel stocks, for which no pre-existing ORF collections are available.

We gratefully acknowledge the U.S. DOE Office of Biological and Environmental Research in the U.S. DOE Office of Science for funding this project; DOE-DE-SC0007078.

91

Modeling of Cellulose, Hemicellulose and Lignin-Carbohydrate Complex Formation and Regulation to Understand Plant Cell Wall Structure

Vincent Chiang[1]* (vchiang@ncsu.edu), Ron Sederoff,[1] David Muddiman,[1] Cranos Williams,[1] Fikret Isik,[1] Joel Ducoste,[1] Chris Smith,[1] John Ralph,[2] and Sunkyu Park[1]

[1]North Carolina State University, Raleigh and [2]University of Wisconsin, Madison

Project Goals: The goal of this proposal is to use a systems approach to increase the fundamental understanding of the plant secondary cell wall. We will use multiple transgenic perturbations in *Populus trichocarpa* and measure effects on plants using advanced quantitative methods of genomics, proteomics, and structural chemistry. The combination of quantitative analysis, transgenesis, statistical inference and systems modeling provides a novel and comprehensive strategy to investigate the regulation, biosynthesis and properties of the secondary cell wall. We expect to answer two major questions: (1) To what extent can the abundance of individual transcripts and proteins of specific genes predict the quantity and composition of monomers and polymers of the cellulosic and lignin components? and (2) To what extent are there other new and necessary genes that regulate secondary cell wall properties?

This is a 3-year project and was recently funded. Here we focus our report on proposed research objectives.

Specific Objectives

1. **Transgene Perturbation:** We will generate transgenic *P. trichocarpa* with downregulated expression of 38 targeted cellulose and hemicellulose metabolic and *TF* genes. Artificial microRNA (amiRNA) and RNAi mediated gene specific knock-down experiments will be carried out in a differentiating xylem-specific manner at up to three levels: (a) the individual gene level, (b) the phylogenetic gene-pair level, and (c) the gene family level. For each transgene construct we will select transgenic lines with three distinct levels (highest suppression and two intermediates) of target gene knock-down to quantify transcriptomic, proteomic, and cell wall structural property responses to perturbations.
2. **Transcriptome Analysis:** Transgenics for each gene will be analyzed to test for specificity and interactive effects on transcripts, using new generation sequencing (Illumina GAIIX) and quantitative real-time PCR. We anticipate discovering the extent of specific and comprehensive feedback and feed-forward (loop) regulation associated with the organization and biosynthesis of cellulosic components in xylem secondary cell walls.
3. **Proteomic Analysis:** Changes in abundance of protein components will be determined by absolute protein quantification using Protein Cleavage coupled with Isotope Dilution MS (PC-IDMS)-based LC-MS/MS. Recombinant proteins will be produced from each target genes and purified for the identification of protein specific peptides (called surrogate peptides). Stable isotope labeled surrogate peptides will be synthesized and used as LC-MS/MS internal standards for absolute quantification of the target proteins in each transgenic line. The approach allows us to quantify the response of target proteins to specific gene perturbations and to correlate changes in the proteome and transcriptome with changes in secondary cell wall properties.
4. **Quantitation of Cellulosic Characteristics:** After quantifying Klason lignin content in stem wood of each transgenic line, the remaining mass in solution will be separated by HPLC to quantify the sugar components and the polysaccharide contents. Wood cellulose crystallinity will be estimated by ^{13}C-CP/MAS solid state NMR. Composition and inter-unit linkages of LCCs will be quantified by 2D NMR.
5. **Database and Website:** A database/website will be set up at the beginning of the project. All project data will be deposited into a relational database accessible through

our project website, forming the framework for statistical analysis and modeling.

6. **Statistical Analysis and Neural Networks:** Multivariate statistical (e.g. path analysis) methods will be explored to describe the degree, direction, and significance of relationships among all regulatory elements (TFs, gene transcripts and proteins) and cellulosic characteristics. Multilayer neural networks (MNNs) will be used as a quantitative approach to predict how variation in process protein concentrations influence cellulosic characteristics.
7. **Regulatory Signaling Graph**: Based on transgenic perturbations and statistical inference, modeling techniques will integrate experimental results to develop computational representations of cellulosic biosynthesis and structure. A mechanistic interaction graph will be developed to describe new regulation and to guide tests of feedback and feed-forward connections to verify functional interactions within the regulatory framework.
8. **Genome Wide Coregulated Network Analysis:** The large amount of new whole transcriptome and whole proteome information will be used for analysis of gene/protein variation to identify new coregulated associations of nontarget genes with known genes affecting secondary cell wall biosynthesis. The purpose of this analysis is to support and extend the structure of the signaling graph and the data-driven MNN models.

The outcome of this work will be a new and comprehensive description of the biosynthesis of cellulosic components of the plant secondary cell wall. This description will have predictive value and will serve to guide more efficient synthesis and degradation of the cell wall of energy crops for improved biofuels and biomass feedstocks.

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

92

Regulation of Neutral Lipid Accumulation in Vegetative Tissues of Plants

Kent D. Chapman[1]* (chapman@unt.edu), John M. Dyer,[2] and Robert T. Mullen[3]

[1]Center for Plant Lipid Research, University of North Texas, Denton; [2]USDA-ARS-ALARC, Maricopa, Ariz.; and [3]University of Guelph, Ontario, Canada

http://www.biol.unt.edu/~chapman/investigators/chapman/Projects/projects_frames/projectframeset.html (select Lipid Droplet)

Project Goals: Our research project aims to examine the cellular machinery that influences the accumulation of lipid droplets in non-seed tissues of plants. Our initial efforts are focused on a homolog of a mammalian hydrolase, *CGI-58*, that in humans is the causative gene for a neutral lipid storage disease. A loss-of-function mutation in this gene in Arabidopsis results in accumulation of triacylglycerols (TAGs) in leaf tissues that normally do not store neutral lipids. We proposed to 1) examine the role of AtCGI58 gene products in lipid droplet formation in Arabidopsis vegetative tissues 2) characterize the biochemical and cellular properties of CGI58 in the context of lipid metabolism. An improved understanding of the factors that regulate the accumulation of neutral lipids in vegetative tissues of plants may help to identify new strategies to enhance the energy content of crop plants.

Confocal laser scanning fluorescence imaging experiments indicated that *Arabidopsis cgi-58* loss-of-function mutants had significantly more lipid droplets in leaf mesophyll cells compared with wild-type cells. These lipid droplets were determined to be cytosolic and isolated lipid droplets were shown to contain triacylglycerols (TAG) with leaf-type fatty acids. While the numbers of lipid droplets were somewhat variable in leaf tissues of mature plants, there was a ten-fold increase in TAG levels in *cgi-58* mutants compared with wild-type plants. This lipid droplet/TAG phenotype was similar to that reported for peroxisomal mutants deficient in the uptake or catabolism of fatty acids (e.g., acylCoA oxidase, *acx1, acx2*; FA transporter, *cts1, pxa1*). There were no statistical differences in TAG levels of *cgi-58* seeds suggesting that the CGI-58 protein may act to influence lipid droplet abundance in non-seed tissues only, unlike the known peroxisomal mutants that confer a sugar-dependent germination phenotype. TAGs of *cgi-58* mutants were specifically enriched in molecular species that contained 18:3 (linolenic acid) fatty acid moieties. The enrichment in 18:3-containing lipids also was observed in major membrane galactolipids and phospholipids, suggesting that CGI-58 may participate specifically in the turnover of 18:3-containing glycerolipid species.

In animal systems, the CGI-58 protein is known to activate triacylglycerol lipases, and to transiently interact with the lipid droplet surface to mediate its functions. While the details of lipase activation in animal cells are not clear, we asked if the AtCGI-58 protein interacts with partner proteins to accomplish its function in lipid droplet biogenesis in plants. We implemented an unbiased approach to screen for interacting partners of CGI-58 using a yeast two-hybrid system to screen an *Arabidopsis* cDNA library. Several candidate interacting proteins were identified and we are currently following up one in detail—At4g39850 (*PXA1*, encoding peroxisomal fatty acid transporter 1, previously named *comatose, cts*, above). The interaction of CGI-58 with PXA1 might not be entirely unexpected, since the phenotype of the loss of *pxa1* (*cts*) function mutant shows an increase in cytosolic lipid droplets in leaves, similar to the *cgi-58* knockouts. Our interaction assays revealed that only the C-terminus of PXA1 interacts with CGI-58, specifically the so-called Walker B motif of nucleotide binding domain 2 (NBD2). This domain has been hypothesized by others to be involved in the regulation of the PXA1 transporter activity. Interaction between CGI-58 and PXA1 *in planta* was demonstrated through nuclear re-localization assays. In these experiments, the so-called Walker B motif of NBD2 (but not a mutated form of the Walker B motif, the Walker A motif nor the NBD1 domain of PXA1) containing a nuclear localization sequence (NLS) was able to redirect AtCGI-58-GFP protein to the nucleus when these two proteins were co-expressed in tobacco BY-2 cells, indicating that the Walker B motif of the PXA1 NBD2 domain is sufficient for CGI-58 interaction and that this interaction indeed occurs in the plant cytosolic environment. Additional transient expression assays in leaves with prolonged incubation have shown that in addition to a cytosolic location, CGI-58-GFP co-localizes to peroxisomes, consistent with a PXA1 interaction *in vivo*.

To test the functional association of CGI-58 and peroxisomal PXA1 beyond TAG accumulation, we turned to another "reporter" of PXA1 function, which is the uptake of 12-*oxo*-phytodienoic acid (oPDA) for its conversion to jasmonic acid. Indeed, whole seedlings or mature leaves of *cgi-58* knockout plants, when wounded, produced significantly less jasmonic acid compared with wild type plants. The reduction in wound-inducible jasmonate was to about 60% of wild-type levels, and there were no significant differences in jasmonate levels in unwounded tissues, suggesting that plants may have mechanisms in place to generate jasmonate independent of CGI-58. However, CGI-58 clearly influences the efficiency of jasmonate formation, likely through facilitating the uptake of oPDA by PXA1. These results corroborate our protein-protein interaction findings, and broaden the importance of CGI-58 function in plants to include modulation of signaling lipid metabolism in addition to regulating neutral lipid accumulation.

References (supported by this project)

1. James CN, Horn PJ, Case CR, Gidda SK, Zhang D, Mullen RT, Dyer JM, Anderson RG, Chapman KD (2010) Disruption of the *Arabidopsis* CGI-58 homologue produces Chanarin-Dorfman-like lipid droplet accumulation in plants. *Proceeding of the National Academy of Sciences (USA)* 107(41):17833-8.
2. Horn PJ, Ledbetter NR, James CN, Hoffman WD, Case CR, Verbeck GF, Chapman KD (2011) Visualization of lipid droplet composition by direct organelle mass spectrometry. *Journal of Biological Chemistry*, 286(5):3298-306.
3. Chapman KD, Ohlrogge JB (2011) Compartmentation of triacylglycerol accumulation in plants. *Journal Biological Chemistry*- Minireview- Thematic series Nov 16. [Epub ahead of print] PMID:22090025.
4. Chapman KD, Dyer JM, Mullen RT (2011) Biogenesis and functions of lipid droplets in plants. *Journal of Lipid Research*- Thematic Review Series Nov 1. [Epub ahead of print] PMID:22045929.
5. Horn PJ, Chapman KD (2011) Lipidomics in tissues, cells, and subcellular compartments. *Plant Journal* Nov 24. doi: 10.1111/j.1365-313X.2011.04868.x. [Epub ahead of print]. PMID: 22117762. Special Issue on High-Resolution Analysis of Plants.
6. Ohlrogge JB, Chapman KD (2011) The seeds of green energy—expanding the contribution of plant oils as biofuels. *The Biochemist* 33:34-38.
7. Dyer J, Mullen R, Chapman K (2012) Oil in biomass: a step-change for bioenergy production? *Inform (in press, scheduled for April 2012).*

Supported by the Office of Biological and Environmental Research in the DOE Office of Science (DE-SC0000797).

93

Investigations of Clusters of Cellobiose-Acid-$(H_2O)_n$

Madeleine Pincu[1]* (mpincu@uci.edu) and **Robert B. Gerber**[1,2]

[1]Department of Chemistry, University of Irvine, Irvine, Calif. and [2]The Fritz Haber Research Center, The Hebrew University, Jerusalem, Israel

The interactions between cellobiose and acids of different strengths, from very strong (H_2SO_4) to very weak (HCOOH), are studied in clusters of the type cellobiose-acid-$(H_2O)_n$. Ab initio molecular dynamics (DFT / BLYP with dispersion) is used to analyze the interaction between cellobiose and the micro-hydrated acids, in the gas phase, at room temperature. We compare these in terms of the probability for proton transfer (determined from Mulliken charges), to the active sugar site, as well as in terms of complete acid ionization. Initial findings show that in the case of cellobiose-HCl-$(H_2O)_n$, the sugar does not facilitate complete acid ionization as compared with the acid in pure water. The sugar, however does affect the pH of the acid. Detailed information on the preferred protonation site of cellobiose and the competition between the sugar and water protonation are presented. Differences in the ionization behaviour of different acids in complex with the sugar are discussed.

94

Production and NMR Analysis of Deuterated Cellulose and Lignocellulosic Biomass and Its Utilization for Neutron Scattering Studies

Barbara R. Evans[1]* (evansb@ornl.gov), Marcus Foston,[2] Art Ragauskas,[2] Hugh M. O'Neill,[1] Juhong He,[1] Sai Venkatesh Pingali,[1] Volker Urban,[1] Paul A. Langan,[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

http://www.ornl.gov

Project Goals: Lignocellulosic biomass from plant cell walls has the potential to play a major role in generation of renewable biofuels if cost-effective conversion can be achieved. However, it is a complex composite material that shows significant recalcitrance towards the structural deconstruction necessary for production of bioethanol. A fundamental understanding of the structural changes and

associations that occur between its component polymers at the molecular level during deconstruction is essential for enabling cost-effective 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" to provide this fundamental information about the structure and deconstruction of plant cell walls by integration of neutron science and computational simulation with physical and chemical characterization by NMR techniques.

In support of neutron science experiments, NMR methods were developed and demonstrated for analysis of deuterium incorporation in cellulose and plant biomass produced by culture of target species under deuterating conditions. The methods provide crucial supporting information for visualization of lignocellulose structure by neutron techniques and computational simulation.

Small angle neutron scattering (SANS) provides useful tools for investigation of lignocellulosic structural complexity and its changes during pretreatment and hydrolysis to sugars. The use of deuterated materials in which the non-exchangeable hydrogen atoms are replaced with deuterium can greatly increase the value of these studies, as the scattering patterns of the different components can be separated by phase contrast, enabling simultaneous observations of each component. Partially deuterated biomass samples were produced by cultivation of duckweed (*Lemna minor*), and the grasses annual rye (*Secale cereale*) and switchgrass (*Panicum virgatum)* in D_2O/H_2O mixtures. Samples of cellulose with specific levels of deuterium incorporation were obtained by cultivating *Acetobacter xylinus* in a defined deuteration medium. Using these samples, NMR methods were developed for quantifying the degree of deuterium substitution in the complete and individual components of the cell wall. A second NMR methodology that requires no sample preparation and is non-destructive was specifically designed and demonstrated to analyze deuterium incorporation in whole biomass samples. It was possible to measure the ratio of the integrals resulting from the subsequent ^1H and ^2H spectra. These values were correlated with known molar ratios of protons to deuterons from a calibration curve that were generated by taking several calibration standards formed from mixtures of glucose/glucose-d_7. The first SANS experiments have been carried out using the deuterated cellulose and plant biomass materials. Combined with ^{13}C-NMR, X-ray diffraction (XRD), and gel permeation chromatography (GPC), the new NMR techniques provide crucial chemical and structural information needed to leverage the full potential of SANS and building of real-time computational models of biomass deconstruction.

The Scientific Focus Area Biofuels is supported under FWP ERKP752 by the Office of Biological and Environmental Research in the U. S. Department of Energy Office of Science.

95

In situ Small-Angle Neutron Scattering and Computer Simulation Investigate Lignin Aggregation During Biomass Pretreatment

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

***In situ* small-angle neutron scattering (SANS) and extensive molecular dynamics (MD) computer simulation were used in a combined approach to examine real-time breakdown of biomass and the temperature dependence of lignin structure and dynamics.**

Lignin, a major polymeric component of plant cell walls, forms aggregates in vivo during pretreatment of lignocellulosic biomass for ethanol production. The aggregates are thought to reduce ethanol yields by inhibiting enzymatic hydrolysis of cellulose. Here, we report on real-time SANS experiments during dilute acid pretreatment of switchgrass and poplar biomass using a temperature-pressure reaction cell we developed. The temperature of the cell was ramped-up from 20 to 180 °C, then maintained at 180 °C for 60 min and finally the cell was cooled down to room temperature. A very clear characteristic structural feature appeared in the SANS data for $Q < 0.06$ Å^{-1} between 120–130 °C. This structural feature progressively moves to smaller-Q with increasing temperature and residence time at 180 ° C, indicating a growth in the particle size. We have identified the particle that appears at 120 °C as being lignin aggregates. To understand the temperature-dependent structure and dynamics of individual lignin polymers in aqueous solution we performed extensive (17μs) MD simulations. Between

150 °C and 210 °C, the lignin transitions from a glassy, compact state to a mobile, extended state. In all cases, the polymers were found to be globular particles, inside of which the R_g of a polymer segment is a power-law function of the number of monomers comprising it. In the low temperature states studied, the blobs are inter-permeable. In contrast, the particles at high temperatures become spatially separated, leading to a fractal crumpled globule form. Our simulations showed that the low-temperature collapse is thermodynamically driven by the increase of the translational entropy and density fluctuations of water molecules removed from the hydration shell. This combination of effects distinguishes lignin collapse from enthalpically driven coil-globule polymer transitions and provides a thermodynamic role for hydration water density fluctuations in driving hydrophobic polymer collapse. The detailed characterization obtained here provides insight at atomic detail into processes relevant to biomass pretreatment for cellulosic ethanol production. Specifically, our combined approach indicates that lignin aggregation occurs rapidly during heating—not during the cool-down phase, as was previously assumed.

96

Study of Plant Cell Wall Polymers Affected by Metal Accumulation Using Stimulated Raman Scattering Microscopy

Shi-You Ding[1]* (Co-PI: Shi.you.Ding@nrel.gov), and **X. Sunney Xie**[2] (PI: xie@chemistry.harvard.edu)

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

Project Goals: This project employs newly-developed chemical imaging techniques to measure, in real-time, the concentration, dynamics and spatial distribution of plant cell wall polymers during biomass growth with inoculation of transgenic symbiotic fungi. The proposed new pathway of biomass production will: 1) benefit metal and radionuclide contaminant mobility in subsurface environments, and 2) potentially improve biomass production and process for bioenergy. The goal of this project will explore a new pathway of delivering detoxicified metal to plant apoplast using transgenic symbiotic fungi, which will enhance metal accumulation from soil, and potentially these metals may in turn be used as catalysts to improve the efficiency of biomass conversion to biofuels. We further develop chemical imaging tools to quantitatively analyze the key plant cell wall polymers such as cellulose and lignin.

In our previous project funded by BER lignocellulose imaging program, we have demonstrated that stimulated Raman scattering (SRS) microscopy, a new imaging method, which allows real time observation of biomass conversion processes (1-4). SRS microscopy offers chemical contrast based on the intrinsic Raman vibrational frequencies in a sample with much shorter imaging time and easier spectral identification. 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 propose to further the recently emerging SRS techniques to imaging the distribution of plant cell wall polymers (e.g., cellulose and lignin) and cell wall architecture with different growth conditions including in heavy metal added soils and inoculated with symbiotic fungi. Specifically, we propose to:

- Establish a model system of plant-symbiotic fungi that produce metal-bound biomass by inoculating the poplar plants with transgenic fungus *Trichoderma harzianum* that carries secretory metal-binding protein.
- Evaluate fungal expression system growing in plant, and the efficiency of inoculation of fungi in intact plant cell walls. Specifically, we will examine if a closer contact between metals and cellulosic and lignin polymers caused by the entrapment of metals within plant cell, will alter plant cell wall structure and lead to a higher efficiency for cell wall degradation.
- Develop imaging tools to quantitatively analyze the distribution and deconstruction kinetics of plant cell wall polymers.

References

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).

This work is funded by the U.S. Department of Energy, the Office of Biological and Environmental Research (BER).

97

Defining Determinants and Dynamics of Cellulose Microfibril Biosynthesis, Assembly, and Degradation

Jocelyn Rose,[1]* (jr286@cornell.edu) Thomas Brutnell,[2] Seth DeBolt,[3] Detlef Smilgies,[1] and Larry Walker[1]

[1]Cornell University, Ithaca, New York; [2]Donald Danforth Plant Science Center, St. Louis, Missouri; and [3]University of Kentucky

Project Goals: The goals of this multidisciplinary project are to: (1) Establish platforms through reverse and forward genetics to identify and manipulate candidate genes that influence cellulose microfibril synthesis and structure; (2) Characterize the effects of altered candidate gene expression on cellulose microfibril synthesis and structure, and develop a mechanistic model for microfibril crystallization; (3) Determine the consequences of altering microfibril architecture on digestibility and integrate this information with nano-scale observations of enzymatic hydrolysis.

The central paradigm for converting plant biomass into soluble sugars for subsequent conversion to transportation fuels involves the enzymatic depolymerization of lignocellulosic plant cell walls by microbial enzymes. Despite decades of intensive research, this is still a relatively inefficient process, due largely to the recalcitrance and enormous complexity of the substrate. A major obstacle is still insufficient understanding of the detailed structure and biosynthesis of major wall components, including cellulose. For example, although cellulose is generally depicted as rigid, insoluble, uniformly crystalline microfibrils that are resistant to enzymatic degradation, the *in vivo* structures of plant cellulose microfibrils are surprisingly complex. Crystallinity is frequently disrupted, for example by dislocations and areas containing chain ends, resulting in "amorphous" disordered regions. Importantly, microfibril structure and the relative proportions of crystalline and non-crystalline disordered surface regions vary substantially and yet the molecular mechanisms by which plants regulate microfibril crystallinity, and other aspects of microfibril architecture, are still entirely unknown. This obviously has a profound effect on susceptibility to enzymatic hydrolysis and so this is a critical area of research in order to characterize and optimize cellulosic biomass degradation.

The entire field of cell wall assembly, as distinct from polysaccharide biosynthesis, and the degree to which they are coupled, are relatively unexplored, despite the great potential for major advances in addressing the hurdle of biomass recalcitrance. Our overarching hypothesis is that identification of the molecular machinery that determine microfibril polymerization, deposition and structure will allow the design of more effective degradative systems, and the generation of cellulosic materials with enhanced and predictable bioconversion characteristics.

We believe that the most effective way to address this long standing and highly complex question is to adopt a broad 'systems approach'. Accordingly, we have assembled a multidisciplinary collaborative team with collective expertise in plant biology and molecular genetics, polymer structure and chemistry, enzyme biochemistry and biochemical engineering. Our team will use a spectrum of cutting edge technologies, including plant functional genomics, chemical genetics, live cell imaging, advanced microscopy, high energy X-ray spectroscopy and nanotechnology, to study the molecular determinants of cellulose microfibril structure. Importantly, this research effort will be closely coupled with an analytical pipeline to characterize the effects of altering microfibril architecture on bioconversion potential, with the goal of generating predictive models to help guide the identification, development and implementation of new feedstocks. This project therefore spans core basic science and applied research, in line with the goals of this program.

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

98

Reduced Lignification Enhances Cellulose Accessibility in Wood Cell Walls

Catherine I. Lacayo,[1] Shi-You Ding,[2] Mona Hwang,[1] and **Michael P. Thelen**[1]* (mthelen@llnl.gov)

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

Project Goals: Effective deconstruction of plant cell wall lignocellulose is crucial to producing economical biofuels. The objective in this section of our SFA project (J. Pett-Ridge, Lead PI) is to dynamically monitor and quantitate the changes in lignin and polysaccharide content during various treatments of wood cells. We have focused in particular on investigating ways to reduce lignin content in walls, and how this influences enzyme accessibility to cellulose.

Most studies on cellulase enzymes use purified model substrates, usually either insoluble, crystalline cellulose, or soluble, chemically modified cellulose. While these substrates are good for assays such as measuring the release of glucose, they have dubious predictive value for the efficiency of chemical and enzymatic processing of plant feedstocks during biofuel production. Also, milled feedstock material such as switchgrass contain a heterogeneous mixture of cell types in different physical states, and therefore are difficult to use for statistical measurements in cell populations. A biologically relevant substrate for analyzing lignocellulose deconstruction is provided by cultured wood cells, which can be examined individually and in populations for the removal of lignin, cellulose, and hemicellulose. Toward this end, we established the *Zinnia* culture system for generating homogeneous suspensions of fully differentiated xylem cells, which are characterized by distinctly organized features found in all wood cell walls and are easily recognized by brightfield, fluorescence, and atomic force microscopy techniques.

Using this system, we characterized the topology of primary and secondary cell walls at nanoscale resolution, providing us with an architectural model of the *Zinnia* xylem cell (Lacayo et al, 2010, *Plant Physiol*). Lignin autofluorescence was diminished by about 90% in wood cells incubated for 6 hours in acidified sodium chlorite at 70 °C. By observing these oxidized cells in an environmental slide chamber under polarized light, we could track the intrinsic cellulose birefringence in secondary walls in real time during incubation with an endoglucanase. The results indicated that most of the cellulose was removed over a 3 hour period. This semi-quantitative analysis was confirmed by similar experiments in which fluorescence intensity was measured in cells probed with a GFP linked carbohydrate binding protein, *Ct*CBM3, which is specific for crystalline cellulose. When cellulose was enzymatically digested during the 3 hour incubation, GFP-*Ct*CBM3 binding decreased, leading to a statistically significant loss of fluorescence when compared to a control lacking cellulase. These studies provided the foundation for further experiments designed to determine the accessibility of cellulase and xylanase enzymes in secondary walls.

The amount of lignin in wood cells was also decreased during xylem development by the use of specific inhibitors of lignin biosynthesis. Culture medium was amended with reduced glutathione (GSH) or diphenyleneiodonium (DPI), and secondary wall differentiation proceeded normally except that autofluorescence and phloroglucinol staining were significantly lower than in control cells. After the treated cells were incubated with cellulase, GFP-*Ct*CBM3 fluorescence measurements indicated a statistically significant correlation of cellulose removal with a decrease in secondary wall lignin. Although the inhibition of lignin synthesis during differentiation was not as pronounced as removing lignin post-differentiation by chemical oxidation, both resulted in increased secondary cell wall accessibility to cellulase. Our studies confirm observations of enhanced saccharification in cellulase treated *Arabidopsis* lignin biosynthetic mutants, and provide a quantitative approach to assessing enzymatic efficiencies in a different model system.

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

99

Live Cell Identification of Redox Regulated Dithiol Sensors by Chemical Probes

Aaron T. Wright[1]* (aaron.wright@pnnl.gov), Natalie C. Sadler,[1] Na Fu,[1] Eric A. Hill,[1] Matthew R. Melnicki,[1] Thomas C. Squier,[1] Allan E. Konopka,[1] Diana J. Bigelow,[1] Erika M. Zink,[1] Michael P. Lewis,[1] Suchitra Datta,[1] Dian Su,[1] Wei-Jun Qian,[1] Don A. Bryant,[2] and **Alexander 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 long-term goal of the PNNL BSFA is to develop a predictive understanding of pathways and regulatory schemes involved in solar energy conversion to biofuel precursors or products. To date, knowledge gaps are not limited to biochemical pathways of energy and carbon flux but to an even larger degree include regulatory events within and between metabolic subsystems that include intracellular signals to which transcriptional regulation is responding. We suggest that "redox sensing," as a means to maintain redox homeostasis in photosynthesizing cells, is likely an equally important mechanism. In this project, we are investigating the central hypothesis that these sensors regulate electron flux and carbon partitioning in cyanobacteria. Specifically, we have developed and applied probe approaches for live cell capture, characterization, and imaging of redox sensors of environmental change using advanced separations, mass spectrometry and traditional biochemical techniques. It is anticipated that the identification of redox-sensitive dithiol linkages and their modulation by thioredoxin (Trx), peroxiredoxin, and other regulators will provide key inputs for understanding the control points of flux distribution.

The primary challenge in identifying redox-regulated dithiol sensors *in vivo* is that cysteine residues are highly reactive and easily oxidized to dithiol linkages following cellular lysis. This eliminates the ability to experimentally measure redox regulation within native physiological settings, and makes requisite the exogenous addition of chemical or biological reductants. Therefore, we developed two different but complementary approaches for synthesizing and employing cell permeable probe reagents that react in live cells (*in situ*) with thiols and permit real-time imaging and mass spectrometric characterization of probe targets. In the first approach we have synthesized click-chemistry enabled chemical probes for fluorescent and mass spectrometric identification of redox regulated dithiol sensors, and in the second approach we used a cell-permeable arsenic probe (TRAP-Cy) that selectively binds to reduced disulfides in close proximity to trap available dithiols in living cells prior to cell lysis. We have used our probes *in situ* to identify proteins in *Synechococcus* sp. PCC 7002 that under-go disulfide exchange in response to changes in cellular conditions.

Click-Chemistry Enabled Probes. Probes were synthesized with three chemical elements: a moiety to impart cell permeability, an iodoacetamide or maleimide group for irreversibly labeling cysteines, and a reporter tag for detection and isolation of probe-labeled proteins. We exploited the multimodal bio-compatible click chemistry (CC) reaction to create "tag-free" probes for profiling proteins in living systems. Probe-labeled proteins were visualized by addition of a complementary azido-tetramethylrhodamine (fluorescence)

* Presenting author

or an azido-biotin tag for enrichment and mass spectrometric analysis (LC-MS).

The maleimido and iodoacetamide probes were added simultaneously *in vitro* or *in situ* to *Synechococcus* sp. PCC 7002 grown in a turbidostat under maximal growth rate conditions. Cells were also removed and placed in the dark for two hours and then labeled *in vitro* or *in situ*. Following probe-labeling cells were lysed and probe-labeled proteins attached to biotin via CC. The probe-labeled proteins were then enriched on streptavidin, digested with trypsin, and the peptides analyzed by LC-MS. Critically, in the cells that were labeled post-lysis (*in vitro*) we found no changes in redox regulation of dithiol sensors, demonstrating that lysis rapidly oxidizes biological samples. However, when we analyzed the *in situ* labeled cells we identified 153 redox-regulated proteins that were statistically different between the light and dark conditions. In a follow-up study using a carbon-limited turbidostat we were able to identify redox changes *in situ* within 30 seconds following the addition of CO_2. A time-course study revealed remarkable changes from 30 seconds to 60 minutes post CO_2 addition. Importantly, these changes cannot be measured by traditional transcriptomic or proteomic measurements because these redox events occur within a time-frame that protein and mRNA content has not changed.

Aresenic Probes. TRAP-Cy contains a cyanine dye to facilitate high-throughput gel-based screening prior to affinity purification and MS based identification of reactive dithiols. Using this trapping reagent, we have explored redox-active dithiol reactivity and their relationship to photosynthetic metabolism in *Synechococcus* sp. PCC 7002 under varied environmental conditions. Upon cell lysis, covalently bound cysteines in association with TRAP-Cy are stable, permitting reduction of all remaining thiols and their alkylation. After treatment with excess ethane dithiol, the TRAP-Cy probe is released, permitting enrichment of these cysteines using a thiol-capture affinity resin. In the vast majority of cases homology models of the structures demonstrate that identified thiols are in close proximity in the tertiary structure. Consistent with our hypothesis that dithiol “switches” (e.g., Cys-Xxx_n-Cys) activated by thioredoxin-dependent pathways regulate electron flux and carbon sequestration to modulate energy partitioning.

Probe-identified proteins from both probe types map well onto multi-subunit supramolecular complexes involving photosynthetic pathways associated with efficient collection of excitation energy (light harvesting), electron transfer reactions linked to formation of electrochemical gradients, carbon dioxide sequestration (dark reactions), and ATP synthesis. Additional redox-dependent pathways include those involving chaperone activity, transcriptional regulation, and antioxidant proteins linked to protein repair. Together, these results provide quantitative information regarding redox-dependent switches associated with photosynthetic regulation, and provide a systems biology tool capable to providing high-throughput information necessary for predictive metabolic modeling. Finally, both cell permeable probe approaches represent the only existing methods for identifying and imaging live cell redox regulation, and they will be critical to informing the predictive models of metabolism needed for bioenergy applications.

This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Biofuels Scientific Focus Area. Mass spectrometry-based proteomics measurements were performed in the William R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by OBER and located at PNNL.

100

Effects of Nutrient Limitation, Light Quantity and Quality on Growth Dynamics and Physiology of *Synechococcus* sp. PCC 7002

Allan E. Konopka[1]* (allan.konopka@pnnl.gov), Grigoriy E. Pinchuk,[1] Eric A. Hill,[1] Leo A. Kucek,[1] Matthew R. Melnicki,[1] Trang Vu,[2] Jennifer L. Reed,[2] and **Alex S. Beliaev**[1]

[1]Pacific Northwest National Laboratory, Richland, Wash. and [2]University of Wisconsin–Madison

Project Goals: The PNNL Biofuels Scientific Focus Area (BSFA) conducts fundamental research of microbial photoautotrophs with specific emphasis on photosynthetic energy conversion, reductant partitioning, and central carbon metabolism. As a model system, we utilize unicellular prokaryotic organism *Synechococcus* sp. PCC 7002, which exhibits one of the fastest growth rates known among cyanobacteria and which is also remarkably tolerant to high light intensities. Understanding the genetic and physiological bases of these properties could provide fundamental new insights that are broadly applicable to the optimization of other biological systems for biofuels development. To that end, we are interrogating fluxes through central metabolic pathways to define the major constraining factors (*i.e.* metabolic and regulatory controls) governing carbon partitioning through the metabolic subsystems of cyanobacteria that can be manipulated to increase yields of specific molecules that are either precursors or fuel molecules themselves.

The initial experimental phase, which consists of analyses of photosynthesis and carbon fixation modules, employs turbidostat cultivation technology which allows the organisms to grow at their unrestricted, maximum growth rate. This approach is suited particularly well for photoautotrophs since attenuation of light by self-shading and mass transfer of CO_2 must be taken into account. We are utilizing a custom-designed photobioreactor (PBR) to rigorously control the physiological state of *Synechococcus* 7002, and analyze the properties of both wild type (and subsequently) discrete mutant strains. PBR illumination is provided by light-emitting diodes (LED) generating 680 nm and 630 nm light for the preferential excitation of chlorophyll *a* and phycobilin pigments. The continuous culture system was modified to control medium input to maintain constant turbidity (turbidostat); this produces a steady state in which growth rate is unrestricted by dissolved nutrient concentrations, but determined by either light irradiance, or cellular capacity (at saturating irradiances). *Synechococcus* 7002 grew when irradiated with only 680 or only 630 nm light, with the latter resulting in higher growth rates; notably, the growth rate differences were greatest at the lowest irradiances. Steady-state dissolved O_2 concentrations were linear functions of irradiance; the slope was 2.3-fold greater when a low amount of 680 nm irradiance was combined with 630 nm light than in cultures grown under 630 nm irradiances alone. When irradiance values were changed, the culture experienced either an increase (shift-up) or decrease (shift-down) in growth rate. After shift-downs, the new steady-state growth rate was usually achieved within 3 h. In contrast, several shift-up transitions required >10 h to reach the new growth rate. The physiological consequences of shifts in light quality and quantity have been investigated using PAM fluorometry to calculate the maximum capacity for photosynthetic electron transport (rETR) and quantum yield of photosystem II (F_v/F_m).

Steady-state chemostat cultures have also been generated and are undergoing analysis of cellular composition, in conjunction with constraint-based modeling and global expression analyses. Significant differences in macromolecular composition have been found among unrestricted growth and light, carbon, or nitrogen-limited cultures and are consistent with previous understanding of macromolecular composition under these conditions. Protein content varied from 33% (N limitation) to 67% (light limitation). Under N limitation, polysaccharide comprised 61% of biomass, but was 10-17% under all other conditions. These measurements will be used to formulate the organism-specific biomass equation and incorporated into developed metabolic model of *Synechococcus* 7002 (*i*Syp612). The initial model was updated with additional genome annotation from pathway/genome databases (PGDBs) and includes 672 genes and 572 reactions. The new information from PGDBs resulted in the modifications of existing gene-protein-reactions associations (GPRs) due to addition of isozymes. A few (~20) new reactions were added to the pathways involved in quinone biosynthesis, carbohydrate, amino acid and nucleotide metabolism. Using the developed model, we will estimate the growth and non-growth associated ATP requirements (GAR, and NGAR, respectively) using the data described above for cultures grown under various light intensities. Model predicted growth, uptake, and secretion rates will be compared with experimental values for cells grown under different chemostat conditions.

This research is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Biofuels Scientific Focus Area.

Biological Hydrogen Production
Systems Biology and Metabolic Engineering Approaches

101

Characterization of H_2 Producing Microbial Communities of the Wood-Feeding Beetle *Odontotaenius disjunctus* by Multi-Scale Measurement of Metabolic Function

Javier A. Ceja-Navarro[1]* (JCNavarro@lbl.gov), Xavier Mayali,[2] Peter K. Weber,[2] Mari Nyyssönen,[1] Zhao Hao,[1] Hoi-Ying Holman,[1] Meredith M. Blackwell,[3] **Jennifer Pett-Ridge**,[2] and **Eoin L. Brodie**[1]

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

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/metagenomic, and computational approaches, to understand nutrient cycling and biofuel (H_2) 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. Our ultimate goal is the development of multi-scale models that can predict ecological and biochemical relationships within multi-trophic microbial systems.

One of our model systems is represented by the hindgut of the passalid beetle *Odontotaenius disjunctus* (Fig. 1A). This wood-feeding beetle is able to survive on a low nutrient diet through symbiotic interaction with its gut microbiome, which provides nutrients from plant polymer decomposition and nitrogen through fixation. Both of these processes result in the production of hydrogen gas.

The digestive system of the passalid beetle is compartmentalized in 4 morphologically differentiated regions (Fig. 1D); each has been characterized with our integrated approach, which consists of:

1. Culture independent techniques used to assess the diversity and composition of the microbial communities inhabiting each gut region by using the PhyloChip, together with analysis of community gene expression by qPCR, and sequencing of enriched cDNA libraries.
2. Use of Chip-SIP, a high-sensitivity, high-throughput stable isotope probing (SIP) method performed on a phylogenetic array to directly measure functional roles of uncultivated microorganisms.
3. Construction of Fosmid libraries for enzymatic assays and sequencing to search for the genes involved in the transformation of specific substrates. We have optimized HMW-DNA extraction and library construction protocols for their application to the beetle gut microbiome.
4. Micron-scale measurement of physicochemical gradients (O_2, H_2, pH) on the distinct gut regions of the passalid beetle using microelectrodes (Fig 1C).
5. Characterization of changes in the chemical composition of digested/undigested substrate (oak wood) using Fourier Transform Infrared Spectroscopy (FTIR).

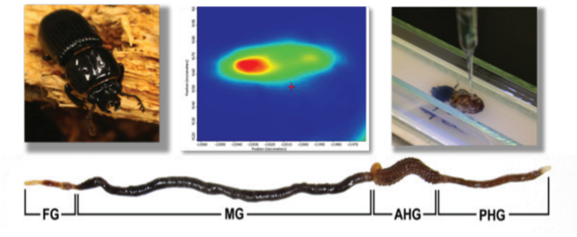


Figure 1. A) The passalid beetle, B) SIR image of undigested oak wood, C) arrangement of the dissected beetle and H_2 microsensor, D) the compartmentalized gut of the passalid beetle, F=Foregut, MG = Midgut, AHG= Anterior hindgut, and PHG =Posterior Hindgut.

By combining these complementary approaches we are attempting to determine the primary processes and organisms contributing to energy flow in this natural system. We have characterized the microbial communities of the passalid beetle, obtaining the highest levels of diversity in the Foregut (FG), followed by Posterior Hindgut (PHG), Midgut (MG), and Anterior Hindgut (AHG). *nifH* expression analysis showed that N_2 fixation was highest in the AHG and sequencing of *nifH* genes identified the dominant N_2 fixing group as members of the Porphyromonadacea (Clostridiales). Concentrations of O_2 and H_2 varied across the different gut compartments; the highest nifH gene expression and lowest bacterial diversity spatially correspond to regions with high concentrations of H_2 production and O_2 is at its minimum.

Using FTIR, we have also characterized changes in the composition of the beetle's substrate (Oak wood); results suggest cellulose and hemicellulose content declines as the substrate passes through the sequential gut regions. We are currently screening fosmid libraries, and performing analysis of enriched cDNA libraries and Chip-SIP microarrays.

This project was funded by the Department of Energy through the Genome Sciences Program. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. This work was supported, in part, by the LDRD project 07-ERD-05,UCRL: POST-

401211. Part of this work was performed at Lawrence Berkeley National Laboratory under the auspices of the University of California—Contract number DE-AC02-05CH11231.

102

Anoxic Carbon and Nitrogen Metabolism in Photosynthetic Microbial Mats as Revealed by Metatranscriptomic and NanoSIMS Studies

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

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

Photosynthetic microbial mats found in coastal environments are complex, stratified microbial communities that fix CO_2 during the day and ferment the fixed carbon under anoxic conditions at night, generating significant amounts of H_2 and organic acids. Fermentation of accumulated photosynthate may be required to provide energy for anoxic N_2 fixation. To understand C and N flux through these mats at a single cell and community level and to relate this flux to the observed rates of H_2 evolution and N_2 fixation, we have analyzed the upper 2 mm of cyanobacterial mats collected from Elkhorn Slough, CA, which have been shown to possess high rates of fermentation and N_2 fixation by metatranscriptomic and NanoSIMS studies.

We characterized the active microbial community from two mats samples collected under dark, anoxic conditions using metatranscriptomic sequencing to define the metabolic pathways for fermentation and N_2 fixation. Clone library and 454 pyrosequencing of the rRNA expressed by this community demonstrated that the active community was dominated by filamentous cyanobacteria including populations closely related to *Microcoelus chthonoplastes* and a novel diazotrophic cyanobacterial isolate, ESFC-1. Approximately 388,000 transcripts (average length ~390 bp) were obtained by sequencing amplified cDNA. Of these transcripts, ~199,000 transcripts (51%) were classified as protein coding sequences by comparison to the SEED database (bit score>40). Transcripts related to *Cyanobacteria* and *Chloroflexales* represent >70% of the total classified sequences in the transcriptome, suggesting that representatives of these phyla are active in the mats under anoxic conditions.

Metabolic reconstruction of the *Microcoleus chthonoplastes* genome identified a complete pathway for fermentation of glycogen to lactate, acetate, formate, ethanol and H_2. Fragment recruitment of metatranscriptomic reads indicated that all the genes in this fermentation pathway were expressed, consistent with organic acid and H_2 evolution measurements. A large number of metatranscriptomic reads affiliated with the *Chloroflexales* were annotated as acetate/lactate permeases and polyhydroalkonoate synthases, suggesting that the *Chloroflexales* may metabolize the organic acids produced by the cyanobacteria further and also synthesize carbon storage compounds to be utilized in photoheterotrophic growth. Assembly of the metatranscriptomics reads revealed that partial *nif* operons were reconstructed with the structural nitrogenase genes (*nifHDK*) for *Microcoleus*, ESCF-1 and an uncultivated member of the *Chloroflexales*, suggesting that these populations were responsible for the observed N_2 fixation activity.

These metatranscriptomic studies were complemented by NanoSIMS measurement of mat samples collected from Elkhorn Slough and incubated with isotopically labeled substrates. Incubation of mat samples with $^{13}HCO_3^-$ provided evidence for the fixation and storage of carbon by *Microcoleus* populations during the day and release of the stored carbon under anoxic conditions. Addition of ^{13}C-labeled acetate resulted in isotopic enrichment of the *Chloroflexi*, providing further evidence for the anoxic food web hypothesis generated by analysis of the mat metatranscriptome. Incubation of the mats with $^{15}N_2$ demonstrated that cyanobacterial populations related to ESFC-1 were highly active in N_2 fixation under dark anoxic conditions; however no evidence was found for incorporation of $^{15}N_2$ into *Microcoleus* and *Chloroflexi*.

103

Development and Validation of Novel Techniques to Understand and Address Limitations to Algal Hydrogen and Biofuels Photoproduction

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

[1]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, nonpolluting 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. In Task 1, we developed a Rhodobacter-based H_2-sensing assay that is useful for conducting high-throughput screening for algal H_2-production mutants under fermentative conditions. The assay incorporates an agar overlay of H_2-sensing *R. cap-*

sulatus cells that fluoresce in response to H_2 produced by an agar underlay of *C. reinhardtii* colonies. The assay is sensitive enough to distinguish one H_2-producing colony from out of thousands of non H_2-producing colonies as present on a single Petri dish. We will focus next on adapting the assay for operation under photobiological H_2-production conditions and use it to screen for microbial libraries with increased H_2 photoproduction. In parallel work, we have transferred an exogenous hydrogenase from *Clostridium acetobutylicum,* along with its attendant assembly proteins, into the H_2-reporter bacterial strain. This bacterium produces hydrogen from the exogenous hydrogenases and emits a fluorescent signal in response to its H_2 production. We will focus our work on using this system to screen for O_2-tolerant [FeFe]-hydrogenases generated through directed-evolution techniques.

In green algal chloroplasts, ferredoxins (FDXs) play a central role in the allocation of low-potential electrons from photosynthesis to the hydrogenase enzyme, as well as to numerous competing assimilatory pathways. These pathways include CO_2 fixation, nitrite reduction, glutamate synthesis, sulfite reduction, cyclic electron transport around Photosystem I, and reduction of thioredoxin for regulation of biosynthetic pathways. Under certain physiological states, these processes have the potential to compete with the hydrogenases for electrons. Moreover, there are 6 potential FDXs that could interact with the hydrogenase enzymes (HYDA1 and HYDA2) and it is not clear yet which one is the favored natural electron donor to hydrogenases under different physiological conditions. To address the issue of competitive metabolic pathways with H_2 production, Task 2 has used a two-hybrid assay to screen *C. reinhardtii* expression libraries, with HYDA2 and HYDA1 as well as the 6 FDXs as baits, followed by pull-down assays to confirm *in vitro* the results obtained *in vivo*. We have also examined pair-wise interactions between each hydrogenase and FDX isoform to identify specific protein-protein interactions. These approaches have resulted in the identification of novel protein-protein interactions that are leading to the development of a map of the network centered on ferredoxins and involving multiple metabolic pathways in *Chlamydomonas*.

Publications

1. Wecker MSA, Meuser JE, Posewitz MC, Ghirardi ML. Design of a new biosensor for algal H_2 production based on the H_2-sensing system of *Rhodobacter capsulatus. International Journal of Hydrogen Energy* 2011, 36(17):11229-11237.
2. Peden E, Boehm M, Davis RA, Old W, Ghirardi ML Dubini A. Identification of the chloroplastic ferredoxins metabolic network in *Chlamydomonas reinhardtii. In preparation.*

104

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

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

[1]Departments of Biology and [2]Energy, Environmental and Chemical Engineering, Washington University, St. Louis, Mo.; [3]Department of Biological Sciences, Purdue University, West Lafayette, Ind.; and [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.

Comparative genomics and hydrogen production in Cyanothece: Analysis of the complete genome sequences of six *Cyanothece* strains (ATCC 51142, PCC 7424, PCC 7425, PCC 7822, PCC 8801, and PCC 8802) indicates that this genus maintains a plastic genome, incorporating new metabolic capabilities while simultaneously retaining archaic metabolic traits, a unique combination which provides the flexibility to adapt to various ecological and environmental conditions (Bandyopadhyay et al., 2011). Our study uncovered the presence of a large and contiguous nitrogenase gene cluster in four of the five newly sequenced strains. These strains were also analyzed for their ability to fix nitrogen and produce hydrogen. High rates of aerobic nitrogen fixation and hydrogen production were observed in five of the six strains, distinguishing *Cyanothece* as a genus of unicellular, aerobic nitrogen-fixing cyanobacteria. The contiguity of the nitrogenase gene cluster in *Cyanothece* sp. 7425 is interrupted by a 2.5Mbp insertion and this cluster exhibits a rearrangement of the *nif* genes as seen in some of the anaerobic nitrogen fixing *Synechococcus* strains. *Cyanothece* 7425 also appears to have lost its aerobic nitrogen-fixing ability and exhibits nitrogen fixation and hydrogen production only under anaerobic conditions. In addition, our study revealed the presence of 3 linear chromosomal elements in *Cyanothece* 7822. These linear elements, like the linear chromosome in

* Presenting author

Cyanothece 51142, harbor genes involved in various carbohydrate metabolism pathways and represent a unique feature of the genus *Cyanothece* compared to other sequenced cyanobacteria. A comparison of the genomes of the six *Cyanothece* strains revealed the presence of several pathways involved in fermentative metabolism, traits usually observed in non-oxygenic microbes. The ability of this genus to maintain a suboxic intracellular environment for a significant part of a diurnal cycle is unique and allows the *Cyanothece* cells to provide a platform for both aerobic and anaerobic metabolic processes. Our analysis suggests that members of the genus *Cyanothece* can be appealing as model organisms for studies pertaining to biohydrogen production.

Metabolomic studies: We have developed a web-based platform (MicrobesFlux) for generating and reconstructing metabolic models for annotated microorganisms. The MicrobesFlux is able to load the metabolic network (including enzymatic reactions and metabolites) of over 1,100 species from KEGG database (Kyoto Encyclopedia of Genes and Genomes) and then automatically converting it to a metabolic model. The platform also provides diverse customized tools, such as gene knockouts and introduction of heterologous pathways, for users to redefine the model network. The reconstructed metabolic network can be formulated to a constraint-based flux model to predict and analyze the carbon fluxes in the metabolisms. The simulation results can be output in SBML format (The Systems Biology Markup Language). The MicrobesFlux is an installation-free and open-source platform that enables biologists with little programming knowledge to develop metabolic models for newly sequenced microorganisms. Our system allows users to construct metabolic networks of organisms directly from the KEGG database. It also provides users with predictions of microbial metabolism via flux balance analysis. This prototype platform can be a springboard to advanced and broad-scope metabolic modeling of complex biological systems by integrating other "omics" data or ^{13}C-assisted metabolic flux analysis results. This platform is being used for generating metabolic models of the *Cyanothece* strains and for predicting gene knockouts and pathway modifications. MicrobesFlux is available at http://tanglab.engineering.wustl.edu/static/MicrobesFlux.html.

Proteomic Studies: A comparative analysis of the proteomes of the six *Cyanothece* strains was carried out to identify similarities and differences in the metabolism of the strains grown under nitrogen sufficient conditions. Among the five strains, the largest number of proteins (3967) could be detected in *Cyanothece* PCC7424 and the lowest (2728) in *Cyanothece* PCC7425. Our results reveal considerable similarities in proteins involved in conserved metabolic and biochemical pathways such as photosynthesis, respiration, N_2-fixation, H_2-production as well as proteins involved in housekeeping functions. Out of a total of 1369 predicted orthologs common to all *Cyanothece* strains, 644 have been detected in the comparative study. However, considerable differences were also evident, particularly in *Cyanothece* PCC7425 where significantly higher numbers of proteins (682 proteins) were uniquely expressed. This may suggest that proteins in *Cyanothece* PCC7425 diverged separately compared to the other strains, possibly as a result of specific selection pressure and subsequent adaptation to environment. Furthermore, in addition to commonly shared phycobilisome (PBS) proteins, we identified 3 phycobiliproteins and 2 PBS-linker polypeptides expressed specifically in PCC7822, PCC7424 and PCC8801 together with phycoerythrin (PE)-operon protein, suggesting their association with PE assembly. While homologs of PSI core proteins PsaAB were identified in all six strains, expression of another PsaB protein (Cyan7822_4989) was only detected in PCC7822.Our results elucidate proteomic footprints of six *Cyanothece* strains, and provide useful targets for future applications in biotechnology.

To determine cellular factors involved in H_2 production, we employed label-free quantitative proteomics in *Cyanothece* ATCC 51142 and PCC 7822 under eight culture conditions. Rates of H_2 production were measured for each condition and compared to changes of protein abundances. Our analysis provides systems level understanding about nitrogenase-mediated H_2 production under aerobic condition providing useful insights into how *Cyanothece* might be optimized for H_2 production.

Publication

1. Bandyopadhyay A, Elvitigala T, Welsh EA, Stöckel J, Liberton M, Min H, Sherman LA, Pakrasi HB (2011). Novel metabolic attributes of *Cyanothece*, a group of unicellular nitrogen fixing cyanobacteria. *mBio* 2 (5): e00214-11. doi:10.1128/mBio.00214-11.

Project number—DE-FC02-07ER64694
This work was supported by funding from DOE-BER (DE-FC02-07ER64694).

105

Contrasting the Metabolic Capabilities of *Cyanothece* 51142 and *Synechocystis* 6803

Rajib Saha[1]* (rus184@psu.edu), Bertram M. Berla,[2] Thomas J. Mueller,[1] Thanura Elvitigala,[3] Lawrence E. Page,[3] Louis Sherman,[4] **Himadri B. Pakrasi**,[2,3] and Costas D. Maranas[1]

[1]Department of Chemical Engineering, The Pennsylvania State University, University Park; [2]Department of Energy, Environmental, and Chemical Engineering, Washington University, St. Louis, Mo.; [3]Department of Biology, Washington University, St. Louis, Mo.; and [4]Department of Biological Sciences, Purdue University, West Lafayette, Ind.

Project Goals: The current project uses systems biology to define the metabolic networks involved in the photobiological production of advanced biofuels and/or their chemical precursors by cyanobacteria. By encasing transcriptomics, metabolic profiling, mutagenesis, physiological analysis and genome-scale metabolic modeling in a systems biology framework, our goals are: i) to develop genome-scale models of two widely used cyanobacterial

* Presenting author

species, *Cyanothece* 51142 and *Synechocystis* 6803, ii) to refine these models based on systems level information obtained from experimental studies (i.e., transcriptome, metabolome, etc.), and iii) to genetically modify these strains based on model predictions, experimentally evaluate their performance and subsequently enhance the production yield of desired products.

Table 1: *Synechocystis* 6803 iRS706 and *Cyanothece* 51142 iRS764 model statistics

	***Synechocystis* 6803 *iRS*706 model**	***Cyanothece* 51142 *iRS*764 model**
Included genes	706	764
Proteins	604	660
Single functional proteins	258	247
Multifunctional proteins	192	210
Protein complexes	3	4
Isozymes	27	30
Multimeric proteins	107	149
Others[a]	17	20
Reactions	1,844	1,910
Metabolic reactions	1,834	1,900
Transport reactions	5	5
GPR associations		
Gene associated (metabolic/transport)	1,798	1,836
Spontaneous[b]	18	19
Nongene associated (metabolic/transport)	18	40
Exchange reactions	5	5
Metabolites	1,894	1,862
Cytoplasmic	1,889	1,857
Extracellular	5	5

[a]Others include proteins involve in complex relationships, e.g. multiple proteins act as protein complex which is one of the isozymes for any specific reaction.

[b]Spontaneous reactions are those without any enzyme as well as gene association.

Cyanobacteria are an important group of photoautotrophic organisms that can synthesize valuable bio-products by harnessing solar energy. Cyanobacteria contribute significantly to biological carbon sequestration, O_2 production and the nitrogen cycle. They exhibit robust growth under diverse environmental conditions and have minimal nutritional requirements. They are also endowed with high photosynthetic efficiencies and diverse metabolic capabilities that confer the ability to convert solar energy into a variety of biofuels and their precursors. However, less well studied are the similarities and differences that exist between the metabolic capabilities of different species of cyanobacteria that may contribute toward their niche biological functions and their suitability as microbial production systems. Towards goal (i) above, here we introduce and compare genome-scale models for two phylogenetically related cyanobacterial species, namely *Cyanothece 51142 i*RS764 and *Synechocystis 6803 i*RS706. Model *i*RS764 is comprised of 764 genes, 1,910 reactions and 1,862 metabolites, whereas model *i*RS706 spans 706 genes, 1,844 reactions and 1,894 metabolites. As many as 1,628 reactions and 1,702 metabolites are common in these two models. All reactions are elementally and charge balanced and localized in three different intracellular compartments (i.e., cytoplasm, carboxysome and thylakoid lumen). GPR (gene-protein-reaction) associations are also established based on the functional annotation information and homology prediction. We contrast flux balance analysis results under different physiological conditions, (i.e., phototrophic, chemotrophic and mixotrophic) and explore their impact on biomass formation, nitrogen fixation and hydrogen yield of these two cyanobacterial species.

Additionally, towards goal (ii) above, we have begun characterizing genetic elements active at different phases of cyanobacterial culture, including during exponential growth and stationary phase. We have identified putative promoters that are highly active at stationary phase and have constructed a library of expression vectors that target genes to areas of the genome that are highly transcribed during stationary phase.

Project number—DE-SC0006870

106

Photophysiology and Energetics of Light-Driven H_2 Production by *Cyanothece sp.* ATCC 51142

Matthew R. Melnicki* (matthew.melnicki@pnnl.gov), Eric A. Hill, Leo A. Kucek, Grigoriy E. Pinchuk, Allan E. Konopka, and **Alex S. Beliaev**

Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: The PNNL Biofuels Scientific Focus Area (BSFA) conducts fundamental research of microbial photoautotrophs with specific emphasis on photosynthetic energy conversion, reductant channeling, and central carbon metabolism. Subsequent to that are the mechanisms of reductant generation and partitioning in cyanobacteria and how relevant pathways are affected by external environmental factors and macronutrient concentrations. We specifically explore the conditions and mechanisms whereby reducing equivalents are funneled to H_2 as well as the physiological and biochemical bases for light-driven proton reduction. Generally, H_2 evolution in cyanobacteria is catalyzed by hydrogenase and/or nitrogenase enzymes and is inhibited by the presence of O_2 or other competing electron acceptors such as CO_2 and N_2. In *Cyanothece* 51142, however, H_2 production occurs simultaneously with photosynthetic release of O_2 under continuous light conditions as a direct result of nitrogenase activity in the absence of exogenous nitrogen and carbon sources. We hypothesize that, under these conditions, sustained and uninterrupted H_2 production depends upon electrons derived from photocatalytic water splitting.

Cyanothece sp. ATCC 51142 is a unicellular cyanobacterium which can tolerate substantial amounts of O_2 during the operation of its dinitrogenase enzyme, despite the strong inhibitory effect of O_2 upon the catalytic mechanism of this enzyme. As the dinitrogenase enzyme has an auxiliary function to reduce protons to H_2 as well as its primary role in N_2 fixation, the organism is capable of producing H_2 gas in the presence of O_2. Nevertheless, *Cyanothece* 51142 is well known for its dynamic physiology which can temporally separate its dinitrogenase activity from the O_2-evolving process of oxygenic photosynthesis. As H_2 is an attractive candidate as a clean renewable fuel, it has been a scientific imperative to elucidate strategies whereby the catalysis of H_2 production can be driven efficiently by the energy in sunlight. Here we show, for the first time, the continuous production of H_2 by a unicellular photosynthetic organism with concomitant evolution of O_2.

Under continuous illumination with a feedback-controllable dual-wavelength light-enclosure, steady-state chemostat cultures of *Cyanothece* 51142 were maintained under NH_4^+-limitation in a custom photobioreactor. Despite the high expression of proteins and transcripts for the dinitrogenase system, no H_2 production could be detected during chemostat growth. Instead, the production of H_2 was shown to be easily induced by these cells upon interruption of the steady-state growth. When transferred from the NH_4-limited chemostat to sealed anaerobic tubes incubated under constant illumination, H_2 production was induced after an initial 12 h phase of O_2 accumulation. The dynamic pattern of alternating H_2 and O_2 evolution showed a periodicity of 24 h, reminiscent to previous descriptions of the temporal separation of N_2 fixation and photosynthesis. However, the level of each gas that accumulated did not significantly decrease between each production phase, and thus it became of interest to monitor the fine-scale dynamics of this process.

By halting the supply of NH_4^+ to the photobioreactor, chemostat growth was interrupted, and again H_2production activity was induced. However, facilitated by the high time-resolution instrumentation of the reactor, in-situ measurements enabled the striking observation that H_2 production can occur concomitantly with photosynthetic O_2 evolution in this unicellular organism. Within 1 h following the halt of NH_4^+ supply, H_2 production appeared with a rate which accelerated steadily over the next 24 h, reaching a maximal productivity of 400 μmol H_2/mg Chl/h. Within 6 h of the treatment, the rate of O_2 evolution had declined to a steady rate of 100 μmol O_2/mg Chl/h. However, following the peak of maximal H_2 production rates, the rate of net O_2 production rose, and the rates for both gases began to oscillate every 12 h, displaying an inverse correlation. However, throughout 100 h of continuous illumination, both H_2 and O_2 production activities maintained a positive net evolution rate, eventually dampening toward stable rates of 125 and 90 μmol/mg Chl/h, respectively. Additionally, dynamic changes in CO_2 evolution were observed, positively correlated with the rising and falling of the H_2 production rate.

By additionally halting the CO_2 supply during the chemostat interruption, *Cyanothece* 51142 induced a H_2 production activity which now displayed a *positive correlation* with changes in O_2 evolution rates, no longer following a 24 h periodicity. The light-dependence of the process was demonstrated by brief periods of darkness within the photobioreactor, leading to an immediate cessation of H_2 production, as well as by performance of in-situ light-saturation curves which showed dose dependence for both gases. Interestingly, the PS II-specific inhibitor DCMU was only capable of suppressing H_2 production in the long-term (hours as opposed to minutes). The yield of H_2 was partially diminished by the presence of increasing amounts of O_2, N_2, or CO_2, suggesting that other electron sinks may compete with the proton-reducing function of nitrogenase when they are available.

This research is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Biofuels Scientific Focus Area.

107

Genomic, Genetic, and Cytological Analysis Reveals Biased Inheritance of PatN, a Negative Regulator of Heterocyst Differentiation in *Nostoc punctiforme*

Douglas Risser,[1] Harry Christman,[1] Elsie Campbell,[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. We are applying "omic", genetic and cytological analyses to define the signal transduction pathway(s) leading to pattern heterocyst differentiation.

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 cyanobacteria of the family Nostocaceae, the heterocysts are present singly (apart from other heterocysts) in the filaments and separated by 10 to15 vegetative cells. 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. We are applying genetic, transcriptomic and proteomic analyses with wild-type and mutant strains to identify the regulatory circuits of free-living heterocyst differentiation and how those circuits have been co-opted during symbiotic growth.

We previously proposed a two stage—biased initiation, competitive resolution—model for the initiation of heterocyst differentiation (Meeks and Elhai, 2002). The biased initiation stage of the model holds that all cells in a filament sense the signal of combined nitrogen limitation (most likely elevated concentrations of 2-oxoglutarate), but only some

* Presenting author

cells can respond to the signal and initiate differentiation. The vegetative cells that can respond occur as a cluster of 2 to 4 cells. What condition specifically allows cells to respond is unknown. The second stage is resolution of the cluster of differentiating cells to a single cell that continues differentiation. Competitive resolution is the consequence of interactions between positive regulatory elements (HetR, HetF) and a negative regulatory element (PatS); overexpression of HetR or HetF, or deletion of PatS results in multiple contiguous heterocysts differentiating at a site in the filament.

We have isolated a mutant strain that displays a heterocyst frequency of about 30% in the free-living growth state. The heterocysts are present singly at relatively regular intervals along the filament, similar to the spacing pattern seen in the symbiotic growth state. This spacing pattern would be near optimal for function, with at least two photosynthetic vegetative cells supplying reductant as photosynthate to each heterocyst for subsequent hydrogen production from nitrogenase or a bidirectional hydrogenase.

The mutated gene has been designated *patN* and, in contrast to other genes involved in positive (e.g. *hetR*, *hetF*) or negative (*patS*) regulation of heterocyst differentiation, *patN* is found only in the genomes of the heterocyst-forming cyanobacteria sequenced to date. Transcription of *patN* is constitutive and PatN localizes to the cytoplasmic membrane in vegetative cells. In filaments supplemented with ammonium the localization is dynamic, with PatN undergoing a round of partitioning and differential inheritance with each cell division, so that the cell with the newest septum receives the majority of PatN. Deletion of the heterocyst patterning gene *patA* is epistatic to deletion of *patN*, and *patA* transcription increases in a *patN* mutant suggesting PatN may function by limiting the level of *patA*. Based on the unequal distribution of PatN among cells in a filament prior to nitrogen deprivation, and the increase in heterocysts observed upon deletion of *patN*, we propose that cells with low levels of PatN at the time of nitrogen starvation may be biased to initiate heterocyst differentiation. Understanding function of PatN is the first step to manipulating *N. punctiforme* to express the symbiotic growth phenotype apart from the plant partner in photo-production of biohydrogen.

Reference

1. Meeks, J. C. and J. Elhai. 2002. Regulation of cellular differentiation in filamentyous cyanobacteria in free-living and plant-associated symbiotic growth states. Microbiol. Mol. Biol. Rev. 66:94-121.

Supported by DOE Genomics:GTL program Grant No. DE-FG02-08ER64693.

108

Biohydrogenesis in the Thermotogales

J.S. Notey[1]* (jsnotey@ncsu.edu), A.D. Frock,[1] M. El-Sayed,[1] A. Petrus,[2] S. Swithers,[2] N. Boucher,[2] D. White,[3] R. Singh,[3] **P.H. Blum**,[3] **K.M. Noll**,[2] and **R.M. Kelly**[1]

[1]Department of Chemical 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

Project Goals: (see below)

The bacterial order Thermotogales consists of obligate fermentative anaerobes, most 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). The objectives of this project are:

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.**

***Thermotoga* genetics.** During the past year our lab has focused on advancing targeted integration technology for *Thermotoga maritima* cell line construction combined with studies on the metabolism of this organism as it pertains to redox homeostasis. An additional promoter, *Msed_1271p* (pilin), has been developed to drive expression of *T. maritima pyrE* that provides expanded versatility to the *T. maritima* genetic system. qRTPCR quantitation of promoter strength has been conducted based on the transcript abundances of *groES*, *Msed_1271*, *hisI*, *ldh*, *rpoD* and transgene *kan* providing insight into transgenic cell lines. While other genetic markers remain in play, an emphasis has been placed on the use of natural and synthetic alleles of *pyrE*, therefore multiple cycles of *pyrE* transformation and recombination have been conducted to paramatize the process. Advances in studies on *T. maritima* metabolism include genetic screens for redox homeostasis mutants and new information on metal reduction that provide insights into regulation of hydrogen formation.

Metal reduction by *T. maritima*. We have discovered that *T. maritima* actively forms magnetic iron (magnetite) using insoluble synthetic ferrihydroxides in a process that is driven by sugar fermentation. Metal reduction is a preferred electron sink that shifts excretion of reductant from synthesis of hydrogen gas to formation of ferrous iron. Negative staining transmission electron microscopy demonstrates synthesis of pili during this process. Genes implicated in this process have been targeted for inactivation to test predictive models.

Phylogenetic analyses revealed likely *ompA* and *ompB* homologs among Thermotogales species. Since OmpA1 and OmpB are the two dominant structural proteins of the *T. maritima* toga and all known species of the Thermotogales have togas, we examined the genome sequences of other members of the Thermotogales to identify possible homologs of these proteins. The closest OmpB pBLAST sequence outside the Thermotogales was from *Vibrio mimicus* with an E value of 8 x 10^{-6} indicating that currently there are no closely related sequences outside the Thermotogales in the GenBank database. OmpB orthologs are present in *Thermosipho melanesiensis*, *Fervidobacterium nodosum*, and most species of the *Thermotoga* genus. However no OmpB homologs were found in *Thermotoga lettingae*, *Thermosipho africanus*, *Kosmotoga olearia, Petrotoga mobilis*, and Thermotogales bacterium mesG.Ag.4. The *ompB* orthologs lie in syntenic regions containing four genes: *secG-tyrS* (tyrosyl-tRNA synthetase)-*ompA1-ompB*. The organisms that lack homologs to *ompB* also have this syntenic region, but have a gene downstream of their *ompA1* homolog that encodes a protein with no homologs in the GenBank nonredundant database (Fig. 3). The exception is that the *K. olearia* protein and the Thermotogales bacterium mesG.Ag.4 protein are homologous to one another, but to no other sequences. All these non-homologous protein sequences were examined for porin characteristics and all had many porin characteristics, except that they do not appear to be globular proteins. Though these proteins are not recognizable homologs of *T. maritima* OmpB, they may be analogs serving the same function.

Mesophilic member of the Thermotogales. A novel mesophilic member of the Thermotogales designated MesG1.Ag.4.2 was isolated from sediments from Baltimore Harbor, Maryland, USA. Cells of strain MesG1.Ag.4.2 were non-motile ovoids with sheath-like structures (togas). The strain was Gram-negative and grew optimally at 37ºC with a doubling time of 16.5 h (specific growth rate = 0.06 h^{-1}) on xylose at 37°C. It grew best on carbohydrates, but proteinaceous compounds also supported growth. This is the first describes species of the mesotoga lineage and provides new possibilities for the development of genetic methods for the Thermotogales.

Thermotoga species differ with respect to sugar utilization. Four hyperthermophilic members of the bacterial genus *Thermotoga* (*T. maritima, T. neapolitana, T. petrophila,* and *T.* sp. RQ2) share a core genome of 1470 open reading frames (ORFs), or about 75% of their genomes. Nonetheless, each species exhibited certain distinguishing features during growth on simple and complex carbohydrates that correlated with genomic inventories of specific ABC sugar transporters and glycoside hydrolases. These differences were consistent with transcriptomic analysis based on a multi-species cDNA microarray. Growth on a mixture of six pentoses and hexoses showed no significant utilization of galactose or mannose for any of the four species. *T. maritima* and *T. neapolitana* exhibited similar monosaccharide utilization profiles, with a strong preference for glucose and xylose over fructose and arabinose. *T.* sp. RQ2 also used glucose and xylose, but was the only species to utilize fructose to any extent, consistent with a phosphotransferase system (PTS) specific for this sugar encoded in its genome. *T. petrophila* used glucose to a significantly lesser extent than the other species. In fact, the XylR regulon was triggered by growth on glucose for *T. petrophila*, which was attributed to the absence of a glucose transporter (XylE2F2K2), otherwise present in the other Thermotoga species. This suggested that *T. petrophila* acquires glucose through the XylE1F1K1 transporter, which primarily serves to transport xylose in the other three Thermotoga species. The results here show that subtle differences exist among the hyperthermophilic Thermotogales that support their designation as separate species.

TM1300 locus in the *T. maritima* genome. The TM1300-1338 locus in the *T. maritima* genome contains a number of hypothetical proteins, including several that correspond to ORFs of less than 100 aa. Analysis of this locus has revealed that several putative bacteriocins and toxin-antitoxins (HicAB) are encoded, suggesting a key role in post-transcriptional regulation and possible anti-microbial actions. The gene encoding TM1312 has been expressed and confirmed to be a toxin (ribonucleases). Purification of TM1316, a putative cyclic peptide and suspected bacteriocin, is underway. Also, gel shift assays have been used to investigate two transcriptional regulators in this locus. For example, TM1330 appears to be involved in regulation of the region upstream of TM1316.

Publications in the past year citing DOE GTL Support

1. Petrus, A. K., K. S. Swithers, C. Ranjit, S. Wu, H. M. Brewer, J. P. Gogarten, L. Pasa-Tolic, and K. M. Noll. Genes for the major structural components of Thermotogales species' togas revealed by proteomic and evolutionary analyses of OmpA and OmpB homologs. PloS One (submitted)
2. Nesbø, C. L., D. M. Bradnan, M. Dlutek, A. K. Petrus, J. Foght, W. F. Doolittle, and K. M. Noll. *Mesotoga prima* gen. nov., sp. nov., the first described mesophilic species of the Thermotogales. Extremophiles (submitted)
3. Swithers, K. S., et al. Genome sequence of *Kosmotoga olearia* strain TBF 19.5.1., a thermophilic bacterium with a wide growth temperature range, isolated from the Troll B oil platform in the North Sea. J Bacteriol. 193: 5566-5667. 2011.
4. Swithers, K. S., et al. Genome sequence of *Thermotoga* sp. strain RQ2, a hyperthermophilic bacterium isolated from a geothermally heated region of the seafloor near Ribeira Quente, the Azores. J Bacteriol. 193:5869-5870. 2011.
5. Boucher, N., K. M. Noll. Ligands of thermophilic ABC transporters encoded in a newly sequenced genomic region of *Thermotoga maritima* MSB8 screened by differential scanning fluorimetry. Appl. Environ. Microbiol. 77:6395-6399. 2011.
6. Frock, A.D., S.A. Gray and R.M. Kelly. Hyperthermophilic Thermotoga species differ with respect to specific carbohy-

drate transporters and glycosidase hydrolases. Appl Environ Microbiol (accepted pending minor revisions)

109

Mutational Analysis of the Enzymes Involved in the Metabolism of Hydrogen by *Pyrococcus furiosus*

Gerrit J. Schut[1]* (gerti@uga.edu), William J. Nixon,[1] Gina L. Lipscomb,[1] **Michael W.W. Adams**,[1] and **Robert A. Scott**[1,2]

Departments of [1]Biochemistry and Molecular Biology and [2]Chemistry, University of Georgia, Athens

Project Goals: The overall goal is to incorporate a systems-level approach to improve our understanding of microbial regulatory and metabolic networks related to hydrogen production through the development of new model organisms for microbial hydrogen production. We are using targeted approaches for the identification and characterization of enzymes and biochemical pathways relevant to biological hydrogen production. Advancing fundamental scientific knowledge in these areas are critical to characterizing enzymatic structure and function, modeling regulatory and metabolic networks, and engineering of enzymes and organisms to improve biological hydrogen production for a future hydrogen economy.

The hyperthermophilic archaeon *Pyrococcus furiosus* grows optimally near 100°C by fermenting carbohydrates to acetate, carbon dioxide and hydrogen gas (H_2). In the presence of elemental sulfur (S^0) it produces hydrogen sulfide rather than H_2. *P. furiosus* contains three distinct hydrogenases, which are enzymes that metabolize H_2. Two are cytosolic and are termed SHI and SHII. Both use NADP(H) as an electron carrier and each is encoded by a 4-gene operon. The third hydrogenase, MBH, is membrane-bound and uses the redox protein ferredoxin as an electron carrier. It is encoded by a 14-gene operon. In a previous study (1) we constructed deletion strains lacking the operons encoding SHI or SHII or both and showed that they exhibited no obvious phenotype under the usual growth condition (at 98°C using maltose as the carbon source). This study has now been extended to include biochemical analyses and growth studies under a variety of conditions using the ΔSHI and ΔSHII deletion strains together with strains lacking a functional MBH or more specifically, lacking the catalytic subunit (MbhL) of MBH. While deletion of either one or both cytosolic hydrogenases does not produce any obvious growth phenotype, the levels of hydrogenase activity in the cytoplasmic extracts are not affected in the ΔSHII strain but are strongly reduced in the ΔSHI strain (<10% compared to the parental strains). In the SHI and SHII double deletion strain, the cytosolic hydrogenase activity was below the detection limit. These data indicate that SHI is responsible for most of the hydrogenase activity in the cytoplasm. In contrast, the strain lacking the membrane-bound hydrogenase catalytic subunit showed no growth in the absence of S^0. This confirms the hypothesis that, in the absence of S^0, MBH is the enzyme that produces H_2. Moreover, these results show that in the absence of S^0 *P. furiosus* can only dispose of reductant generated from sugar oxidation in the form of H_2 gas. The deletion strain devoid of all three hydrogenases (ΔSHI ΔSHII Δ*mbhL*) grows only in the presence of S^0 and did not produce any detectable H_2. When the hydrogenase deletion strains and their parental strains were grown in the presence of limiting S^0 (0.5 g/L vs 2 g/L), both S^0 is reduced and H_2 is produced in the parental strains (ca. 50% H_2 produced compared to no added S^0). Interestingly, in Δ*mbhL*, a significant amount of H_2 is produced (ca. 20% compared to the parental strains) and growth was compromised (ca. 50% lower cell yield as compared to the parental strains). Therefore, the H_2 produced in Δ*mbhL* must be catalyzed by SHI, showing that this cytosolic 'uptake' hydrogenase can also produce H_2 from NADPH *in vivo*, even though this reaction is thermodynamically unfavorable. However, SHI cannot compensate for the absence of MBH since it does not enable significant growth of the Δ*mbhL* strain. We propose that the *in vivo* function of SHI is to recycle H_2 and provide a link between external H_2 and the intracellular pool of NADPH needed for biosynthesis. *P. furiosus* only uses a low potential ferredoxin in its glycolytic pathway, which is linked to MBH for the disposal of all reducing equivalents as H_2, generating an ion gradient for ATP generation in the process. The ability to recycle H_2 might have a distinct energetic advantage in the environment, but it is clearly not required for heterotrophic growth of the organism under the usual laboratory conditions.

Reference

1. Lipscomb, G. L., Stirrett, K., Schut, G. J., Yang, F., Jenney, F. E, Jr., Scott, R. A., Adams, M. W. W. and Westpheling, J. (2011) Natural competence in the hyperthermophilic archaeon *Pyrococcus furiosus* facilitates genetic manipulation: use for constructing multiple markerless deletions of genes encoding two cytoplasmic hydrogenases. *Appl. Environ. Microbiol.* 77, 2232-2238

This work was supported by grants from the Office of Biological and Environmental Research (DE-FG02-08ER64690) and the Chemical Sciences, Geosciences and Biosciences Division (DE-FG05-95ER20175), Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy.

110

Pathway of Fermentative H_2 Production by Sulfate-Reducing Bacteria

Kimberly L. Keller[1]* (kellerkl@missouri.edu), Steve D. Brown,[2,3] Barbara J. Rapp-Giles,[1] Iris Porat,[3] Adam Deutschbauer,[4,3] Jennifer Kuehl,[4,3] Timothy J. Tschaplinski,[3] and **Judy D. Wall**[1,3]

[1]University of Missouri, Columbia; [2]Oak Ridge National Laboratory, 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 abundant soil anaerobes, the sulfate-reducing bacteria (SRB) of the genus *Desulfovibrio*, produce H_2 from organic acids. These apparently simple pathways have yet to be clearly established and their elucidation forms the goals for our research efforts. Information obtained may facilitate the exploitation of other anaerobes not yet readily amenable for examination by molecular tools and will be useful in consideration of practical applications. Energy metabolism, flux through the pathways, and regulation are likely to be limiting factors in channeling reductant for H_2 production, the process which we are examining. Our research has focused on: (1) the determination of the path of electrons during pyruvate fermentation growth in order to identify possible electron sinks, (2) confirmation of the enzymes that are essential during pyruvate fermentation, and (3) determination of the role of the pyruvate enzymes during lactate respiration.

Strains of *Desulfovibrio* can ferment organic acids in the absence of additional terminal electron acceptors and produce rather large amounts of H_2. We believe a study of the limitations to H_2 production in our model organism, *Desulfovibrio alaskensis* G20 (Hauser et al, 2011) may be informative to decipher the flow of electrons in those organisms chosen for industrial application for H_2 production.

We proposed to tease apart the contribution of fermentation to the respiratory energy budget to determine the dependence of the bacterium on this process. In studying growth, we have realized an important role for the utilization of fumarate via the fumarate reductase complex (Dde_1258 – Dde_1256) during various growth modes. *D. alaskensis* G20 is able to grow robustly on fumarate through disproportionation after adaptation when subcultured from cells grown on lactate/sulfate. However, growth on fumarate is inhibited by formate, lactate, pyruvate, H_2, or CO_2 indicating that the use of fumarate by the cells may be a secondary or tertiary option as an electron acceptor. Interestingly, proteomic analysis reveals that even during growth on lactate/sulfate or by pyruvate fermentation, the proteins of the fumarate reductase complex are expressed at relatively abundant levels. Yet, a very small proportion of reductant appears as succinate in these culture conditions.

Mutant analysis of the genesencoding the fumarate reductase complex, the quinone-reducing complex, and the type-1 tetraheme cytochrome c_3 indicate possible components playing a role in the flow of electrons, as all are unable to grow on fumarate. In addition, several mutants lacking formate dehydrogenase isozymes have an impaired growth phenotype on fumarate. Microarray, proteomic, and metabolite analysis is currently underway to determine changes that may be critical to robust growth of *D. alaskensis* G20 by disproportionation of this substrate and in the channeling of reductant for H_2 production.

References and Publications

1. Hauser, L. J., M. L. Land, S. D. Brown, F. Larimer, K. L. Keller, B. J. Rapp-Giles, M. N. Price, M. Lin, D. C. Bruce, J. C. Detter, R. Tapia, C. S. Han, L. A. Goodwin, J-F. Cheng, S. Pitluck, A. Copeland, S. Lucas, M. Nolan, A. L. Lapidus, A. V. Palumbo, and J. D. Wall. (2011) The Complete Genome Sequence and Updated Annotation of *Desulfovibrio alaskensis* G20. *J Bacteriol.* 191:4268–4269.
2. Keller, K. L., and J. D. Wall. (2011) Genetics and molecular biology of the electron flow for sulfate respiration in *Desulfovibrio*. *Front Microbiol.* 2:1-15.

This work was supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research, Genomics Program:GTL BioH2 Production and BioEthanol contract DE-FG02-083464691; Office of Science, Office of Biological and Environmental Research, Genomics Program:GTL. Oak Ridge National Laboratory is managed by the University of Tennessee-Battelle LLC for the Department of Energy under contract DEAC05-00OR22725; Office of Basic Energy Sciences, contract FG0287ER13713; and ENIGMA, Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under contract DE-AC02-05CH11231. Proteomic data was provided through an EMSL grant, sponsored by the Department of Energy's Office of Biological and Environmental Research located at Pacific Northwest National Laboratory.

111

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

Ryan Severins,[1] Jonathan Millen,[1] Katherine Chou,[2] Ji Hye Jo,[2] Yuan Zhang,[1] Wendy D.M. Wang,[1] Robert J. Zagursky,[1] Pin-Ching Maness,[2] Yi Qu,[3] Mary Lipton,[3] and **J.H. David Wu**[1]* (davidwu@che.rochester.edu)

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

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.

Due to the novelty of CtHydA3, it was selected for initial study. Genes encoding CtHydA3 (Ct_3003), a ferredoxin-like protein (Ct_3004), and three FeFe-hydrogenase maturation proteins (CtHydE, CtHydF, and CtHydG) have been cloned into three plasmids and co-transformed into *E. coli* strain Rosetta (DE3) and BL21 (DE3) for heterologous expression. In addition, a 6X His-tag sequence was fused to either the C- or N-terminus of CtHydA3. Protein immunoblots confirmed the expression of both the C- and N-terminus His-tagged CtHydA3 (73 kDa band) in *E. coli* Rosetta, but not in *E. coli* BL21, likely due to differences in codon usage between *C. thermocellum* and *E. coli*. The recombinant protein was affinity-purified with *in vitro* activity linking to reduced methyl viologen. However, the low yield of the purified protein prevented its biochemical characterization. The *E. coli* Rosetta strain also displayed a high background hydrogenase activity. We have since initiated transformation using an *E. coli* FTD strain (DE3) lacking all background hydrogenase activity with the intent to determine cofactors requirements and directionality (H_2 production or H_2 uptake) of CtHydA3 in *E. coli* cell-free extract. The outcome will reveal the physiological functions of CtHydA3 during cellulose metabolism.

To identify transcription factors controlling 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, obtained by growing on cellobiose or crystalline cellulose and eluted from the affinity columns, were identified by the MALDI-TOF or LC-MS-MS techniques. We thus identified a REX-like protein that regulates the expression of a hydrogenase gene and many other genes, indicating that it is a global regulator. It is likely the first global regulator experimentally verified in this bacterium. We also carried out the proteomic analysis and identified the cellulosomal proteins and non-cellulosomal glycosyl hydrolases that are up- or down-regulated when the cells were grown on the cellulose substrate vs. cellobiose.

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.

262A

submitted post-press

The Ins and Outs of Algal Metal Transport

Crysten E. Blaby-Haas and **Sabeeha S. Merchant*** (merchant@chem.ucla.edu)

Department of Chemistry and Biochemistry, University of California, Los Angeles

http://www.chem.ucla.edu/dept/Faculty/merchant.html/

Project Goals: (see abstract)

Metal transporters are a central component in the interaction of algae with their environment. They represent the first line of defense to cellular perturbations in metal concentration, and by analyzing algal metal transporter repertoires, we can get insight into a fundamental aspect of algal biology. The ability of individual algae to thrive in environments with unique geochemistry, compared to non-algal species commonly used as reference organisms for metal homeostasis, provides an opportunity to broaden our understanding of biological metal requirements, preferences and trafficking. *Chlamydomonas reinhardtii* is the best developed reference organism for the study of algal biology, especially with respect to metal metabolism; however,the diversity of algal niches necessitates a comparative genomic analysis of all sequenced algal genomes. A comparison between known and putative proteins in animals, plants, fungi and algae using protein similarity networks has revealed the presence of novel metal metabolism components in Chlamydomonas including new iron and copper transporters. This analysis also supports the concept that, in terms of metal metabolism, algae from similar niches are more related to one another than to algae from the same phylogenetic clade.

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

112

Approaches to Systematically Examine Carbon Flux in Microbial Communities Using 'Omics' and Stable Isotope Probing

Annika Mosier,[1] Nicholas Justice,[1] Daniela Aliaga Goltsman,[1] Christopher Miller,[1] Chongle Pan,[2] Ben Bowen,[3] Brian Thomas,[1] Steven Singer,[3] Trent Northen,[3] Robert Hettich,[2] and **Jill Banfield**[1]* (jbanfield@berkeley.edu)

[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.

Earth's climate is regulated by biogeochemical carbon exchanges between the land, oceans and atmosphere that are chiefly driven by microorganisms, which produce and consume carbon dioxide (CO_2), methane, and organic matter and support the growth of higher organisms. Microbial communities are therefore indispensible to the study of carbon cycling and its impacts on the global climate system. However, approaches to examine carbon flux in communities systematically at the molecular level are inadequate. We are developing methods to track the flow of carbon into and through microbial communities using a well-characterized model system—acid mine drainage (AMD) biofilms.

We are using stable isotope probing (SIP)-proteomics and SIP-metabolomics to measure and characterize the incorporation of carbon compounds into proteins and metabolites from specific organisms. SIP-proteomics and SIP-metabolomics can determine ^{13}C or ^{15}N atom% of thousands of proteins and hundreds of metabolic features in the AMD community. Together with comprehensive community metagenomics, these novel approaches demonstrate which compounds are being consumed and which organisms are consuming them. Protein data from individual organisms, community metabolite profiles and measurements of carbon pools and fluxes will allow us to generate a model for carbon cycling in the community.

We have developed and validated a ^{15}N-based SIP-proteomic method using laboratory-grown AMD communities (Pan et al., 2011) and that has since been applied to further experiments (see below). We are now using ^{13}C-based SIP-proteomics to track the incorporation of $^{13}CO_2$ into microbial proteins. Newly developed methods for SIP-metabolomics with ^{13}C and ^{15}N constrain the biological origin and specific formulas of the metabolites generated by the AMD community, making metabolite identification more tractable. Together, these approaches will enable us to follow the movement of carbon through the community as it shifts from primary production in early growth stages to a mixture of heterotrophic and autotrophic metabolisms in the later stages of development.

The initial colonizer of the AMD biofilms, *Leptospirillum rubarum*, dominates early community development and carbon fixation in the system. *Leptospirillum ferrodiazotrophum* and other lower abundance community members appear later as biofilms develop and diversify. Using community genomic data collected over a period of 7 years, we assembled the partial genome of a new bacterial species that expands the *Leptospirillum* clade: *Leptospirillum* Group IV. The new species shares 97% 16S rRNA sequence identity and 70% identity between shared proteins with its closest relative, *Leptospirillum ferrodiazotrophum*. The presence of nitrogen fixation and reverse TCA cycle proteins suggest an autotrophic metabolism similar to that of *Leptospirillum ferrodiazotrophum*, while hydrogenase proteins unique to *Leptospirillum* Group IV suggest an active role in the anaerobic setting.

Fungi often colonize late-developmental-stage biofilms at low relative abundance and may play an important role in recycling carbon in the community. We are employing genomics, transcriptomics, and proteomics to link functional activities encoded and expressed by fungi with biogeochemical processes within the ecosystem (Miller et al., in prep). We reconstructed the near-complete composite genome (27 Mbp) of the dominant fungal AMD community member, *Acidomyces richmondensis*. Approximately 900 unique fungal proteins were detected by proteomics in field samples, covering many important metabolic pathways. Genes involved in heterotrophic carbon cycling, including several glycosyl hydrolases involved in polysaccharide hydrolysis, were expressed at varying levels based on transcript abundances inferred from ESTs recovered from field samples.

We have shown that Archaea from the Thermoplasmatales lineage dominate submerged anaerobic biofilms in the AMD system and play an important role in nutrient cycling in the anaerobic and degradative portion of the carbon cycle (Justice et al., in prep). Members of the novel ARMAN nanoarchaeal lineage are especially abundant in the sunken biofilms, comprising at least 10% of the community based on FISH and metagenomic data analysis. Comparative community proteomic analyses show a persistence of bacterial proteins in sunken biofilms, but there is clear evidence of deamidation caused by acid exposure. Given the low representation of bacterial cells in sunken biofilms, we infer that deamidated proteins are derived from populations of lysed cells. Culture experiments demonstrated heterotrophic growth of A-plasma and *Ferroplasma* sp. on peptone, betaine, casein-derived peptides, and natural biofilm material, all coupled to iron reduction. The results demonstrate anaerobic archaeal growth via protein degradation, and possibly other organic carbon compounds. ^{15}N-based SIP-proteomics showed that archaea are the dominant species incorporating ^{15}N in sunken biofilms. The SIP-proteomic techniques also allowed us to profile the archaeal metabolic activities preferentially enriched for newly synthesized protein. These findings expand our understanding of the roles of Archaea in anaerobic nutrient cycling.

We are determining how a key parameter of global climate change, elevated temperature, regulates the flow of carbon through the microbial-based AMD ecosystem. Proteomic and metabolomic data highlight which community members and which modes of carbon cycling are affected by elevated temperature. This approach is establishing whether the carbon cycling pathways in the system are robust to the effects of climate change or, if not, where the cycling may break down.

113

Estimation of Hydrogen Isotope Enrichment in Proteins and Lipids in a Microbial Community

Ben Bowen[2]* (BPBowen@lbl.gov), Curt Fischer,[1] Chongle Pan,[3] Bob Hettich,[3] Trent Northen,[2] and **Jillian F. Banfield**[1,2]

[1]University of California, Berkeley; [2]Lawrence Berkeley National Laboratory, Berkeley, Calif.; and [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.

In the biosynthesis of lipids and amino acids, microorganisms may preferentially use one of the hydrogen stable isotopes (^{2}H and ^{1}H or D and H). This results in different fractions of hydrogen stable isotopes in synthesized biomass versus water in the growth medium. Isotope preference partially stems from the physiology of the microorganisms. Autotrophic microorganisms, for example, preferentially incorporate ^{1}H during lipid biosynthesis, possibly resulting from reactions involving NAD(P)H that forms isotopically light acetyl-CoA. Stable isotope fractionation is traditionally measured using isotope ratio mass spectrometry. However, it is challenging to measure and identify unknown molecules in a complex sample using this technique. Here we present a new approach for measuring hydrogen fractionation in complex samples of proteins and lipids from an acidophilic microbial community in acid mine drainage.

The analyses targeted coexisting bacteria and archaea grown in a 4 atom% ^{2}H-enriched medium. To measure average hydrogen fractionation in amino acids, the community proteome was digested using trypsin and measured using 2-dimension liquid-chromatography coupled with high-resolution tandem mass spectrometry analysis on an Orbitrap Velos. The Sipros algorithm (code.google.com/p/sipro) was used to search each MS^2 spectrum against all peptide sequences in the AMD protein database at every ^{2}H atom% level in 1% precision. The ^{2}H enrichment and sequence of 4,477 unique peptides from 36,090 spectra were identified. Further, to more precisely quantify isotopic enrichment, a MATLAB program was developed to find the optimum ^{2}H values which best fit the isotopic distributions of identified peptides in MS and MS^2 spectra. The results indicated that the microbial community has an average ^{2}H depletion of -360 ‰ relative to water in the growth medium for proteins. We applied a similar technique to the analysis of lipids identified using LCMS, and the lipids were found to be isotopically light by -260 ‰ relative to water in the growth medium.

This difference between -260 ‰ ^{2}H fractionation in lipids and -360 ‰ in proteins suggests that additional factors, that do not involve NAD(P)H, are also contributing to ^{2}H/^{1}H fractionation. Further, autotrophic bacteria sharing 94% 16S rRNA gene sequence identity displayed statistically significant differences in protein hydrogen isotope fractionation. This suggests these microbes have different metabolic traits determined by their distinct ecological niches. In addition, it was found that heterotrophic, archaeal members of the community had isotopically light protein and were significantly different from coexisting bacteria. Potentially, this could be attributed to metabolite transfer from autotrophs and unknown aspects of fractionation associated with iron reduction. Differential fractionation of hydrogen stable isotopes into metabolites and proteins may reveal trophic levels of members of microbial communities. The approach developed here provided insights into the metabolic characteristics of organisms in natural communities, and may be applied to analyze other systems.

This work was supported by Genome Sciences Program in Carbon Cycling (contract number DE- SC0004665).

114

Microbial Food Web Mapping: Linking Carbon Cycling and Community Structure in Soils Through Pyrosequencing-Enabled Stable Isotope Probing

Nathalie J.A. Curlevski,* Ashley N. Campbell, and **Daniel H. Buckley** (dbuckley@cornell.edu)

Cornell University, Ithaca, N.Y.

Project Goals: This research program explores fundamental aspects of carbon cycling in soil microbial communities. Major goals are to develop and validate methods for pyrosequencing-enabled stable isotope probing (SIP) and to use this approach to dissect the microbial food web in soil. 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. With this approach we will examine connections between microbial community composition and soil carbon cycle dynamics. Specific objectives include 1) determine whether carbon input parameters (composition, quantity, timing of carbon additions) alter the route of carbon through the soil community, 2) determine whether these shifts interact with respect to microbial community structure, and 3) evaluate whether microbial community structure is functionally equivalent across edaphically similar soils that differ in management history.

The terrestrial biosphere contains a large fraction of global C and nearly 70% of the organic C in these systems is found in soils. Much of the organic C in soils is respired and on an annual basis soil respiration produces 10 times more CO_2 than anthropogenic emissions, but it remains difficult to predict the response of soil processes to anthropogenic changes in the environment. Our difficulty in predicting how soil processes will respond to environmental change suggests a need for a greater understanding of the biotic mechanisms that govern the soil C-cycle. It is important to examine the internal dynamics of soil microbial communities, and the manner in which they influence community function, in order to understand the how the terrestrial C-cycle responds to environmental change.

While strides have been made 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. The pyrosequencing enabled stable isotope probing approach that we are developing will allow for pulse chase style experiments that allow ^{13}C-isotopes to be tracked through the soil community over time. The approach will involve the application of synthetic biomass containing a mix of carbon sources designed to approximate the plant biomass. The use of synthetic biomass allows substitution of ^{13}C-labeled substrates into the mixture to track the manner in which different types of C (ie: polymers and sugars) are metabolized by different components of the community.

Initial experiments have explored 1) what are the dynamics of degradation for ^{13}C-cellulose plant simulant and how does the ^{13}C move through community nucleic acids over prolonged incubation, 2) how does the community respond to the addition of ^{13}C-cellulose plant simulant relative to the addition of only ^{13}C-cellulose, and 3) how does the ^{13}C assimilation into microbial nucleic acids from ^{13}C-xylose plant simulant vary from that of ^{13}C-cellulose plant simulant when all aspects of the experiment are identical except for the isotopic label. $^{12}CO_2$ and $^{13}CO_2$ generated from microbial respiration is determined over a month long incubation and soils are sampled destructively over time. Samples from different times are subject to DNA and RNA stable isotope probing and 454 pyrosequencing of gradient fractions. These data are used to determine the buoyant density profile for individual OTUs in ^{13}C treatments relative to unlabeled controls and to observe how the degree of ^{13}C-label incorporation by individual OTUs changes over time.

Future experiments will explore how food web dynamics differ between edaphically similar soils that differ in land management practice. To prepare for these experiments we have identified a series of suitable sites that represent a gradient of management impact resulting in accumulation of soil organic matter. The sites include an intensively managed corn field, fields that were removed from intensive management for 10 or 20 years and subsequently managed for organic grain crops, and long term pasture. Initial characterization of the soil bacterial and fungal communities by 454 pyrosequencing is currently underway in order to determine the community composition in these sites and develop protocols for pyrosequencing enabled ^{13}C-SIP. Using these data, OTUs identified by ^{13}C-SIP can then me mapped back to the landscape in space and time to evaluate their distribution and importance in native soil systems.

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.

115

From Community Structure to Function: Metagenomics-Enabled Predictive Understanding of Microbial Communities to Climate Warming at the Temperate Grassland Ecosystems in Oklahoma

Jizhong Zhou[1]* (jzhou@ou.edu), Liyou Wu,[1] Kai Xue,[1] Lei Cheng,[1] Mengting Yuan,[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] Xia Xu,[1] Dejun Li,[1] 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: Determining the response, adaptation and feedback mechanism of biological communities to climate change is critical to project future states of the earth and climate systems, but poorly understood in microbial com-

munities. 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 response of microbial community structure, function and activity 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 response to climate change. We have carried out our studies at two contrasting long-term experimental facilities, the temperate grassland ecosystems in OK (this poster) and tundra ecosystems in Alaska (Poster led by Schuur et al).

Feedback responses of microbial communities to climate warming. We have used integrated metagenomic technologies to analyze the responses of microbial communities in a long-term (10 years) experimental warming grassland ecosystem in Oklahoma. Our results showed that microorganisms play crucial roles in regulating soil carbon (C) dynamics through three primary feedback mechanisms: (i) shifting microbial community composition, which most likely led to the reduced temperature sensitivity of heterotrophic soil respiration, (ii) differentially stimulating genes for degrading labile but not recalcitrant C so as to maintain long term soil C stability and storage, and (iii) enhancing nutrient cycling processes to promote plant nutrient use efficiency and hence plant growth. Elucidating microbially mediated feedbacks is fundamental to understanding ecosystem responses to climate warming and provides a mechanistic basis for C-climate modeling.

Responses of microbial communities to clipping within the context of climate warming. We have also examined the responses of microbial communities to clipping within the context of climate warming using the GeoChip-based metagenomics technology. Various statistical analyses showed that clipping had significant impacts on microbial community structure, and altered the directions of the warming effects. The results clearly indicated that clipping substantially mediated warming effect on soil microbial community. The stimulated genes in relative more recalcitrant C decomposition under warming with clipping may have important implications for the stability of soil C storage. The diminished warming effect by clipping for N genes may affect plant production and further influence soil C cycling.

Metagenomic sequencing analysis of microbial communities under long-term warming. We employed the Illumina Hi-Seq2000 platform (2 X 100 paired end) to obtain shotgun metagenomes from 12 replicate samples, with 6 each for warming and control. Each sample yielded 10-15 Gb of sequence data with >60 Gb per community. Using an assembly-free bioinformatic pipeline, our results showed that more than 90% of the genes and organisms did not differ in abundance between the two communities while a higher abundance of genes related to sporulation was observed under warming. Significant differences were observed among the top four most abundant phyla (Proteobacteria, Acidobacteria, Planctomycetes, and Bacteriodetes) between warming and control. Interestingly, the microbial populations of these phyla from the warmed samples showed significant increase in G+C% content. Furthermore, several metabolic pathways were significantly increased under warming, including pathways directly related to the emission of greenhouse gases (e.g., CH_4, NO, and CO_2), nitrogen cycling (e.g., fixation, nitrification, nitrate/nitrite reduction), and organic carbon utilization (e.g., mixed acid fermentation, mannitol utilization).

Belowground net primary productivity under warming and clipping. The dynamics of belowground net primary productivity (BNPP) is of fundamental importance in understanding carbon (C) allocation and storage in grasslands. Thus we have examined the changes of BNPP under warming and clipping. Warming increased BNPP by 42-67% with a significant increase observed in wet years. Clipping also had significant positive impacts on BNPP. Overall, f_{BNPP}, the fraction of BNPP to NPP, increased under both warming and clipping treatments, more in dry years. Water availability (either precipitation or soil moisture) was the most limiting factor for both BNPP and f_{BNPP}. It strongly dominated the interannual variability of NPP and f_{BNPP} and their responses to warming and clipping. Our results suggested that water availability might regulate tallgrass prairie's responses to warming and land use change, which may eventually influence the global C cycling.

The stability of organic carbon in deep soils. The majority of C in deep soils is recalcitrant and old with turnover times of hundreds to thousands of years, whose fate is critical to project future climate warming. Therefore, we determined the warming effect on old, recalcitrant organic C decomposition by combining long-term (9 yr) in-situ field and short-term (9 wk) laboratory incubations. Our results showed that warming of approximately 2°C significantly facilitated the loss of C with several thousand years old in the deep soil layer. Coupled stable isotope probing and meta-genomic analysis indicated that warming-induced old C decomposition was closely related to changes in the functional structure of microbial communities. Our findings suggest that warming may significantly reduce the size of the vast pool of old C in global soils and thus reinforce the positive feedback between the C cycle and climate.

Development of amplicon-sequencing approaches for uncovering functional gene diversity. Eco-functional genes involved in nitrogen and carbon cycling were targeted for pyrosequencing in order to resolve changes in microbial functional community structure underlying process changes. For nitrogen cycling, the functional genes nifH (N fixation), nirK (denitrification), amoA (nitrification), and Archaeal-amoA (nitrification) were initially targeted. Bacterial laccases, which catalyze the oxidation of various substituted phenolic groups were also sequenced. The FunGene Pipeline and Repository has been enhanced to provide the full line of capability of processing/analyzing sequences of these eco-functional genes. In addition, primers are currently being developed for the nitrogen cycle (nosZ and nirS) and the carbon cycle (pmoA, ligE, and fungal peroxidase).

116

From Community Structure to Functions: Metagenomics-Enabled Predictive Understanding of Microbial Communities to Climate Warming in Alaskan Tundra

Edward A.G. Schuur[1]* (tschuur@ufl.edu), **Jizhong Zhou**,[2] Liyou Wu,[2] Kai Xue,[2] Lei Cheng,[2] Mengting Yuan,[2] Jin Zhang,[2] Ye Deng,[2] Joy D. Van Nostrand,[2] **Zhili He**,[2] Ryan Penton,[3] Jim Cole,[3] **James M. Tiedje**,[3] Rosvel Bracho-Garrillo,[1] Chengwei Luo,[4] **Konstantinos Konstantinidis**,[4] Xia Xu,[2] Dejun Li,[2] and **Yiqi Luo**[2]

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

Determining the responses, adaptations and feedback mechanisms of biological communities to climate change is critical to project future states of the earth and climate systems, but poorly understood for microbial communities. We have carried out our studies at two contrasting long-term experimental facilities, a tundra ecosystem in Alaska (this poster) and a grassland ecosystem in Oklahoma (Poster led by Zhou et al).

Effects of experimental and natural warming of deep soil and permafrost on ecosystem carbon balance in Alaskan tundra. Approximately 1670 Pg (billion tons) of soil carbon are stored in the northern circumpolar permafrost zone, more than twice as much carbon than currently contained in the atmosphere. Permafrost thaw, and the microbial decomposition of previously frozen organic carbon, is considered one of the most likely positive feedbacks from terrestrial ecosystems to the atmosphere in a warmer world. Here we report results from two experimental systems that examine this question: **1)** a natural gradient of permafrost thaw where *minimal*, *moderate*, and *extensive* permafrost degradation has been observed over a decadal time scale, and **2)** a new ecosystem warming manipulation—the Carbon in Permafrost Experimental Heating Research (CiPEHR) project—where we increased air and soil temperature, and degraded the surface permafrost. Within these experiments, net ecosystem C exchange and the radiocarbon age of ecosystem respiration were measured to determine the influence of old C loss on ecosystem C balance. By partitioning respiration sources across the thaw gradient, we determined that areas that thawed over the past 15 years had 75% more annual losses of old C compared to minimally thawed areas, but had overall net ecosystem C uptake as increased plant growth offset these losses. In contrast, sites that thawed decades earlier lost an additional 25% more old C annually, which contributed to overall net ecosystem C release despite increased plant growth. These findings were mirrored by the warming experiment where increased plant uptake appears to compensate for microbial release of carbon, at least in the three years of warming that we have observed. Together, these data document significant losses of soil C with permafrost thaw that, over decadal time scales, overwhelms increased plant C uptake at rates that could make permafrost a large biospheric C source in a warmer world, similar in magnitude to current C fluxes from land use change.

Patterns of permafrost thaw influence on tundra microbial communities. Microbial communities from the permafrost thaw gradient were analyzed by Geochip 4.0. Six cores were taken from each of the minimal, moderate and extensive permafrost thaw sites; each core contained 6-7 sections by depth (2-3 organic and 3-4 mineral fractions). Various statistical analyses (detrended correspondence analysis, dissimilarity tests and multiple regression tree) all showed that the thaw gradient (sites) was more important than depth in influencing soil microbial community structure, though both had significant impacts. The Simpson diversity index was significantly different across the sites, with the highest value at minimal thawing site and lowest at moderate thawing site. For genes involved carbon degradation, there was no consistent trend across the thaw gradient, though significant differences were observed in some genes. Both methane production and oxidation genes were significantly affected by permafrost thaw, with highest abundances at minimal site and lowest at moderate site. For nitrogen cycling, the abundance of denitrification genes was generally higher at minimal site with only one exception of *narG*, likely due to a more anaerobic condition at minimal site. The genes in assimilatory nitrogen reduction were generally higher at the moderate thaw site, though *nasA* did not differ among three sites, indicating high plant and microbes activity at the moderate site.

Laboratory determination of microbial temperature sensitivity. To determine the temperature sensitivity of microbial respiration (Q_{10}) in soils with different mixtures of labile versus recalcitrant carbon, we are conducting soil incubations of soils from different depths at two constant temperatures (15°C and 25°C). These soils came from a warming experiment in a tundra ecosystem in Alaska and a warming experiment in a grassland in Oklahoma. Three different layers were incubated from Alaska: two surface soils (0 – 15 cm and 15 – 25 cm with high carbon content) and a horizon deeper (>50 cm) within the surface permafrost. From the Oklahoma site, we used soils from control and warmed plots combined with a root exclusion treatment that had kept new root inputs out over the 8 years of the experiment. Instantaneous Q_{10} was measured by exposing soils to 6 different temperatures ranging from 5 to 30°C while measuring CO_2 flux over this range. For Alaskan soils, warmed soils from 0 – 15 cm layer had a slightly higher, but not significant, average Q_{10} than the control plots. No other differences in average Q_{10} were detected for different soil layers or treatments. However, carbon fluxes in the first 15 cm were 10 and 35 times higher than fluxes at 15 – 25cm or from the permafrost soil, respectively.

---

* Presenting author

Modeling integration and development. A new model was developed based on the soil C dynamics model developed at the early stage of the project. The main objectives of the model are to calculate Q_{10} using a data-model fusion technique, and to evaluate dynamics of Q_{10} with recalcitrance of soil organic C (SOC). Q_{10} and turnover rates for different C pools (labile, slow and passive pools) and fraction of each C pool can be optimized using Bayesian probability inversion and a Markov chain Monte Carlo (MCMC) technique. This approach generates posterior probability density functions of model parameters. A previously published dataset using three incubation temperatures (15, 25 and 35 ˚C) and an incubation period of 588 days was used to test the model. Preliminary results show that Q_{10} values for labile and slow C pools can be constrained very well at temperature regimes of both 15-25 ˚C and 25-35 ˚C, but only the Q_{10} value at 25-35 ˚C can be well constrained for the passive pool. Based on the dataset, Q_{10} values increase with recalcitrance of SOC. This preliminary result also indicates that longer incubation studies are needed in order to assess the temperature sensitivity of slower turnover pools, especially at low temperature regimes. Further work will push towards integrating microbial community into an ecosystem modeling framework.

117

Microbial Communities Generating Greenhouse Gases in Thawing Permafrost

Rhiannon Mondav,[1] Suzanne B. Hodgkins,[2] Virginia I. Rich,[3] Carmody K. McCalley,[3] Tyler Logan,[4] Rick Wehr,[3] Changsheng Li,[5] Steve Frolking,[5] Jeffrey Chanton,[2] Patrick M. Crill,[4] Scott R. Saleska,[3] and **Gene W. Tyson**[1]* (g.tyson@uq.edu.au)

[1]University of Queensland, Brisbane, Australia; [2]Florida State University, Tallahassee; [3]University of Arizona, Tucson; [4]Stockholm University, Stockholm, Sweden; and [5]University of New Hampshire, Durham

http://www.ecogenomic.org/melting-permafrost

Project Goals: The IsoGenie project ("Genes, isotopes, and ecosystem biogeochemistry: dissecting methane flux at the leading edge of global climate change") focuses on integrating microbial and geochemical studies to improve our understanding of carbon cycling in a subarctic wetland where climate change induced permafrost thawing is transforming carbon sinks into sources. Our research goals are: (1) to discover relationships between microbial community composition and metabolism to ecosystem carbon cycling; (2) to learn how these relationships are affected by shifting environmental variables, and (3) apply this knowledge to better understand and predict changing carbon budgets in subarctic ecosystems experiencing substantial climate change. One of the aims of this research was to characterize microbial communities along a degradation gradient using 16S rRNA gene amplicons, meta-genomics, -transcriptomics and -proteomics.

High northern latitudes are at the leading edge of global climate change with the effects of warming already evident in degrading permafrost. Increased thawing of permafrost, a significant global carbon sink, releases previously sequestered labile carbon. Thawing initiates a transition from intact permafrost (Palsa hummock) through an intermediate thaw state (collapsed Palsa; hollow) to fully degraded inundated end state (fen). The transition to a fen state has been associated with dramatic increases in biogenic methane production and other greenhouse gases (GHGs). GHGs produced by peat microorganisms are 10-fold and 100-fold higher in the hollow and fen samples respectively, as compared to hummock samples (100 year CO_2 equivalents). The exact combination of *in situ* ecological conditions triggering GHG efflux is unknown, however the rising water table associated with thawing permafrost is believed to lead to anoxic conditions favourable to microbial methane production.

Here, microbial communities along a degradation gradient in Stordalen Mire (Abisko National Park, Sweden) were characterized using 16S rRNA gene amplicon pyrosequencing. Sampling sites and depths were selected based on geochemical data, including GHG flux. Changes in microbial community structure along the degradation gradient were substantial and occurred at the high water mark (middle sampling depth) in the hollow sites (Figure 1). Microbial communities below the waterline in hollow peats (positive methane flux) were relatively low complexity and were dominated by a single archaeal species within the order *Methanomicrobiales* and two bacterial species of the order *Acidobacteriales*. Microbial community richness and composition in the hummock samples (low to negative GHG flux) approached that of non-permafrost soils. *Acidobacteria* were dominant across all sample sites and depths, comprising between 20% and 42% of the OTUs observed. The ubiquity of *Acidobacteria* across all samples is consistent with members of this phylum being adapted to low energy, low nutrient, highly acidic and water stressed environments.

* Presenting author

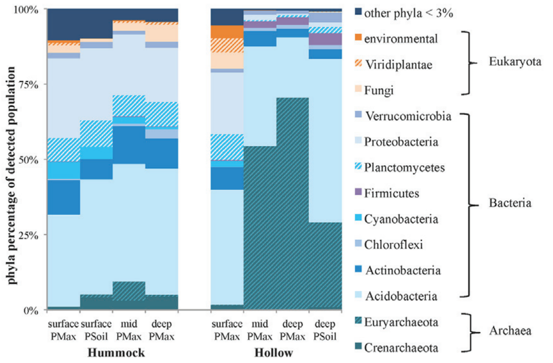


Figure 1. Microbial community composition of active layer peat in (low GHG) Hummock and (mid GHG) Hollow. Two extraction methods, PMax (PowerMax kit, MoBio) and PSoil (PowerSoil Total RNA/DNA co-elution kit, MoBio), were tested in this study.

To further elucidate these emerging patterns in microbial community structure and diversity, samples collected during 2010 and 2011 (June, July, August, and October) are currently being sequenced. The third year of sample collection is scheduled for the 2012. Metagenomic sequencing of select samples are being undertaken to facilitate the recovery of the dominant *Methanomicrobiales* and *Acidobacteriales* genomes. In addition, microbial community profiles generated from all samples, together with GHG data will be used to select samples for combined metagenome, metatransciptome and metaproteome analysis. Integration of meta'omic data with the isotopic geochemistry will clarify ecosystem variables related to GHG production. Insights gained regarding the ecological triggers of GHG emission over the degradation gradient will inform emission projections, filling a gap in current climate modelling scenarios.

The IsoGenie Project is supported by the Office of Biological and Environmental Research in the U.S. DOE Office of Science, Project Grant DE-SC0004632. Fellowships for RM and GWT are funded by the Australian Research Council.

118

The Impact of Global Warming on the Carbon Cycle of Arctic Permafrost: An Experimental and Field Based Study

T. Vishnivetskaya,[2] N. Mykytczuk,[4] J. Allan,[4] K. Cheng,[2] K. Chourey,[3] R. Hettich,[3] A. Layton,[2] X. Liu,[3] J. Murphy,[2] T.J. Phelps,[3] S. Pfiffner,[2] G. Saarunya,[2] B. Stackhouse,[1] C.Y.M. Lau,[1] R. Sanders,[1] S. Myneni,[1] L. Whyte,[4] R. Wilhem,[4] W. Pollard,[4] G. Lamarche-Gagnon,[4] M. Kidd,[5] S. Elliott,[5] P. Hatcher,[6] R.L. Sleighter,[6] P. Bennett,[7] L. Ziolkowski,[8] and **T.C. Onstott**[1]* (tullis@princeton.edu)

[1]Princeton University, Princeton, N.J.; [2]University of Tennessee, Knoxville; [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [4]McGill University, QC, Canada; [5]Los Alamos National Laboratory, Los Alamos, N.M.; [6]Old Dominion University, Norfolk, Va.; [7]University of Texas, Austin; and [8]McMaster University, Hamilton, ON, Canada

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

Project Goals:

1. **Perform ~2 year long, thawing experiments under water saturated and water under saturated conditions on well-characterized, intact cores of Arctic active-layer and permafrost from a proposed reference ecosystem site where CO_2 and CH_4 fluxes, temperatures, humidity, soil moisture, nutrients, microbial diversity and activities, as well as C isotopic analyses, are currently being measured in the field.**
2. **Perform phylogenetic, metagenomic, transcriptomic and proteomic analyses of these cores prior to and during the experiments.**
3. **Characterize the composition of the solid and dissolved organic matter, the inorganic geochemistry, the vertical flux of volatile organic acids, O_2, H_2, CO_2 and CH_4 and the isotopic systematics of CO_2 and CH_4 during the experiments.**
4. **Compare the results of these experiments with field measurements, and based upon both data sets construct a 1D biogeochemical reaction/transport model that predicts the CO_2 and CH_4 release into the atmosphere as permafrost thaws.**
5. **Development of a high sensitivity ^{14}C RNA isotope microarray and a CRDS for C and H isotopic analyses of atmospheric CH_4.**

Forty 1 meter long cores were collected from a 7 meter diameter polygon located near the McGill Arctic Research Station during April 2011 where the depth to permafrost is 70 cm. Total community genomic DNA (cgDNA) was isolated from the 2011 samples using four different commercially available kits to determine the best DNA extraction protocol for metagenomics. The v1 region of the 16S rRNA gene was amplified directly from cgDNA or re-amplified from 16S rRNA amplicons. Taxonomic analysis revealed similarity in bacterial communities derived from different extraction kits. The significant variation in the bacterial community structure as a function of depth in the active layer previously reported in the 2010 core samples does not appear to be an artifact of extraction protocols. Gemmatimonadetes is the dominant bacterial phylum in the uppermost active layer, whereas Firmicutes and Actinobacteria are the dominant phyla near the permafrost table and within the permafrost consistent with geochemical evidence of increasing anaerobicity with depth. The organic matter from these cores ranged from 6% total organic C at the surface to 1% at 10 cm depth with C/N = 13-16. NMR and FT-ICR-MS reveal the composition of this organic matter is significantly different from that reported in boreal soils and peat deposits (1). ^{14}C analyses are underway.

Microcosms of subsamples from the cores amended with select nutritional amendments revealed significant CO_2 production and CH_4 consumption and the presence of stress responsive proteins (e.g. DnaK, GroEL) and proteins essential for energy production and survival under carbon starvation (e.g. F0F1 ATP synthase, acyl-CoA dehydrogenase). Proteins from the genera *Bradyrhizobium, Sphingomonas, Lysinibacillus* and *Methylophilaceae* were detected, and these bacteria were also identified by 454 pyrosequencing on the same samples. Metaproteomics of the pristine core samples, however, yielded relatively few protein identifications suggesting a relative lack of microbial activity and limited microbial biomass. Metaproteome efforts have been focused on evaluation and optimization of experimental protocols to efficiently extract proteins from these low biomass cores. Lipidomic analyses of the core samples are underway.

Continuous *in situ* gas flux analyses of CO_2 and CH_4 fluxes from permafrost and ice-wedge active layer (AL) soils were conducted using CRDS's during July 2011. Polygon AL soil flux showed a net outward CO_2 flux (175 to 3,155 mg/m^2-day) and consumption of atmospheric CH_4 (-1.2 mg/m^2/day). Gas flux from the ice-wedge AL surface was in a similar range as the polygon, having slightly higher maximal CO_2 flux (3670 mg/m^2/day) and net CH_4 consumption (-2.0 mg/m^2/day). The $\delta^{13}C$ of the CO_2 efflux from the surface were consistent with microbial activity, ranging from -10.6 ‰ to -15.5 ‰ for the polygon and ice-wedge soils, respectively. In both AL soils, gas flux fluctuates diurnally. The CO_2 out flux was found to be anti-correlated to surface temperature (R = -0.67), whereas the CH_4 in flux was found to be correlated to surface temperature (R = 0.76). Using a vertical gas probe, the CO_2 concentrations increased with depth and corresponded to a CO_2 flux 19 to 41 mg/h in the polygon AL soils *vs.* 18 to 54 mg/h in the ice-wedge soils. Through the same profile, the CH_4 concentration decreased from 0.59 ppmv to <0.1 ppmv within 30 cm of the surface in the ice wedge and from 0.62 to 0.18 ppmv at the base of the polygon AL. Below the surface, the $\delta^{13}C$ of CO_2 was more ^{13}C depleted than at the AL surface, reaching -18.9 ‰ and -21.3‰ at the base of the AL in the polygon and ice wedge soils, respectively. These data suggest that both polygonal and ice-wedge AL soils contribute net efflux of CO_2 and consumption of CH_4 during the summer season and that the more saturated ice-wedge soils may have a more favorable environment for methanogenic bacterial activity with depth with greater stratification in fluxes of gases through the profile. Defining these relationships is critical for accurately modeling the extent and rate of + or - feedback in global climate models.

Thawing experiments on the 2011 cores have begun with an initial active layer thaw over a one-month period to simulate Arctic spring thaw. Previous reports of dramatic shifts in the phylogenetic and functional gene structure of active layer and permafrost after only 2 weeks of warming (2, 3) motivated us to relate this sudden change to initial C and N fluxes in our thawing experiments. Once the community and fluxes have stabilized we will initiate permafrost thawing.

All components for the Scintillator Layered Imaging Microscope for Ecological Research (SLIMER) have been assembled and the spatial accuracy and resolution validated with an α source. The next step will be to increase S/N for low energy β sources. The sensitivity and reproducibility of the PU-CRDS has been significantly increased to that it is now capable of measuring the $\delta^{13}C$ of 2 ppmv CH_4 with a precision of ±2‰. Modifications are underway to make the PU-CRDS robust under field conditions for transport to the Arctic.

Two posters given at the Fall Meeting of AGU and data shared with the RCN on Permafrost Carbon stimulated much discussion and new collaborations.

References

1. R.L. Sanders, R.L. Sleighter, P.G. Hatcher, S.C.B. Myneni, 2011. Strikingly Different form of organic carbon in soils of Polar regions. Science, *submitted for review*.
2. M.J.L. Coolen,* J. van de Giessen, E.Y. Zhu and C. Wuchter, 2011, Bioavailability of soil organic matter and microbial community dynamics upon permafrost thaw. Environmental Microbiology, doi:10.1111/j.1462-2920.2011.02489.x
3. R. Mackelprang1, M.P.Waldrop, K.M. DeAngelis, M.M. David, K.L. Chavarria, S.J. Blazewicz, E.M. Rubin and J.K. Jansson, 2011. Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. Nature, doi:10.1038/nature10576.

This project is supported by U.S. Department of Energy Program on Biological Systems Research on the Role of Microbial Communities in Carbon Cycling grant number DE-SC0004902 to Princeton University, University of Tennessee-Knoxville, McGill University, Oak Ridge National Laboratory and Los Alamos National Laboratory.

119

IsoGenie: Microbes, Ecosystem Biogeochemistry, and Climate Change: Dissecting Methane Flux in a Thawing Permafrost Peatland in Northern Sweden

Carmody K. McCalley,[1]* Suzanne B. Hodgkins,[2]* Rhiannon Mondav,[3] Tyler Logan,[4] Rick Wehr,[1] David D. Nelson,[5] J. Barry McManus,[5] Mark S. Zahniser,[5] **Changsheng Li**,[6] **Steve Frolking**,[6] **Jeffrey Chanton**,[2] **Virginia Rich**,[1] **Patrick M. Crill**,[4] **Gene Tyson**,[3] and **Scott R. Saleska**[1] (saleska@email.arizona.edu)

[1]University of Arizona, Tucson; [2]Florida State University, Tallahassee; [3]University of Queensland, Brisbane, QLD, Australia; [4]Stockholm University, Stockholm, Sweden; [5]Aerodyne Research Inc., Billerica, Mass.; and [6]University of New Hampshire, Durham

---

* These authors contributed equally to this work

Project Goals: As permafrost thaws, increasing CH_4 emissions from northern wetlands are likely to cause positive feedback to atmospheric warming. The IsoGenie project seeks to discover how functional relations between biogeochemical processes, particularly methane production, and the underlying microbial community dynamics are affected by climate change-induced permafrost thaw. This goal is enabled by recent transformative methodological advances in both ecosystem science and molecular biology that allow high-volume tracking of isotopic composition

of carbon gases, as well as the genetic potential and gene expression in the microbial communities which produce those gases.

To achieve this goal, we measured concentration profiles and surface fluxes of carbon gases (CH_4 and CO_2), along with their isotopic compositions, across a gradient in permafrost thaw from palsa (with underlying permanently frozen peat), to recently collapsed and flooded palsa dominated by *Sphagnum spp.* (intermediate thaw), to fully inundated sites dominated by *Eriophorum angustifolium* (fully thawed). At the same time, we sampled soil microbial communities in these sites to characterize composition, metabolic potential and gene expression. Both isotopic composition of carbon gases, and meta-genome/-transcriptome/-proteome data can be used to identify active metabolic pathways of methane production, and we hypothesized that the two methods would give consistent results. This poster focuses on biogeochemical and isotopic results, and a companion poster focuses on microbial community results (Mondav et al, "Microbial Communities Generating Greenhouse Gases in Thawing Permafrost").

Along this environmental gradient, from permafrost to fully thawed sites, the lability of the peat increases significantly as determined in incubations of peat material and monitoring of methane and carbon dioxide production rates. Coincident with this trend is an increase in methane surface fluxes and an increase in $_{13}$C isotopic composition of methane (from ~ -80‰ in intermediate thaw sites to ~-65‰ in fully thawed sites) suggesting a shift in methane metabolism towards acetate fermentation and away from CO_2 reduction. We also observed significant temporal variation during the 2011 growing season in the carbon isotopic composition of methane fluxes, indicative of shifts in methanogenic and methanotrophic activity. These observations will be used to test a biogeochemical model (the DNDC model) of methane production that includes alternative production pathways under different redox conditions.

Together, these initial results suggest that thaw-induced changes in hydrology and plant community composition increase peat lability, stimulating acetate fermentation and yielding increased methane emissions. We conclude that the biological controls on metabolic pathways of methanogenesis, though poorly represented in most ecosystem models, may nonetheless be important, in interaction with permafrost thaw dynamics, in determining future CH_4 emissions under changing climate.

The IsoGenie Project is supported by the Office of Biological and Environmental Research in the U.S. DOE Office of Science, Project Grant DE-SC0004632.

120

The Impact of Precipitation and Nitrogen Deposition Changes on the Microbial Community and Greenhouse Gas Cycles of a Southern California Grassland

Emma L. Aronson* (earonson@uci.edu) and **Steven D. Allison**

University of California, Irvine

Project Goals: The overall project goal is to describe the impact of changes in precipitation regimes and nitrogen deposition levels, predicted by regional and global models for Southern California, on grassland leaf litter decomposition and associated microbial community. In this presentation we describe a piece of the project, the goals of which are to investigate the impact of these treatments on the structure and function surface soil microbial composition, with specific focus on greenhouse gas fluxes. Further, this project seeks to strengthen the connection between microbial community structure and function, and potentially enhance our understanding of the biospheric role in greenhouse gas cycles.

Fluctuations in greenhouse gas (GHG) concentrations in the atmosphere can lead to profound climatic and environmental changes. Likewise, environmental changes can influence variation in soil consumption and release of key greenhouse gases. Carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O) are currently increasing in the atmosphere at high rates due to human activities. Accurately understanding natural and anthropogenic roles in ever-changing greenhouse gas cycles is necessary for a more complete understanding of biosphere-atmosphere interactions. It is only with this understanding that decisions can be made at local, national and international levels to mitigate potentially harmful climatic changes. This project examines the biological mechanisms behind the response and resilience of soil microbial communities to change. This is examined in both community composition as well as function, in terms of production and consumption of CH_4, CO_2 and N_2O, thereby enhancing our understanding of the biospheric role in greenhouse gas cycles.

Southern California provides an excellent natural laboratory for investigating the relationships between climate, plant composition, net primary production, soil microbial ecology, and ecosystem function. California's coastal climate is Mediterranean, with a mild wet winter, a warm or hot dry summer, and extreme inter-annual precipitation variation associated with ocean currents. Soil trace gas fluxes may be variable in this ecosystem due to this high climatic variability on seasonal and inter-annual scales, with intense pulses of precipitation and extensive drought. This region is likely to experience profound environmental change in the next 50 years, due to high densities of fossil fuel combustion, NOx emission, and nitrogen deposition (Fenn et al. 2003). Total nitrogen deposition is 2.5 to 4.0 gN $m^{-2}yr^{-1}$ in many regions of the San Gabriel and San Bernardino Mountains, and increasing (Fenn et al. 2003). Global Circulation Model (GCM) and Regional Climate Model (RCM) runs almost uniformly project significant warming in California (IPCC 2007), which can lead to increased drought conditions. Projections of future precipitation are less certain, with many GCMs projecting increased winter precipitation, and some models predicting a dramatic change (IPCC 2007). Hence the study of the impacts of increased N deposition within the context of both increased and decreased precipitation is relevant to understanding the response of this system to projected changes in this environment.

This experiment, performed at Loma Ridge in the Irvine Ranch Conservancy, manipulates N and rainfall in a grassland community in a factorial combination of replicated plots. Nitrogen is added in fall and mid-spring as slow-release granular ammonium nitrate (osmocote) pellets at 6 gN m^{-2} yr^{-1}. This level of N addition in treatment plots is comparable to the higher levels of nitrogen deposition in the region (Fenn et al. 2003) and is similar to that used in other studies. The 50% precipitation increase and decrease treatments, which have been in place since 2007, are achieved by manipulating both the average size and number of storms (Hanson 2000), to simulate anticipated precipitation shifts. The experiment uses rainout shelters with clear polyethylene retractable roofs to remove ~50% of the annual precipitation from the low water plots by selective closure during a subset of the winter storms. The water draining off the shelters is collected with metal gutters and PVC pipe, stored and applied to increased precipitation plots.

Gas flux rates of N_2O, CH_4 and CO_2 have been measured across the start of the rainy season of 2011-2012. All gases are measured monthly using static vented chambers and gas chromatography. In addition, CO_2 is measured constantly using automated flow-through chambers. The initial data show some striking trends. As predicted, decreased rainfall is associated with lower ecosystem respiration across plots, but increased rainfall was also shown to coincide with lower respiration rates than controls. Increased N deposition was found to correlate with increase N_2O release, as expected. No trends have yet been observed in CH_4 flux, but both consumption and release have been observed in the site.

Concurrent with gas flux measurements, *in situ* soil moisture and temperature data are being collected. Soil cores are removed from each plot when gas flux measurements are performed. Cores are frozen immediately and returned to the lab where DNA and/or RNA are extracted. Pyrosequencing (454) analysis of the microbial community has been performed on a subset of samples using the same 16S primers and bioinformatics pipeline currently employed to examine litter microbial communities. The initial data from this subset of samples will be presented.

This research has only just started, with several molecular techniques to be employed later on on samples across time and treatments. Changes in the relative abundances the functional genes *pmoA* (for methanotrophy), *nirS*, *nirK* and *nosZ* (for denitrifiction), and *amoA* (for nitrifiction), along with associated mRNA activity, will be measured by qPCR and rt-qPCR, respectively. Also, samples of pooled replicate plots are being analyzed by a full meta-genomic Illumina sequencing protocol, with data forthcoming.

References

1. Fenn M. E., J. S. Baron, E. B. Allen et al. 2003. Ecological effects of nitrogen deposition in the western United States. Bioscience 53(4):404-420.
2. Hanson P. J. 2000. Large-scale Water Manipulations. In Sala OE, RB Jackson, HA Mooney, RW Howarth (editors). Methods in Ecosystem Science, Springer, NY.
3. IPCC. 2007. Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (eds Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M. and Miller, H.L.). Cambridge University Press, Cambridge and New York, NY.

This project is supported by the DOE Genomic Sciences program as well as the NOAA Climate and Global Change Postdoctoral Fellowship.

121

Reconciling Phylogeny and Function During Plant Litter Decomposition by High-Throughput Functional Metagenomics

Mari Nyyssönen[1]* (mjnyyssonen@lbl.gov), Claudia Weihe,[2] Michael L. Goulden,[3] Kathleen K. Treseder,[2] Jennifer B.H. Martiny,[2] Adam C. Martiny,[2,3] **Steven D. Allison**,[2,3] and Eoin L. Brodie[1]

[1]Ecology Department, Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; [2]Department of Ecology and Evolutionary Biology, University of California, Irvine; and [3]Department of Earth System Science, University of California, Irvine

Project Goals: The overall aim of this project is to link functional traits that influence carbon cycling with individual microbial taxa in order to build predictive trait-based models of ecosystem responses to global change. The functional traits are connected to microbial taxa with DNA sequencing, novel functional assays, experimental manipulations and mathematical models that are used to investigate the distribution of extra-cellular enzyme genes and functions involved in plant litter decomposition across taxonomic groups of microbes. The resulting trait-based frameworks are then used to model how altered water and nitrogen availability as well as plant species composition impact plant litter decomposition rates in annual grassland.

This presentation focuses on one aspect of the discovery and identification of functional traits involved in plant litter decomposition, the functional screening of metagenomic libraries. A total of twenty-four metagenomic fosmid libraries have been constructed from high molecular weight DNA isolated from decomposing plant litter collected over one year period from a global change experiment undergoing rainfall and nitrogen manipulations. The libraries encompass over 20 Gb of metagenomic DNA and cover a range of microbial taxonomic groups including Actinobacteria, Bacteroidetes and different classes of Proteobacteria as well as Fungi.

The 24 libraries are currently being screened for activities involved in cellulose, hemicellulose, chitin, lignin and starch degradation as well as peptide breakdown and mineralization of organic phosphate using novel functional assays. These automated high throughput assays are based on colorimetric and fluorescent substrates and have been optimized to allow multiplexed screening of both endolytic and exolytic hydrolysis activities required during different stages of decomposition. Several cellulose, hemicellulose, chitin and starch degrading as well as phosphatase and proteinase producing clones have been identified.

Clones expressing the targeted activities are sequenced with the Illumina sequencing platform and the identified hydrolytic genes are characterized for enzyme kinetics and given phylogenetic context for incorporation into trait-based microbial decomposition models.

This work is funded by the U.S. Department of Energy, Office of Science, BER Biological Systems Sciences Division. Part of this work was performed at the Lawrence Berkeley National Laboratory under contract number DE-AC02-05CH11231.

122

Atmospheric Nitrogen Deposition and Microbial Mechanisms Enhancing Soil Carbon Storage

Sarah D. Eisenlord[1]* (seisen@umich.edu), Ivan P. Edwards,[1] Elizabeth E. Entwistle,[1] **Kurt S. Pregitzer**,[2] **Cheryl R. Kuske**,[3] and **Donald R. Zak**[1]

[1]School of Natural Resources and Environment, University of Michigan, Ann Arbor; [2]College of Natural Resources, University of Idaho, Moscow; and [3]Bioscience Division, Los Alamos National Laboratory, Los Alamos, N.M.

http://www.webpages.uidaho.edu/nitrogen-gradient/Default.htm

Project Goals: Since the industrial revolution, the amount of reactive nitrogen (N) deposited from the atmosphere has increased across large areas of the Northern Hemisphere (e.g., from 50-100 to 1500-2000 mg N m^{-2} y^{-1}). Agricultural and industrial activities are largely responsible for this increase, which has the potential to disrupt many of the fundamental ways in which ecosystems function—ranging from the growth and vigor of trees to the microbial mechanisms mediating the cycling and storage of carbon (C) in soil. Our long-term experiment was established to understand the mechanisms by which atmospheric N deposition could alter the ecosystem processes in sugar maple dominated forests of the Great Lakes region. Exposure to 18 years of experimental N deposition, at a rate expected mid-century, has slowed the decay of plant litter, which has increased soil C storage as well as export of phenolic DOC; this biogeochemical response has occurred in the absence of a change in litter biochemistry or production. Presently, we are testing the hypothesis that the slowing of litter decay and the accumulation of soil organic matter in response to anthropogenic N deposition is governed by the environmental regulation of microbial gene expression. We predict that chronic experimental N deposition has down regulated the transcription of fungal genes encoding lignocellulolytic enzymes, thereby slowing litter decay and opening a niche for other, less effective lignocellulolytic soil microorganisms to occupy. To understand whether these hypothesized mechanisms are at work in our experiment, we have employed molecular genetic approaches to examine the community and physiological responses of saprotrophic fungi and *Actinobacteria* residing in forest floor and surface soil.

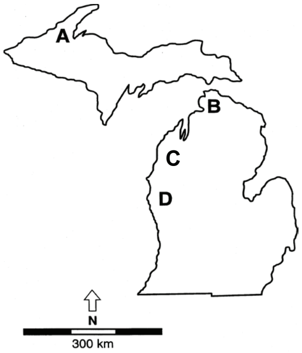


Figure 1. Distribution of replicate sites spanning the north-south range of northern hardwood forests in the Upper Great Lakes region.

Experimental Design: Since 1994, we have experimentally simulated increased atmospheric N deposition in replicate stands of a northern hardwood forest ecosystem stretching across a 500-km climatic gradient (Fig. 1). Our study sites deliberately encompass the north-south geographic range of the of the sugar maple-dominated (*Acer saccharum* Marsh.) northern hardwood forest in the Great Lakes region of North America, enabling us to generalize our experimental results across this geographic region. These sites are floristically and edaphically matched (>80% sugar maple on Typic Haplorthods), but they differ in climate along the north-south latitudinal gradient (Fig. 1). The study sites also span an atmospheric N deposition gradient, over which NO_3^--N composes ~60% of wet-plus-dry deposition. There are six 30-m x 30-m plots at each study site; every plot is surrounded on all sides by a 10-m wide treated buffer. Three plots at each site receive ambient atmospheric N deposition. The other three plots at each site receive ambient N deposition plus 3 g NO_3^--N m^{-2} y^{-1}, a rate approaching that expected by 2050 across large portions of North America and other regions of the Earth. The additional N is delivered over the growing season in six equal applications (0.5 g N m^{-2} $month^{-1}$) of solid $NaNO_3$ pellets, which are broadcast over the forest floor.

Hypothesis Testing: Basidiomycete and ascomycete fungi are the primary agents of plant cell wall decay, and a change in their composition or decline in activity could slow litter decay under experimental N deposition. To characterize the actively metabolizing community of these organisms, we used cDNA clone libraries constructed from 28S fungal rRNA. The active basidiomycete communities under ambient and experimental atmospheric N deposition differed significantly in terms of membership as well as their dispersion across a phylogenetic tree. Furthermore, suggestive, albeit nonsignificant, differences in the fraction of unique phylogenetic branch length (i.e., the UniFrac metric) between ambient and experimental atmospheric N deposition were observed for forest floor basidiomycetes. In contrast, the active ascomycete communities under ambient and experimental atmospheric N deposition did not exhibit significant differences in these same metrics. Collectively, our observations indicate that experimental N deposition has altered the composition of litter decaying fungi and that these changes have ecosystem-level implications for the cycling and storage of C in forest ecosystems.

Actinobacteria also are one of the few groups of saprotrophic microorganisms which oxidatively depolymerize lignin, producing substantial soluble polyphenolics during the process. These organisms could plausibly become more important agents of lignin decay, if atmospheric N deposition suppresses the activity or alters the composition of lignolytic basidiomycete and ascomycete fungi. To test this idea, we quantified actinobacterial abundance and community composition under ambient and experimental N deposition. Actinobacterial abundance was assessed using quantitative PCR of 16S rRNA and community composition was evaluated using clone libraries and phylogenetic community analyses (i.e., Libshuff and Unifrac). Contrary to our expectation, experimental atmospheric N deposition had no effect on actinobacterial abundance in the forest floor (~10^{10} gene copies/g); however, it significantly decreased actinobacterial abundance by 47% in surface mineral soil. Our analyses revealed experimental N deposition further elicited a significant membership change in forest floor and surface soil communities, as well as significant differences in the phylogenetic diversity of forest floor *Actinobacteria*. This shift in community composition occurred in concert with the slowing of plant litter decay, the accumulation of soil organic matter, and the greater production of phenolic DOC.

To determine whether slower decomposition rates resulted from down-regulation of the transcription of key lignocellulolytic genes, we quantified the expression of fungal genes encoding key cellulolytic (cellobiohydrolase, *cbhI*) and lignolytic (laccase, *lcc*) enymes. Our results indicate that the community-scale expression of *cbhI* under experimental N deposition did not differ significantly from that under ambient N deposition. In contrast, expression of *lcc* was significantly down-regulated by a factor of 2-4 fold relative to its expression under ambient N deposition. Our results suggest that chronic atmospheric N deposition may lower decomposition rates through a combination of reduced expression of ligninolytic genes such as *lcc*, as well as compositional changes in the fungal community. More importantly they indicate that ecosystem response to atmospheric N deposition, a wide-spread agent of global change, is controlled by the environmental regulation of fungal gene expression.

Our research is funded by grants from the U.S. Department of Energy, Office of Biological and Environmental Research, and the National Science Foundation.

123

Microbial Response to Modified Precipitation Patterns in Tallgrass Prairie Soil: Molecular Mechanisms, Activity Rates, and Organic Matter Dynamics

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

[1]Oregon State University; [2]Lawrence Berkeley National Laboratory; [3]Oak Ridge National Laboratory; [4]Kansas State University; and [5]DOE Joint Genome Institute

http://cropandsoil.oregonstate.edu/soils/research/myrold/konza

Project Goals: Identify microbial physiological responses to modified precipitation in Great Plains prairie soils, and assess the implications for carbon cycling dynamics.

A significant amount of carbon (C) is processed and stored in prairie soils: grasslands cover 6.1-7.4% of the earth's land surface and hold 7.3-11.4% of global soil C. Global change models predict that the future precipitation regime across the North American Great Plains will entail less frequent but larger rainfall events. The response of prairie soil microbial C processing and allocation to this scenario of higher hydrologic variability is not known, but will be a key determiner of the future capacity for prairie soil C sequestration. We are approaching this problem by assessing soil microbial function (respiration, C utilization efficiency, extracellular enzyme activity) and molecular indicators of dominant C allocation pathways (soil transcriptome, proteome and metabolome) under ambient and experimentally modified precipitation regimes.

The Rainfall Manipulation Plots (RaMPs) at the Konza Prairie Long-Term Ecological Research (LTER) site in north-eastern Kansas, USA is a replicated field manipulation of the timing and magnitude of natural precipitation that was established in 1998. This experiment does not modify the total amount of growing season rainfall, it imposes extended dry periods and larger, less frequent rainfall events. We collected soil before, during and after rainfall events in both ambient and extended precipitation interval (more "droughty") treatments and measured microbial growth, respiration and potential organic matter degradation responses. Notable results include (1) Equivalent rainfall

* Presenting author

events caused equivalent microbial respiration responses in ambient and interval manipulation plot soils, but biomass increased after the rainfall in the extended precipitation interval plots only. This implies a greater C use efficiency, or greater potential for belowground C retention, in "droughty" soils. (2) C:N ratio of biomass was increasingly high as soil water content decreased. This implies a physiological and/or population-level shift in the microbiota at low soil water content. (3) Extracellular enzyme activity responses were mixed across the suite of functional groups measured, with one consistent response: cellulose hydrolysis potential was always lower 5 days after rainfall. This implies a decreased dependency on soil organic matter degradation, with a lag period, after rainfall events; perhaps related to plant activity and root exudate deposition belowground.

These results lead to hypotheses regarding microbial physiological adaptation to drought stress in prairie soils. We are collecting molecular data (454 sequencing and QPCR of bacterial 16S rRNA and fungal ribosomal genes and transcripts, full transcriptomes and proteomes) to test these hypotheses. (H1a) Microbial taxa that respond quickly to increased water availability after drought are more active in soil with an altered precipitation regime history. (H1b) Transcripts and proteins from COGs indicative of growth, not maintenance, will be more abundant after rainfall in the "droughty" plots. (H2a) In soils with low water contents, transcripts and proteins driving trehalose (or other compatible solute) production will be more abundant. (H2b) In soils with low water contents, fungal cells will be more abundant. (H3) Higher root exudate uptake and metabolism 5 days after rainfall events will be evidenced by higher abundance of sugar and amino acid transporters. Directly addressing these mechanistic hypotheses would not be possible without "Omics" approaches.

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.

Early Career Award

124

Microbial Communities in Restored Wetland Sediments

Shaomei He[1]* (she@lbl.gov), Mark Waldrop,[2] Lisamarie Windham-Myers,[2] Stephanie Malfatti,[1] Tijana Glavina del Rio,[1] and **Susannah G. Tringe**[1]

[1]DOE Joint Genome Institute, Walnut Creek, Calif. and [2]U.S. Geological Survey, Menlo Park, Calif.

http://www.jgi.doe.gov/sequencing/why/CSP2011-tringe.html

Project Goals: In this project, we use high-throughput sequencing tools to characterize microbial communities in restored wetland sediments, aiming to understand how the biotic and abiotic environmental factors govern microbial community structure and how microbial communities influence carbon flux, and thus impact long-term biological carbon sequestration.

On peat islands previously drained for agriculture, aerobic decomposition of peat has led to substantial land surface subsidence, thereby increasing risk of levee failure. Restoration of wetlands has a great potential to reverse land subsidence by slowing down decomposition and favoring new peat accumulation. In addition, the high primary production and slow decomposition rates found in restored wetlands may result in a net atmospheric CO_2 sequestration. However, one major concern is the emission of CH_4 that could potentially offset the carbon captured due to primary production. In wetland ecosystems, microbial communities play key roles in governing greenhouse gas flux, yet they are poorly characterized due to their high complexity. By using powerful high-throughput sequencing tools, we aim to identify community patterns, indicator species, genes or pathways that are associated with peat accretion rates and CH_4 flux, and these will providing vital information for better modeling of wetland carbon flux and management.

In this preliminary study, we collaborated with scientists at the U.S. Geological Survey (USGS), and collected belowground samples from a restored wetland from a USGS pilot-scale restoration project on Twitchell Island in the Sacramento/San Joaquin Delta, CA. The wetland is continuously fed by water from San Joaquin River, and is primarily vegetated with cattails (*Typha* spp.) and tules (*Schoenoplectus acutus*). We selected three sites that have varied proximity to the inflow, thus exhibiting gradients in physicochemical conditions and peat accretion rates. From each site, we collected three sample types, including the bulk decomposed material, cattail rhizomes and tule rhizomes. Pyrosequencing of amplified V8 regions of 16S rRNA genes was used to generate microbial community profiles. In parallel, mesocosm anaerobic incubation was conducted to evaluate CO_2 and CH_4 flux.

Our sequencing data indicate that wetland community composition is primarily governed by sampling site, and secondarily by sample type. Particularly, wetland communities from these three sites transited in a direction largely consistent with the physicochemical gradients along these sites. The mesocosm incubation experiment showed that CO_2 flux was significantly higher in the rhizome samples than in the bulk samples. By contrast, difference in CH_4 flux was more related to sample sites. Low CH_4 flux communities were associated with the site closest to the inflow, correlated to higher availabilities of electron acceptors, particularly sulfate and nitrate. Some of their more abundant microbial populations, as compared to high CH_4 flux communities, are likely reducers of these electron acceptors, as suggested by their closest microbial isolates. High CH_4 flux communities were associated with sites further from the wetland inflow, which have shown higher peat accretion rates. These sites harbored more abundant methanogenic archaeal populations, which likely contributed to the higher methane flux observed.

Currently, additional samples collected in a different season are being analyzed to identify seasonal effects, and comparative metagenomic analyses are being conducted to reveal differences in community functional profiles.

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

125

Great Prairie Grand Challenge Soil Metagenome Sequencing Project

Janet K. Jansson[1,2]* (jrjansson@lbl.gov), Titus Brown,[3] Patrick Chain,[2] James Cole,[3] Maude David,[1] Jordan Fish,[3] Adina Howe,[3] Regina Lamendella,[1] Kostas Mavromattis,[2] Susannah Tringe,[2] and James M. Tiedje[3]

[1]Lawrence Berkeley National Laboratory; [2]DOE Joint Genome Institute; and [3]Michigan State University

Project Goals: Our overarching goal is to use deep metagenome sequencing to gain an understanding of the soil microbial community phylogenetic and functional gene repertoire in the U.S. great prairie and adjacent cultivated soils. This information will be used to determine the impact of land management (i.e. cultivation) on microbial community structure and function. In addition, this information will provide a database of microbial genes in native prairie soils that can serve as a baseline for understanding the impacts of perturbation and climate change on processes including soil carbon cycling. In order to achieve these goals the consortium is developing novel tools for metagenome sequencing, assembly, annotation and gene mining.

The United States Great Prairie contains approximately 35% of the total soil organic carbon stocks in the continental U.S. These carbon pools may be impacted in unknown ways by altered climate conditions and by land management. At the Joint Genome Institute, the U.S. Great Prairie was chosen as a “grand challenge” project for deep metagenome sequencing of a complex microbial community. This project is particularly challenging because of the high microbial diversity in soil that complicates metagenome assembly.

To date we have sequenced nearly 2 terabases of DNA from eight locations: 1) Wisconsin native prairie, 2) Wisconsin cultivated corn, 3) Wisconsin cultivated switchgrass, 4) Wisconsin restored prairie, 5) Iowa native prairie, 6) Iowa cultivated corn, 7) Kansas native prairie and 8) Kansas cultivated corn. Each site was sampled along a quadrant for a total of 8 samples per site. All of the samples were first profiled for their microbial community compositions by 454 pyrotag sequencing. In addition, the DNA from a central core from each location was extensively shotgun sequenced using the Illumina platform. A new algorithm was developed for assembly of the metagenome data. Different screening tools were developed to screen the assembled data and raw reads for key genes of interest in the carbon and nitrogen cycle.

454 pyrotag sequence data. We found that cultivation has a major impact on the composition of the soil microbial communities at all locations. There was similarity in the microbial community structures from the prairie soils in the three different locations. Some specific microbial species were more or less abundant in prairie soils, compared to cultivated soils, thus potential indicators of land use history. For example some Bradyrhizobia were more abundant in all of the native prairie soils. The restored prairie site had a microbial community composition that was intermediate between the cultivated corn and the native prairie soils, suggesting that this prairie community was progressing towards a restored (native) state.

Metagenome assembly. A novel approach was developed for metagenome assembly based on data reduction by filtering and normalization, followed by dividing the data into smaller disconnected sets to aid in assembly. This approach was tested on the Iowa corn and Iowa prairie metagenomes and is currently being applied to the remaining metagenomes. For more details about this approach, please see the poster presented by Adina Howe.

Metagenome mining. We first used a Hidden Markov Model (HMM) approach to screen for *nif*H genes in the Iowa corn and prairie assemblies. In addition, we developed a set of validated genes for different key pathways, including cycling of carbon and nitrogen. This gene database consists of “maudules” that represent individual pathways of interest. Currently we are using the maudules to mine the Iowa corn and prairie and the Kansas prairie metagenomes.

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

126

Development of Integrated “Omics” Approach for Assessing Microbial Cycling of Carbon in Prairie Soil Using a Model Soil Bacterium: *Arthrobacter chlorophenolicus*

Maude M. David[1]* (MMDavid@lbl.gov), Lydia Zeglin,[2] Peter Bottomley,[2] Robert Hettich,[3] Kristen Corrier,[3] Ari Jumpponen,[4] Konstantinos Mavromatis,[5] Charles Rice,[4] Manesh Shah,[3] Susannah Tringe,[5] Nathan VerBerkmoes,[3] David Myrold,[2] and **Janet K. Jansson**[1,5]

[1]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]Oregon State University, Corvallis; [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [4]Kansas State University, Manhattan; and [5]DOE Joint Genome Institute, Walnut Creek, Calif.

http://www.lbl.gov/

Project Goals: Soils of the U.S. 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 and more extreme precipitation events,

* Presenting author

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 impact of altered rainfall patterns on the carbon cycling dynamics of the soil microbiota in Kansas native prairie soil using a combination of omics approaches.

To develop protocols for RNA and protein extraction from soil, we spiked the soil with a soil bacterium that has been genome sequenced, *gfp* (green fluorescent protein)-tagged *Arthrobacter chlorophenolicus*. The *gfp* gene was used as an internal standard for accurate quantification. The model strain was inoculated into sterile and non-sterile Kansas prairie soil, amended or not with general (acetate) and specific (chlorophenol) carbon substrates.

Total RNA was extracted from the samples and is currently being sequenced using the Illumina platform to obtain metatranscriptomes. Target genes of interest were quantified by quantitative PCR and RT-QPCR. Total proteins were also extracted from the same samples to obtain metaproteomes. Metaproteomics proved particularly challenging due to the high humic acid content in the Kansas prairie soil. Therefore, we tested and optimized methods for extraction of proteins from the soil for subsequent measurement by shotgun metaproteomics via 2d-LC-MS/MS on an LTQ Velos mass spectrometer.

The first RNA-based results confirmed that the *gfp* transcript could be detected under most conditions and was thus a good estimator of *A. chlorophenolicus* abundance and activity. The initial metaproteome data indicated that several of the enzymes involved in acetate and chlorophenol degradation pathways were expressed in soil. In addition, several proteins involved in response to stress (thioredoxin, chaperonin, cold-shock proteins, etc.) were expressed. We also detected high levels of a flagellin protein in the soil amended with chlorophenol. Comparison of *A. chlorophenolicus* protein yields from sterile and non-sterile soil showed the impact of high background soil diversity on complicating the proteomic results.

These methods are now being applied to the samples acquired from the rainfall manipulation plots at the Konza Prairie Long-Term Ecological Research station. At JGI we have sequenced over 150 Gb of DNA from Kansas soil. To analyze the data using omics we have developed a new comprehensive functional database. We also focused on improving the sequence annotation by screening sequences belonging to particular KO (KEGG Orthology) families. We used this information to build a HMM (Hidden Markov Model). The database was structured into "maudules", which were chosen to represent specific functions with ecological context (such as denitrification, methanogenesis, etc.). This validation approach will enable us to directly explore omics data for key functions of interest, and to integrate the different omics analyses to improve our understanding of carbon cycling processes in the prairie.

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.

127

Microbial Mediation of Litter Decomposition in Soil: The Role of Plant Roots

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

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

Project Goals: Our project (*Plant Stimulation of Soil Microbial Community Succession: How Sequential Expression Mediates Soil Carbon Stabilization and Turnover*) focuses on a fundamental understanding of C cycling in soil as mediated by soil microorganisms and their interactions with plants. Our work investigates how the interactions between roots and soil microorganisms affect transformations of root derived C, decomposition and loss as CO_2, as well as C sorption and stabilization in soil at ambient and elevated levels of atmospheric CO_2. Through our research we seek to provide a mechanistic understanding of the conversion of root C to stabilized soil C, clarify the impacts of increased concentration of CO_2 on soil C sequestration, and substantially expand our understanding of molecular regulation of terrestrial C cycling.

Soil organic C is the largest component of the terrestrial C cycle, with fluxes through the C pool mediated by soil microorganisms and modulated by their interactions with plant roots and root exudates. We characterized the metabolic profiles of *Avena barbata* root exudates collected in sterile hydroponic systems using gas chromatography-mass spectrometry (GC-MS). A variety of compounds were detected in exudate samples, including carbohydrates (e.g. glucose, fructose, galactose), low molecular weight organic acids (e.g. oxalic, malic, maleic acids), amino acids and amides (e.g. lysine, serine, glycine), fatty acids (e.g. arachidic, lauric, oleic acids), sterols (e.g. cholesterol) and others (e.g. hydroxylamine, glycerol). We are using this library of identified exudate compounds to inform spatially explicit analysis of *A. barbata* root exudates patterns in soil. To do so, *A. barbata* seedlings grown in microcosms were pulse labeled with $^{13}CO_2$ and the exudates were sorbed to an initiator-treated silicon wafer. Organic compounds on the wafer were then analyzed by nanostructure-initiator mass spectrometry (NIMS). Using the hydroponically-generated library, we are currently identifying compounds detected using the NIMS approach.

A concurrent effort examines the effect of live *A. barbata* on the mineralization of ^{13}C-labeled root litter in soil over two growing seasons. The mineralization rates of labeled root litter in the presence of *A. barbata* were determined and compared with those in a no-plant treatment by measuring total CO_2 and $^{13}CO_2$ fluxes. $^{13}CO_2$ flux in the no-plant treatment was significantly higher than in the presence of live *A. barbata* in the early litter decomposition stage, suggesting an initial negative priming effect of live roots on litter mineralization. However, the trend changed after 40 days; after 70 days, the $^{13}CO_2$ flux rates from decomposing root litter became higher in the presence of live *A. barbata* plants, suggesting a positive priming effect.

The metabolic data and mineralization rates measured in the presence and absence of live plant roots will guide future investigations of the microbial metabolic pathways responsible for rhizosphere C processing. This work will be extended to distinguish the C-cycling transcriptome of rhizosphere microbial communities utilizing ^{13}C-labeled exudates to provide a mechanistic basis for understanding organic matter priming in the rhizosphere.

Funding for this research was provided under contract FOA DE-PS02-09ER09-25 through Genomic Sciences program of the Office of Biological and Environmental Research, U.S. Department of Energy to the University of California, Berkeley. Part of this work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344 and at Lawrence Berkeley National Laboratory under the auspices of the University of California—contract DE-AC02-05CH11231.

128

Linking Microbial Identity and Succession Patterns to Uptake of Plant-Derived Carbon in the Rhizosphere

Jeremy Bougoure[1]* (bougoure1@llnl.gov), Erin Nuccio,[2] Xavier Mayali,[1] Shengjing Shi,[2] Donald Herman,[2] Zhili He,[4] Jizhong Zhou,[4] Ulisses Nunes da Rocha,[3] Eoin Brodie,[3] Jennifer Pett-Ridge,[1] and **Mary Firestone**[2,3]

[1]Chemical Sciences Division, Lawrence Livermore National Laboratory, Livermore, Calif.; [2]University of California, Berkeley; [3]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; and [4]University of Oklahoma, Norman

Project Goals: Our project (*Plant Stimulation of Soil Microbial Community Succession: How Sequential Expression Mediates Soil Carbon Stabilization and Turnover*) focuses on a fundamental understanding of C cycling in soil as mediated by soil microorganisms and their interactions with plants. Our work investigates how the interactions between roots and soil microorganisms affect transformations of root derived C, decomposition and loss as CO_2, as well as C sorption and stabilization in soil at ambient and elevated levels of atmospheric CO_2. Through our research we seek to provide a mechanistic understanding of the conversion of root C to stabilized soil C, clarify the impacts of increased concentration of CO_2 on soil C sequestration, and substantially expand our understanding of molecular regulation of terrestrial C cycling.

Soil organic C is the largest pool within the terrestrial C cycle and fluxes within this pool are regulated by complex interactions between edaphic factors, plants, soil microorganisms and minerals. Plant roots and their exudates exert control over the microorganisms mediating decomposition of complex 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.

In this research we are assessing structure and successional patterns within the microbial community mediating carbon transformations in the rhizosphere of *Avena barbata* (slender wild oat), by assessing community transcription and substrate (i.e., root exudates or root litter) use profiles along actively-growing roots (i.e., root tip to fully mature root) and through time as roots grow and eventually decompose. Initial results suggest that the rhizosphere microbial community differs along a root age gradient and we hypothesize that these differences will be reflected in resource use measured by stable isotope probing (SIP). We have constructed a suite of stable isotope labeling growth chambers and microcosms, which we are using to pulse label *A. barbata* plants and monitor the fate of added $^{13}CO_2$ as it moves into the soil C pool both as root exudates and decomposing root litter. We are specifically interested in the differential response of rhizosphere communities exposed to either independent or simultaneous additions of ^{13}C-live roots and ^{13}C root litter.

High throughput next generation sequencing is being used to measure rhizosphere microbial community diversity (16S rRNA) as well as being combined with community separation based upon uptake of added ^{13}C labels (SIP). 16S rRNA survey also forms the basis for development of probe sets for phylogentic microarray chips to be used for stable isotope probing via Chip-SIP, a new method that combines community identification with high density RNA microarrays and substrate use profiling by NanoSIMS isotopic analysis of the array spots. We are also trialing existing probes and using the acquired 16S rRNA data to develop probes for fluorescent in-situ hybridization (FISH) to spatially examine the rhizosphere microbial community. The FISH method will be combined with NanoSIMS isotopic analysis to simultaneously observe the fate of added ^{13}C label in intact soil/root samples.

Funding for this research was provided under contract FOA DE-PS02-09ER09-25 through Genomic Sciences program of the Office of Biological and Environmental Research, U.S. Department of Energy to the University of California, Berkeley. Part of this work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344 and at Lawrence Berkeley National Laboratory under the auspices of the University of California—contract DE-AC02-05CH11231.

129

Soil Fungal Community Responses to Long-Term Elevated CO_2 and N Deposition Conditions in a Temperate Pine Forest (DOE Duke Forest FACE site)

Carolyn F. Weber,[1] Andrea Porras-Alfaro,[2] Kuan-Liang Liu,[1] Gang Xie,[1] Monica Moya Balasch,[1] La Verne Gallegos-Graves,[1] John Dunbar,[1] Rytas Vilgalys,[3] and **Cheryl R. Kuske**[1]* (kuske@lanl.gov)

[1]Los Alamos National Laboratory, Los Alamos, N.M.; [2]Western Illinois University, Macomb, Ill.; and [3]Duke University, Durham, N.C.

Project Goals: The simultaneous increase of atmospheric CO_2 and nitrogen (N) deposition to forest ecosystems is predicted to alter plant productivity and, consequently, to change the amount and quality of above and belowground carbon entering forest soils. It is not known how such changes will impact the composition and function of soil microbial communities, particularly fungal communities, that play a key role in degrading complex carbon. Our recent studies explored the composition of soil fungal biomass after ten years of elevated CO_2 conditions, and under five years of combined elevated CO_2 and increased soil N conditions, in a temperate pine forest (DOE Duke Forest FACE site). We used a combination of DNA- and RNA-based surveys and sequencing of target genes to compared features of the resident and active soil fungal community.

Ecological studies: First, we conducted a comparative study of fungal cellobiohydrolase I genes (*cbh*I), representing the resident (DNA-based) and expressed (cDNA-based) communities in surface soil (0-10 cm depth) across the elevated CO_2 and N-fertilization treatments at the Duke Forest FACE site. Our study demonstrated that the richness and composition of the soil cellulolytic fungal community was distinct between the DNA- and cDNA-based gene surveys. Richness or composition of the cellobiohydrolase-containing fungal community was not altered by elevated CO_2 and/or N-fertilization conditions relative to the ambient controls. The soil fungal community was dominated by members of the Basidiomycota that have minimal or no representation in current sequence databases.

Second, we conducted seasonal surveys of the fungal *cbhI* gene and the fungal LSU gene (phylogenetic marker) across the treatments at the Duke Forest FACE site, over two years and five seasonal time points to determine how fungal community responses to climate change parameters may vary with season. In all seasons, soil fungal community richness (LSU gene) was decreased in elevated CO_2 plots compared to ambient plots. In both ambient and elevated CO_2 plots, N fertilization increased richness in spring and summer sampling points indicating that some taxa are nitrogen limited at this site during these seasons. In contrast, we did not detect a change in *cbhI* richness that correlated with season.

Third, we conducted an intensive LSU sequencing survey of soil fungal communities present in forest floor and in underlying soil to determine the response to elevated CO_2 and N deposition treatments manifest across soil depths of a few cm (forest floor, 0-2 cm, 2-5 cm, 5-10 cm). Soil chemistry and fungal community (LSU gene) richness and composition differed significantly across shallow changes in soil depth. Fungal community response to elevated CO_2, N deposition, and the combined treatment was also highly stratified by depth, illustrating that fungal community roles in this forest ecosystem cannot be accurately predicted using bulk, homogenized samples typically employed in large field studies.

Method development and database resources: Defining the factors underpinning methodological biases is required to optimize the design of gene expression studies in soils. As a prerequisite to the ecological studies described above, we compared the richness and composition of the fungal *cbhI* gene in forest floor and underlying soil, that could be captured using two cDNA preparation methods and two different cDNA priming methods. Richness, composition and reproducibility of gene expression profiles of the fungal *cbhI* gene were examined when amplified from sscDNA, or from dscDNA synthesized using SMART PCR. In the dscDNA libraries from soil or litter samples, richness was significantly reduced and the composition was altered relative to sscDNA libraries. Library composition was significantly more reproducible among replicate sscDNA libraries than among parallel dscDNA libraries from litter. We also performed comparative richness and compositional analyses of the fungal *cbhI* gene amplified from soil cDNAs that had been generated using either oligo(dT) primers or random hexamers. Our results demonstrated that similar *cbh*I richness and composition were recovered using either priming method.

To support taxonomic interpretation of the soil fungal community sequences obtained in the ecological studies, we established a taxonomic database for the *cbhI* gene, and for the fungal LSU gene. The accuracy of the LSU database using multiple sequence lengths was determined using naïve Bayesian classifier and BLAST approaches. These classification resources are publicly available to the scientific community through the Ribosomal Database Project.

This project was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (2009LANLF260) to CRK and a Los Alamos National Laboratory Director's Postdoctoral Fellowship to CFW. Sequencing was made possible through the DOE Joint Genome Institute and the Los Alamos National Laboratory LDRD program.

* Presenting author

130

Arid Land Ecosystem Responses to Long-Term Elevated Atmospheric CO_2 in a Large Manipulated Field Experiment

Blaire Steven,[1] La Verne Gallegos-Graves,[1] Jayne Belnap,[2] R. David Evans,[3] John Dunbar,[1] and **Cheryl R. Kuske**[1]* (kuske@lanl.gov)

[1]Los Alamos National Laboratory, Los Alamos, N.M.; [2]U.S. Geological Survey, Moab, Utah; and [3]Washington State University, Pullman

Project Goals: Atmospheric CO_2 levels are expected to double within the next century. Arid land ecosystems, which comprise over 40% of the earth terrestrial surface, are predicted to be particularly vulnerable to changes in climate. Processes mediated by soil microorganisms responsible for carbon and nitrogen turnover are critical in arid land ecosystems where plants are sparse. We conducted highly replicated surveys of soil microbial communities associated with the dominant shrub species (creosote bush) and in interspace biological soil crusts (biocrusts) at the DOE Nevada Test Site Free Air CO_2 Enrichment (FACE) site to determine the effects of over ten years of elevated CO_2 conditions on the soil biota.

First, the ability of a targeted metagenomic approach (small subunit (SSU) rRNA pyrosequencing) and shotgun metagenome approaches were compared to identify known distinguishing features between the creosote root zone communities and biocrusts, and differences due to the more subtle elevated CO_2 treatment. The biocrust datasets were clearly differentiated from root zone datasets using either of the sequencing approaches. However, different compositional features were identified using the different approaches. The ability to detect possible treatment effects was largely approach-dependent, as the magnitude of resolved differences due to elevated CO_2 was smaller when shotgun metagenome reads were used, compared to pyrosequenced SSU datasets or even SSU reads recruited from the shotgun metagenomes. Based on prior knowledge of the biocrust communities, the SSU-based datasets more accurately identified the dominant biocrust cyanobacteria populations compared to the shotgun metagenome datasets.

Second, using quantitative PCR (qPCR) of cyanobacteria 16S rRNA genes, 16S rRNA pyrotag sequencing, and shotgun metagenome sequencing, we explored the response of biocrust cyanobacteria to elevated CO_2 in more detail. In years with sufficient moisture for growth, higher plants in this ecosystem have responded to elevated CO_2 with increased biomass. We hypothesized that the photosynthetic cyanobacteria that are dominant in the biocrusts would respond similarly. In contrast to our prediction, the relative abundance of cyanobacteria biomass (qPCR) was not significantly different between biocrusts under ambient or elevated CO_2 conditions, and trended toward a decrease in cyanobacteria biomass under elevated CO_2 conditions. Similarly, the proportion of cyanobacteria in 16S rRNA gene libraries or in shotgun metagenomes was either not significantly different or was reduced in biocrusts under elevated CO_2 conditions relative to the ambient controls. Comparison of the shotgun metagenomes provided information to suggest physiological and functional differences in the biocrust cyanobacteria under the elevated and ambient CO_2 conditions. Taken together these results indicate long-term elevated CO_2 produced shifts in both the structure and function of the biocrust cyanobacteria.

Employing highly replicated field sampling and DNA sequencing, we identified that a major effect of elevated atmospheric CO_2 on the biocrusts was a divergence in microbial community composition, resulting in increased spatial heterogeneity. The increased variability of the microbial communities in response to elevated CO_2 was robust using multiple sequencing approaches and was observed in both taxonomic and functional profiles. These results will inform future studies to quantify natural variability of soil microbial communities and for the design of future climate change experiments having predictive potential.

This project was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (2009LANLF260) to CRK. Sequencing was made possible through the DOE Joint Genome Institute and the Los Alamos National Laboratory LDRD program.

131

Direct Interspecies Electron Transfer in Methanogenic Environments: Genome-Scale Analysis of Mechanisms in Defined Co-Cultures and Natural Aggregates

Amelia-Elena Rotaru[1]* (arotaru@microbio.umass.edu), Pravin Mala Shresta,[1] Fanghua Liu,[1] Pier-Luc Tremblay,[1] Jessica Smith,[1] Dan Carlo Flores,[1] Mollie Murnane,[1] Harish Nagarajan,[2] Karsten Zengler,[2] and **Derek R. Lovley**[1]

[1]University of Massachusetts, Amherst and [2]University of California–San Diego, La Jolla

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 predictively 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 digestors, the most effective strategy yet devised for large-scale conversion of organic wastes to fuel. For over 40 years interspecies hydrogen transfer has served as the paradigm for anaerobic interspecies electron transfer. However, our recent studies demonstrated that direct interspecies electron transfer is possible[1] and may be the predominant mechanism for electron exchange in some methanogenic environments[2].

In order to elucidate the mechanisms of direct electron exchange under methanogenic conditions, and the factors controlling the rate of this process, studies were conducted with natural methanogenic aggregates as well as defined co-cultures. Co-cultures were established with *Geobacter metallireducens* as the electron-donating partner and either *Methanosarcina bakeria* or *Methanosaeta harudinacea* as the methanogenic partner. A *Geobacter* species was chosen because *Geobacter* species are often abundant constituents of methanogenic aggregates in anaerobic wastewater digestors as well as in methanogenic soils, such as rice paddies. *Methanosaeta* are the most abundant methanogens in similar methanogenic environments, except when they are replaced by *Methanosarcina* species.

Co-cultures of *G. metallireducens* and *Msr. barkeri* formed aggregates that effectively converted ethanol to methane. Formate could be ruled out as a potential electron shuttle between the two organisms because *Msr. barkeri* is unable to use formate as an electron donor. Although *Msr. barkeri* has the potential to use hydrogen, hydrogen did not appear to be an important intermediate for electron exchange because the cells within the aggregates were not adapted for hydrogen utilization. These results are consistent with direct interspecies electron transfer between *G. metallireducens* and *Msr. barkeri*.

The possibility of direct interspecies electron transfer is being further evaluated with studies in which on or more of the co-culture partners are strains that are deficient in components that are considered to be important for this process. For example, the PilA-pili of *Geobacter sulfurreducens* were recently shown to possess metallic-like conductivity that permits long-range electron transfer along their length and play a role in interspecies electron transfer[3]. Therefore, a strain of *G. metallireducens* that cannot produce PilA-pili was constructed and it is being determined whether this mutant can form syntrophic associations with *Msr. barkeri*. Furthermore, gene expression patterns in the two microorganisms are being compared with gene expression patterns determined in studies with previously described[1] co-cultures of *Geobacter* species conducting interspecies electron transfer.

Co-cultures of *G. metallireducens* and *Mst. harudinacea* that effectively converted ethanol to methane were also established. The ability to these two organisms to form syntrophic aggregates is significant because *Mst. harudinacea* is unable to use either hydrogen or formate as electron donors. Therefore, direct interspecies electron transfer is likely in this co-culture system as well. This is consistent with our previous finding that *Geobacter* and *Methanosaeta* species were the predominant microorganisms in natural methanogenic aggregates that were exchanging electrons via direct interspecies electron transfer[2]. Gene expression analysis of these co-cultures is being initiated and expected to be complete by the time of the meeting.

Analysis of gene expression of natural methanogenic aggregates provided additional evidence for direct interspecies electron transfer to the *Methanosaeta* species, which accounted for over 90% of the methanogens in aggregates[2]. Although *Methanosaeta* are unable to use hydrogen or formate as electron donors, a full pathway for carbon dioxide reduction could be detected in *Methanosaeta* genomes. Genes in the carbon dioxide pathway of the *Methanosaeta* in the natural aggregates were expressed at high levels. These results indicate that *Methanosaeta* receive low-potential electrons that can drive carbon dioxide reduction, consistent with the concept that *Methanosaeta* species can directly accept electrons from other members of the consortia.

As summarized in a companion poster, a genome-scale metabolic model that can describe the growth of a *G. metallireducens*/*G. sulfurreducens* co-culture that functions via direct interspecies electron transfer has been constructed. The gene expression and other physiological data needed for similar genome-scale metabolic models of methanogenic communities functioning via direct interspecies electron transfer is now being collected from the defined co-cultures and natural aggregates.

The discovery of direct interspecies electron transfer is a paradigm shift in anaerobic microbial ecology. The studies summarized here are expected to provide models that will be able to predict rates of electron exchange in methanogenic communities under a diversity of environmental conditions and the impact of environmental perturbations, such a climate change, on the rates of methanogenesis. Furthermore, an understanding of the mechanisms for direct interspecies electron transfer for methane production is leading to new concepts to either promote this process for bioenergy applications or to inhibit undesirable release of methane from terrestrial environments.

References

1. Summers ZM, Fogarty HE, Leang C, Franks AE, Malvankar NS, Lovley DR. 2010. *Science* 330:1413-5.
2. Morita M, Malvankar NS, Franks AE, Summers ZM, Giloteaux L, Rotaru AE, Rotaru C, and Lovley DR. 2011. *mBio* 2:e00159-11.
3. Malvankar NS, Vargas M, Nevin KP, Franks AE, Leang C, Kim BC, Inoue K, Mester T, Covalla SF, Johnson JP, Rotello VM, Tuominen MT and Lovley DR. 2011. *Nat Nanotechnol* 6:573-9.

This research was supported by the Office of Science (BER), U. S. Department of Energy, Award No. DE-SC0004485.

132

Integrated Genome-Scale Modeling of Syntrophic Consortia Reveals Microbial Community Dynamics and Mechanisms of Electron Transfer

Harish Nagarajan[1]* (nh@ucsd.edu), Mallory Embree,[1] Adam M. Feist,[1] Amelia Elena Rotaru,[2] Pravin Shrestha,[2] Bernhard Ø. Palsson,[1] **Derek R. Lovley**,[2] and **Karsten Zengler**[1]

[1]Systems Biology Research Group, Department of Bioengineering, University of California–San Diego, La Jolla and [2]Department of Microbiology, University of Massachusetts, Amherst

http://geobacter.org

Project Goals: The long-term goal of our project, which is entitled "Systems Level Analysis of the Function and Adaptive Responses of Methanogenic Consortia", is to develop genome-scale metabolic models of microbial communities that play an important role in the global carbon cycle that can be coupled with the appropriate physical-chemical models to predict how the microbial communities will respond to environmental perturbations, such as climate change. The short-term objective of the current research is to determine if the syntrophic associations that are central to the functioning of methanogenic terrestrial wetlands can be predictively modeled with genome-scale metabolic models.

The previously described[1] syntrophic co-culture of *Geobacter metallireducens* and *Geobacter sulfurreducens* serves as a genetically tractable model for the process of direct interspecies electron transfer hypothesized to be important in terrestrial methanogenic environments. Therefore, this co-culture is serving as an initial test case to determine if it is possible to use genome-scale metabolic models to predict the physiological responses of microorganisms engaging in direct interspecies electron transfer in complex communities and to characterize the relative efficiencies of direct electron exchange versus other alternatives, such as interspecies hydrogen or formate transfer.

In order to model different modes of electron transfer, it is important to accurately account for the energetics involved in extracellular electron transport. Hence, the existing genome-scale reconstructions of *G. sulfurreducens* and *G. metallireducens* were expanded to reflect the most updated annotations, a distinct periplasm compartment, and detailed biosynthetic pathways. Importantly, both these models now include detailed representation of the energy metabolism involved in extracellular electron transfer. The models now account 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 thermodynamic consistency, appropriate gene association and cellular localization of the different electron carriers. The *G. sulfurreducens* model now consists of 829 genes and 1079 reactions, while the *G. metallireducens* model consists of 974 genes and 1173 reactions.

We further developed a modeling framework to integrate these two genome-scale models into a combined model to study the metabolic interactions. This framework includes a shared metabolite pool to account for the metabolic exchanges between the constituents of the consortia. Computational simulations revealed that the optimal ratio of the constituents contained 21% *G. metallireducens*. This prediction is in accordance with experimental observations of a composition of 15% *G. metallireducens*.

Flux balance analysis simulations indicated that direct interspecies electron transfer is more efficient than interspecies hydrogen transfer for the growth of the consortia. This prediction is consistent with the previous observation[1] that selective pressure for rapid syntrophic growth selected for a mutation that promoted direct electron transfer. Acetate secreted by *G. metallireducens* was used as a carbon source as well as an additional source of electrons for *G. sulfurreducens*.

We also modeled a newly developed co-culture in which the *G. sulfurreducens* strain was incapable of acetate oxidation because the citrate synthase gene was deleted. This eliminated additional grow of *G. sulfurreducens* with acetate as the electron donor. The optimal ratio of the constituents in this co-culture was predicted to be 45% *G. metallireducens*, which compared well with the experimental observation of 50%. Again, direct electron transfer was predicted to be preferred over interspecies hydrogen transfer.

We further performed high-throughput transcriptomic profiling (RNA-seq) and physiological screens of both the wild type and mutant aggregates and analyzed them in the context of the metabolic model to gain insights into the mechanisms of electron transfer. Differential expression analysis revealed that 481 genes had significant changes in expression levels between wild-type and citrate synthase mutants. The global effects of these gene expression changes on the dynamics of the microbial community are being investigated with the aid of the integrated community model. We have also performed genome re-sequencing of the aggregates and identified SNPs that have accumulated as a result of the selective pressure of syntrophy.

In summary, this study represents an integrated multi-omic approach to elucidate the electron transfer mechanism and to characterize the metabolic phenotype of a laboratory evolved syntrophic consortium. The models being developed in these studies will be important for understanding the functioning of anaerobic terrestrial microbial communities and predicting the influence of environmental changes on methane emissions and other aspects of carbon cycling.

Reference

1. Summers, Z. M., H. E. Fogarty, C. Leang, A. E. Franks, N. S. Malvankar, and D. R. Lovley. 2010. Direct exchange of electrons within aggregates of an evolved syntrophic coculture of anaerobic bacteria. Science 330:1413-1415

* Presenting author

This work was funded by Office of Science (BER), U.S. Department of Energy (DE-SC0004485).

133

Systems-Level Insights into Alternate Methane Cycling Modes in a Freshwater Lake via Community Transcriptomics, Metabolomics, and nanoSIMS Analysis

M.G. Kalyuzhnaya,[1] S. Yang,[1] D.A.C. Beck,[1] J. Bickford,[1] M.C. Konopka, L. Chistoserdova,[1] V.J. Orphan,[2] and **M.E. Lidstrom**[1]* (lidstrom@u.washington.edu)

[1]University of Washington, Seattle and [2]California Institute of Technology, Pasadena

Project Goals: The 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-dependent versus oxygen-dependent modes, in order to bridge gaps in understanding the specialized bacterial communities involved in these processes.

The 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.

Our research was focused on the following three objectives:

Objective 1: Identify actively transcribed pathways. Next generation sequencing-based transcriptomic profiling (RNA-seq platform) was used to perform global characterization of C_1-metabolism in lake sediment, as well as community responses to stimulated environmental perturbation, such as low oxygen and additional nitrate**.** Reconstruction of the related C_1-metabolic function was performed. The results demonstrated that the sediment C_1- community is represented by microbial species adapted to low oxygen and low methane flux.

Objective 2: Identify physiologically active pathways. A highly comprehensive ion exchange solid phase extraction (SPE) liked with hydrophilic interaction liquid chromatography (HILIC-MS/MS) and liquid chromatography with a pentafluorophenylpropyl column (LCPFPP-MS/MS) was developed to understandcentral carbon metabolites involved in methane assimilation and adjacent pathways in complex natural mixtures. The recovery of 51 targeted metabolites from five compound classes (amino acids, carboxylic acids, sugar phosphates, nucleotides, and acyl-CoAs) was investigated. The following SPE procedures were employed: (a) mixed mode strong cation exchange, (b) mixed mode strong anion exchange, and (c) mixed mode weak anion exchange. We analyzed 32 of the targeted 51 metabolites using either HILIC-MS/MS or LCPFPP-MS/MS after SPE sediment samples cleanup and pre-concentration. The remaining 19 targeted metabolites were either at, or below, the detection limit. The current approach provides a good workflow for absolute quantification of intermediates in C_1-carbon metabolismin naturalmicrobial communities.

Objective 3: Identify activity of individual cells. As part of Objective 3, 10 individual active cells with unusual phylogeny were selected for whole genome amplification, sequencing, assembly, and functional characterization as a result of metabolic reconstruction. The result couples genomics to function, and therefore, has the potential to reveal newinsights into the physiological capabilities of yet uncultured members of natural microbial communities.

This work was supported by the DOE (DE-SC0005154).

134

Vitamin Biosynthesis and Regulation in Marine Algae

Darcy McRose,[1]* Shuangchun Yan,[1] Emily Reistetter,[1] and **Alexandra Z. Worden**[1,2] (azworden@mbari.org)
PI: Alexandra Z. Worden, Monterey Bay Aquarium Research Institute and UCSC
Co-PIs: Stephen Callister, Richard Smith, Joshua Stuart

[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 use it to investigate gene function, pathways and consequences of environmental perturbations on primary production.

Phytoplankton are important primary producers in marine environments. While the regulatory role of macronutrients on algal growth has been clearly demonstrated, the potential influence of vitamins and small metabolites remains largely unknown. Several groups of organisms have developed independence from vitamin B12, however vitamin B1 (thiamine) is required by all known members of the three domains of life. Still, it is unclear to what extent ecologically relevant marine algae require exogenous sources of B1 and which taxa can synthesize this vitamin. Furthermore, there is little information about the algal genes involved in thiamine biosynthesis, which limits hypothesis development and experimentation.

Here, we present an analysis of the thiamine biosynthetic pathway in eukaryotic algae, with targeted experimental work in the marine green alga *Micromonas*. To explore the evolutionary history of thiamine biosynthesis genes and elucidate which marine algae may have the genetic potential to synthesize B1 we performed a comparative genomics analysis of thiamine biosynthetic genes encoded by algae from different eukaryotic supergroups and by cyanobacteria. These genes were compared to those in bacteria, yeast and land plants. The analysis revealed that green algae are quite different from other eukaryotic algae (e.g. diatoms and pelagophytes), and that even within the green algae there is significant variation. For example, within the *Micromonas* genus, one strain (RCC299) appears to lack key thiamine biosynthetic enzymes, and therefore presumably requires an exogenous source, while another strain (CCMP1545) appears to encode a more complete pathway. In addition, at least one thiamine biosynthetic gene, found in *Micromonas*, and other green algae, is more akin to homologs in fungi than to the gene that performs the same biosynthetic step in land plants.

Because *Micromonas* has a widespread distribution in the ocean, we examined several of the identified putative thiamine biosynthesis genes experimentally. In addition to sampling for complete transcriptome sequencing, qPCR primers were designed to a suite of genes in each strain. Experiments were conducted to explore growth responses to thiamine deprivation in the two *Micromonas* strains. When cultured without thiamine, *Micromonas* CCMP1545 up-regulated key thiamine biosynthetic genes however neither strain seemed able to grow successfully in the absence of thiamine. qPCR primers were also designed to genes that have no known role in thiamine biosynthesis, but, based on genomic analysis, appear to be under the control of a thiamine sensing molecular switch. *Micromonas* RCC299 up-regulated these genes in response to thiamine deprivation and we hypothesize they represent previously unrecognized thiamine transporters. The disparity in expression responses between the two *Micromonas* as well as the unusual pattern of gene homology indicate that thiamine plays a role in algal physiology and ecology. However, experimental validation is critical for developing true understanding of ecological controls and niche differentiation.

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

135

Computational and Experimental Approaches to Systems Biology of Marine Green Algae

Marijke J. van Baren[1]*, Peter Waltman,[2] Daniel Sam,[2] Emily Reistetter,[1] Samuel Purvine,[3] Charles Ansong,[3] Stephen Callister,[3] Joshua Stuart,[2] and **Alexandra Z. Worden**[1,4] (azworden@mbari.org)
PI: Alexandra Z. Worden, Monterey Bay Aquarium Research Institute and UCSC
Co-PIs: Stephen Callister, Richard Smith, Joshua Stuart

[1]Monterey Bay Aquarium Research Institute, Moss Landing, Calif.; [2]Biomolecular Engineering Department, University of California, Santa Cruz; [3]Pacific Northwest National Laboratory, Richland, Wash.; and [4]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 use it to investigate gene function, pathways and consequences of environmental perturbations on primary production.

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

Prasinophytes are a group of unicellular marine green algae that are evolutionarily distinct from the model green alga *Chlamydomonas*, but are related to both the latter and land plants. *Micromonas* is a widespread prasinophyte that is exceptional in its size (<2 micrometer diameter) and having a small genome (21 Mb). The genomes of two *Micromonas* strains share 90% or less (depending on criteria used) of their protein encoding genes, have low gene redundancy and contain 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. Our strategy is to subject the two different strains to ecologically relevant perturbations and use whole transcriptome and proteome profiling, as well as traditional cellular measures, to understand their responses to perturbation and develop knowledge of cellular pathways. The genomes of *Arabidopsis thaliana*, *Physcomitrella patens*, *Chlamydomonas reinhardtii*, *Ostreococcus* RCC809, *Saccharomyces cerevisiae* and *Escherichia coli* are being used for comparative purposes.

To investigate genetic pathways in *Micromonas*, a mixed computational inference strategy is being utilized that integrates data from diverse sources, including the transcriptome and proteome expression. As a first step, we have implemented a scoring system to track supporting evidence for all predicted genes in the two strains. Every possible annotation is scored, in a database that holds all annotations and their evidence, and the highest scoring gene at a locus then selected automatically. We are currently expanding the database to hold all biologically relevant information for different experimental conditions. This database provides a foundation for pathway and perturbation analyses. Interaction orthologs (interologs) are included to help identify putative interaction pathways. An integrative multi-species biclustering algorithm is also being used to identify putatively conserved modules that are shared by the two *Micromonas* species, as well as species-specific differences in these modules and modules that are unique to each organism. In turn, these modules were used in conjunction with several network-inference methods to identify additional putative pathways whose activity levels were identified using PARADIGM, a state-of-the-art pathway prediction algorithm. Given its relationship to plants the dual-*Micromonas* system will enable modeling of more general primary producer responses across ecological and evolutionary scales. Our overall goal is to develop an efficient system for gaining insights to the green lineage, including novel and conserved genetic mechanisms, with high relevance to marine carbon cycling. 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.

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

Early Career Award

136

Linking Phylogeny and Function of Methanotrophic Archaeal-Bacterial Consortia in Deep-Sea Methane Seeps Using *in situ* Targeted Metagenomics, Molecular Ecology, and Stable Isotope Tracer Experiments

Joshua A. Steele[1]* (jsteele@gps.caltech.edu), Rachel S. Poretsky,[1] Anne E. Dekas,[1] Shulei Sun,[2] Abigail Saxena,[1] Patricia L. Tavormina,[1] Mark H. Ellisman,[2] and **Victoria J. Orphan**[1]

[1]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena and [2]Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis, University of California, San Diego

http://www.gps.caltech.edu/people/vorphan/research

Project Goals: We are exploring the biogeochemical functions and ecological relationships between ANaerobic MEthane oxidizing Archaea (ANME) and Sulfate Reducing Bacteria (SRB) and associated prokaryotes. We have pursued the following approaches: 1) Magneto-FISH to isolate the ANME-SRB consortia, 2) molecular investigation of phylogeny and function of the ANME-SRB metagenomes and the *in situ* community, 3) stable isotope tracer experiments plus single-cell FISH-nanoSIMS analysis to link organisms with function, and 4) bulk geochemical measurements to observe biogeochemical significance. Targeting ANME-SRB metagenomes has allowed comparison with other metagenomes and supported the reverse-methanogenesis hypothesis for methane oxidation. *In situ* measurements and stable isotope tracer experiments enabled testing the functionality of genes identified by metagenomic analysis within the ANME-SRB consortia, and revealed an unexpected role for these organisms in the nitrogen cycle. Although valuable in isolation, combining these techniques yields new insights into these uncultured methane-based syntrophic partnerships.

Targeting the uncultivable ANME-2c and their bacterial consortia in methane seep sediment from Eel River Basin, we have built on the Magneto-FISH technique pioneered by our lab group (Pernthaler et al 2008). We have generated and analyzed more than 10 times the sequence from our initial effort and, using the Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA) pipeline, identified a greater number of ORFs and gene/protein families. Due to the unique environment of the methane seep, the proportion of sequence assigned to COG or pfam database sequences has remained low (e.g. ~8% currently compared to 4% previously); also, when compared to the NCBI RefSeq database only 50-70% of ORFs in assembled contigs were similar to any protein sequences of characterized organisms. Similar to what we found in earlier targeted metagenomes, there was a wide variety of ANME-2c-associated bacteria, dominated by δ- and γ-*Proteobacteria*. δ-*Proteobacteria*-derived sequences were primarily affiliated with *Desulfobacterales*, *Desulfuro-monadales*, and *Syntrophobacterales*. Many different γ-*Proteobacteria* were present with no single group predominating. Archaeal sequences made up a smaller percentage of the metagenome and were mostly from the *Methanosarcinales, Methanomicrobiales*, and *Methanocellales* and other methanogenic groups. Using 16S rRNA gene quantitative PCR, we were able to confirm our ability to enrich for specific syntrophic partners with Magneto-FISH, with genomic DNA extracts from the *Desulfobulbus*-targeted magneto-FISH revealing a greater proportion of *Desulfobulbus* 16S rRNA genes compared with the original sediment.

To generate better contig assembly and to overcome limitations in annotation we have combined our individual ANME-2c targeted metagenome libraries from a single sample in the Eel River Basin into a single contig assembly. This combined metagenome was compared to publicly available metagenomes from methane seep sediments within the Integrated Microbial Genome (IMG) database and to recently published genomes of ANME-1, *Desulfosarcina*, and *Desulfobulbus*. Recruiting our contigs to previously published methane seep fosmid libraries, revealed up to 30% coverage of ANME-2 and SRB fosmids and little to no recruitment to taxonomically identified ANME-1 fosmids. All of the genes necessary for the reverse methanogenesis pathway are present in the ANME-2c enriched metagenome, confirming earlier findings. Genes for carbon fixation, sulfate reduction, nitrate reduction, and nitrogen fixation are also present. Grouping the reads and contigs using sequence-based categorization (e.g. tetranucleotide correlation and self organizing map algorithms) will increase confidence in these characterizations. Further association of identity and function using 16S rRNA genes and other housekeeping genes (e.g. RecA/RadA, RecG, leuS etc.) correlated to functional genes involved in sulfate reduction (aprBA, dsrAB), nitrogen metabolism (nifHDK, nirK, napG, etc.) and methanotrophy (e.g. mcrA) in the ANME-

2c enriched metagenome and in other methane seep habitats. Combining these techniques should provide phylogenetic characterization that is less dependent on databases from cultivated organisms.

We leveraged information from the ANME-2c metagenome to show ANME archaea are involved in N_2 fixation at methane seeps. We initially identified the potential for ANME related *nif* genes in the metagenome (Pernthaler et al. 2008) and bolstered this with a survey of the *nifH* genes in the ANME-2c enriched sample and the microbial community at the Eel River Basin (ERB) methane seep. We showed direct incorporation of $^{15}N_2$ into ANME-2 archaea collected at the ERB using FISH-nanoSIMS (Dekas et al. 2009). We have expanded our study to other methane seeps (Costa Rica; Hydrate Ridge, USA; Monterey Canyon, USA) and to potential bacterial diazotrophs. Nitrogen fixation rates are highly spatially variable at methane seeps, both laterally and with sediment depth, with the highest rates observed within the methane-sulfate transition zone. Quantitative comparison between the EA-irMS bulk rates and the NanoSIMS single-cell rates of nitrogen fixation suggest that the ANME are responsible for the majority (~80%) of seep nitrogen fixation, but that the remainder is fixed by other diazotrophic organisms, likely including free-living sulfate reducing bacteria.

Publications

1. Pernthaler, A., Dekas, A. E., Brown, C. T., Goffredi, S. K., Embaye, T., Orphan, V.J. (2008) Diverse syntrophic partnerships from deep-sea methane vents revealed by direct cell capture and metagenomics. *PNAS* 105(19) 7052-7057
2. Dekas, A.E., Poretsky, R.S., and Orphan, V.J. (2009) Deep-Sea Archaea Fix and Share Nitrogen in Methane-Consuming Microbial Consortia. *Science* 326: 422-426

This research was funded by a DOE Office of Biological and Environmental Research Early Career Grant to V. Orphan.

137

Correlative Compositional Imaging and Protein Profiling in Marine Anaerobic Methane Oxidizing Microbial Communities

Shawn E. McGlynn[1]* (smcglynn@caltech.edu), Jeff J. Marlow,[1] Grayson Chadwick,[1] Tomas J. Deerinck,[2] Shulei Sun,[2] Karuna Chourey,[3] Chongle Pan,[3] Robert L. Hettich,[3] Mark H. Ellisman,[2] and **Victoria J. Orphan**[1]

[1]California Institute of Technology, Pasadena, Calif.; [2]University of California–San Diego, La Jolla; and [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://www.gps.caltech.edu/people/vorphan/research

Project Goals: Multiple experimental lines of inquiry are being applied in an effort to shed light on the metabolic activity and genomic capacity of organisms responsible for the anaerobic oxidation of methane. Characterization by secondary ion mass spectrometry, transmission electron microscopy, and fluorescence *in situ* hybridization reveals organism specific composition and structure, while metaproteomics and histological approaches reveal biochemical functionality and distribution within these uncultured multicellular consortia.

Large amounts of methane are oxidized in anoxic marine sediments by symbiotic microorganisms whose physiology is largely unknown. These uncultured organisms, comprised of consortia of methane-oxidizing archaea (ANME) and sulfate reducing bacteria (SRB), form highly structured physical associations and participate in a syntrophic relationship. To characterize these metabolic partnerships, complimentary microscopic and spectrometric techniques are being developed, providing high resolution compositional and phylogenetic information. These analyses indicate significant differences between cell types and create the opportunity for conducting comparative and quantitative image analysis between samples experiencing shifts in environmental conditions. Integration of these approaches with metagenomics and metaproteomics has the potential to result in direct correlation of form and function in these organisms.

Methodologies appropriate for isolating high concentrations of biological material from environmental samples have been recently developed by our team, making possible the application of high resolution imaging techniques including electron microscopy and spatial mass spectrometry. These techniques have proven suitable for the analysis of microbial consortia with a diversity of imaging methodologies. Characterization of ANME/SRB consortia by transmission electron microscopy (TEM) reveals a diversity of cellular morphologies and intercellular components. Many of these intercellular features can be understood partially from comparison with those present in organisms in pure culture, however other features appear to be novel and undescribed. In addition, TEM analysis reveals the presence of an inorganic iron-bearing phylosilicate crust which surrounds some, but not all of the aggregate types. Material composition of these features, as well as compositional differences between organism types are being probed by secondary ion mass spectrometry (nanoSIMS) at spatial resolutions previously not possible. These investigations indicate the presence of internal polyphosphate storage granules in ANME cells, and numerous storage granules with unknown composition in the associated SRB partner. In addition, high resolution imaging reveals close physical contact between cells of different types, underscoring the possibility of metabolic coupling between them. Concomitant to these imaging techniques, taxonomic identification of individual microorganisms in prepared thin sections of the consortia is made possible by employing florescence *in situ* hybridization (FISH). The pairing of these techniques gives the ability to study these highly structured consortia at unprecedented spatial resolution and deliver information as to individual organism roles within this complex ecosystem.

As an extension to these studies, histological approaches are being developed to query the presence of functional gene products. The enzymes methyl-coenzyme M reductase

* Presenting author

(Mcr) as well as nitrogenase (Nif) are being targeted to probe enzyme distribution within microbial consortia. These enzymes represent carbon and nitrogen entry points into the metabolism of ANME/SRB consortia and their identification holds the potential to directly reveal specialization and resource partitioning at the cellular level.

In tandem to conducting cell specific image based analyses, environmental metaproteomic data from methane-seep sediments and microcosm experiments containing ANME/SRB consortia are being used to survey the distribution of expressed proteins. Preliminary proteomic profiling of methane seep sediments reveal many enzymes known to be involved in methane oxidation and sulfur reduction pathways. These analyses indicate that it is possible to extract a significant number of proteins from environmental samples, and furthermore that these proteins can be identified using a locally created metagenomic database. We have found that the efficiency of protein extraction from soil microbes is influenced strongly by the type of soil matrix in which they reside and early results show that protein extraction is inhibited by the presence of clays. These findings will help tailor microcosm setups for in-depth exploration and validation of partnership between ANME and SRB consortia in deep sea methane seep sediments. The ability to profile proteomes from samples of variable environments allows for the identification of gene products of known and unknown function, and will thus be of value in identifying the metabolic modes utilized in the anaerobic oxidation of methane.

Collectively, these investigations indicate that many of the processes occurring in anaerobic methane oxidizing microbial communities are discernable utilizing culture independent techniques and aid in understanding the role of these organisms in global elemental cycling.

Funding Opportunity Announcement Number: DE-PS02-09ER09-25
Supported by the Office of Biological and Environmental Research.

138

How Monomer Availability Impacts the Scramble for Carbon During Plant Polymer Decomposition: Using Pure Cultures as a Predictive Model System

Shifang Hsu, Nivedita Clark, and **Christopher B. Blackwood*** (cblackwo@kent.edu)

Department of Biological Sciences, Kent State University, Kent, Ohio

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

Project Goals: Microbes are responsible for decomposing plant materials in terrestrial systems, which has an important impact on carbon sequestration. During the decomposition of plant materials, microorganisms take up monomers for use in respiration and biosynthesis. Plant cell wall polymers, such as cellulose, pectin and lignin, are degraded into monomers by microorganisms capable of producing extracellular enzymes. However, production of extracellular enzymes is an energy-intensive process. Therefore, our project goal is to describe in detail how monomer availability affects microbial physiology and gene transcription. Furthermore, from this information we will predict the identity and importance of "cheating" organisms during decomposition. In this project, we categorize microorganisms involved in decomposition as 1) Investors, the microorganisms that produce extracellular enzymes to break down complex plant material into simple molecules; 2) Obligate cheaters, who do not carry genes of extracellular enzymes but who take up monomers released from plant polymers; 3) Opportunistic cheaters, who do have genes of extracellular enzymes, but suppress expression of these enzymes and still take up degradation products.

In order to understand the interaction between decomposer activity and carbon supply during the decomposition process, we grew *Talaromyces stipitatus* NRRL 1006, a potential investor/opportunistic cheater, in sand microcosms with cellulose as the sole carbon source. We amended microcosms with monomers in two experiments to monitor shifts in gene expression and microbial physiology in response to different amounts and types of labile carbon monomers. Here we describe patterns in respiration; responses in extracellular enzyme activity, biomass, and gene expression will also be presented. In the first experiment, xylose was amended at four concentrations ranging from 10 mM to 120 mM in order to examine concentration thresholds resulting in a physiological and transcriptomic response to monomers. Respiration responded to all xylose amendments, but was highest for the 60 and 120 mM amendments. In the second experiment, we amended microcosms with 60 mM solutions of glucose, xylose, galacturonic acid, and vanillin to determine the effect of switching from growth on cellulose and corresponding glucose monomers, to other monomer types. Surprisingly, respiration increased the most due to xylose amendment, followed by glucose and galacturonic acid. Respiration was suppressed by vanillin.

These experiments will be repeated with 13 additional model fungal and bacterial species with completed genomic sequences. These organisms include potential investors and cheaters. In initial experiments, the respiration rate of investors growing on cellulose in sand microcosms showed periodic increases and decreases during 20 days incubation, whereas the respiration rate of cheaters peaked quickly and then went flat. Small subunit RNA copy numbers of cheaters and potential investors continued to shift following leveling off of respiration rate.

Combining data from these experiments, it is hypothesized that "cheaters" will be identifiable as the organisms with the highest monomer concentration thresholds resulting in suppression of extracellular enzyme genes.

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

139

Methanogenic Archaea and the Global Carbon Cycle: A Systems Biology Approach to the Study of *Methanosarcina* Species

Matthew N. Benedict,[1,2] Matthew C. Gonnerman,[1,2] Ping Ma,[1] Rachel J. Whitaker,[1] Nathan D. Price[1,2]* (nprice@systemsbiology.org), and **William W. Metcalf**[1]

[1]University of Illinois, Urbana and [2]Institute for Systems Biology, Seattle, Wash.

http://mcb.illinois.edu/faculty/profile/1191
price.systemsbiology.net

Project Goals: 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. Strains from additional well-studied and important ecosystems will be isolated and examined at later stages of the project. 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.

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. One of the initial successes of the first year of funding was to develop a genome-scale model of *Methanosarcina acetivorans* metabolism.

Methanosarcina acetivorans strain C2A is a marine methanogenic archaeon notable for its substrate utilization, genetic tractability, and novel energy conservation mechanisms. To help probe the phenotypic implications of this organism's unique metabolism, we have constructed and manually curated a genome-scale metabolic model of *M. acetivorans*, iMB745, which accounts for 745 of the 4540 predicted protein coding genes (16%) in the *M. acetivorans* genome. The reconstruction effort has identified key knowledge gaps and differences in peripheral and central metabolism between methanogenic species. Using flux balance analysis, the model quantitatively predicts wild type phenotypes and is 96% accurate in knockout lethality predictions compared to currently available experimental data. The model was used to probe the mechanisms and energetics of byproduct formation and growth on carbon monoxide, and the nature of the reaction catalyzed by the soluble heterodisulfide reductase HdrABC in *M. acetivorans*. The genome-scale model provides quantitative and qualitative hypotheses that can be used to help iteratively guide additional experiments to further the state of knowledge about methanogenesis.

Reference

1. Benedict, M.N, Gonnerman, M.C, Metcalf, W.W., and Price, N.D., 2011. Genome-scale metabolic reconstruction and hypothesis tesing in the methanogenic archaeon *Methanosarcina acetivorans* C2A. *Journal of Bacteriology*. Epub ahead of print.

This work was funded by the DOE Office of Science Genomes to Life Program (# ER64999).

140

Identification of S-layer Proteins in the Methanosarcinaceae

R.P. Gunsalus[1,2]* (robg@microbio.ucla.edu), D. Francoleon,[3] R. Loo,[3] J. Loo,[1,3] L. Rohlin,[1,2] U. Kim,[2] and M. Arbing[1]

[1]UCLA-DOE Institute of Genomics and Proteomics, [2]Department of Microbiology, Immunology, and Molecular Genetics, and [3]Department of Chemistry and Biochemistry, University of California, Los Angeles

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

Project Goals: One major goal of this collaborative project addresses the identity, structure and function of archaea envelopes that provide cell protection from environmental challenges. Our model organisms include key methanogenic species involved in anaerobic carbon cycling and methane production. An understanding of their ability to adapt and thrive in changing habitats impacts fundamental processes of microbial biomass transformations, CO_2 sequestration, and energy generation by anaerobic microorganisms.

The cell envelopes of many archaeal species[1] have a proteinaceous surface or lattice termed the surface-layer (S-layer). It is typically composed of only one or two abundant, often post-translationally modified proteins that self-assemble to form a highly organized surface-exposed array. Currently, very little is known about the properties of such surface arrays in any archaean. Surprisingly, over a hundred proteins were annotated to be S-layer or surface associated components in the *Methanosarcina mazei, Methanosarncina acetivorans,* and *Methanosarcina barkeri* genomes, reflecting limitations of current bioinformatics predictions[2,3]. To experimentally address what proteins are present, we devised an *in vivo* biotinylation technique to affinity tag all surface-exposed proteins that overcame challenges in working with these fragile microorganisms. The *Methanosarcina* species were adapted to growth under N_2 fixing conditions to minimize the level of free amines that would interfere with the NHS-label acylation chemistry used[4]. A

3-phase separation procedure was then employed to isolate the intact labeled cells from any lysed-cell derived proteins. The Streptavidin affinity enrichment was followed by stringent wash to remove non-specifically bound proteins, and LC-MS-MS methods were employed to identify the labeled surface proteins. The major surface layer protein was identified in all three species to belong to a small highly conserved group of hypothetical proteins. They were shown to be present in multiple glycosylated forms by using SDS-PAGE coupled with glycoprotein-specific staining, and by interaction with the lectin, Concanavalin A. This family of related S-layer proteins/genes identified in all the sequenced *Methanosarcina* genomes exhibited similar features including a signal P sequence, tandem DUF1608 domains, and a C-terminal hydrophobic transmembrane helix. To address S-layer structure and function, crystallographic studies were performed whereby the *M. acetivorans* S-layer protein DUF1608 domain structure was determined at 2.3 A. This structure provides a model for S-layer protein assembly onto the cell surface to form a lattice. Finally, several pore types were revealed that would allow for movement of small molecules to the cytoplasmic membrane. In conclusion, these studies reveal a conserved protein signature within the *Methanosarcinaceae* having distinct protein features and implied architecture that is absent in other archaea.

References

1. Albers, S-J., and B.H. Meuer. 2011. The Archaeal Cell Envelope. Nat Microbiol. Rev. 9:414.
2. Galagan JE, Nusbaum C, Roy A, Endrizzi MG, Macdonald P, FitzHugh W, Calvo S, Engels R, Smirnov S, Atnoor D *et al*: 2002. The genome of *Methanosarcina acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res* 12:532-542.
3. Maeder DL, Anderson I, Brettin TS, Bruce DC, Gilna P, Han CS, Lapidus A, Metcalf WW, Saunders E, Tapia R, Sowers KR. 2006. The *Methanosarcina barkeri* Genome: Comparative Analysis with *Methanosarcina acetivorans* and *Methanosarcina mazei* Reveals Extensive Rearrangement within Methanosarcinal Genomes. J Bacteriol. 188:7922-31.
4. Francoleon, D.R., P. Boontheung, Y. Yang, U-M. Kim, P. Denny, J.A. Loo, R.P. Gunsalus, and R.R. Ogorzalek Loo. 2009. S-layer, surface-accessible, and Concanavalin-A binding proteins of *Methanosarcina acetivorans* and *Methanosarcina mazei*. J. Proteome Research. 8:1972-1982.

This project was supported in part by the Department of Energy Grant DE-FC02-02ER63421 to UCLA and the Department of Energy Biosciences Division grant DE-FG02-08ER64689.

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

141

Influence of Habitat on Diversity and Evolution of Surface Ocean Microbes

Andrew E. Allen[1]* (aallen@jcvi.org), Lisa Z. Allen,[1,2] Jonathan H. Badger,[1] Christopher L. Dupont,[1] John P. McCrow,[1] Ian T. Paulsen,[4] Mathangi Thiagarajan,[5] Doug B. Rusch,[5] Jed Fuhrman,[6] Kenneth H. Nealson,[1] Shibu Yooseph, and **J. Craig Venter**[1]

[1]Microbial and Environmental Genomics, J. Craig Venter Institute, San Diego, Calif.; [2]Scripps Institution of Oceanography, University of California-San Diego, La Jolla; [3]Division of Biological Sciences, University of California-San Diego, La Jolla; [4]Macquarie University, Sydney, NSW, Australia; [5]Informatics, J. Craig Venter Institute, Rockville, Md.; and [6]Marine and Environmental Biology, University of Southern California, Los Angeles

Project Goals: The development of computational tools to assess and compare marine metagenomic samples on the basis of phylogenetic and functional diversity, and to examine genome level adaptations to resource availability in different size classes of marine bacteria. The Global Ocean Survey (GOS) continues to sample and analyze marine microbial life around the world from the Sorcerer II research vessel. Comparisons of microbial populations in the Indian Ocean to those found during GOS phase I have resulted in a better understanding of variation in ecological niches available in marine surface waters.

In open ocean pelagic ecosystems bacterial cells account for approximately half of the organic biomass and bacterial activity is responsible for consumption of a large fraction of photosynthetically-derived carbon. Genomic adaptations that underlie bacterial physiological adaptations to environmental gradients such as latitude, temperature, productivity, and particle association have not been clearly elucidated on a large scale. Through analyses of metagenomic data derived from size-fractionated bacterial communities, large scale phylogenomics, and genome size-normalized analyses of genomic contents we elucidate several fundamental adaptations to environmental gradients in the surface ocean. Our results show that marine bacterioplankton in the surface ocean adapt to environmental variability in three ways; molecular-level alterations, changes in overall genomic content, and transitions in community structure. We report on the influence of habitat on genomic properties such as estimated genome size, gene family composition, transporter repertoire, and carbon to nitrogen ratio of the predicted proteome. Small free-living picoplanktonic open-ocean bacterioplankton compared to coastal, larger size-class, or particle associated bacteria have smaller genomes that are enriched for transporters, depleted in regulatory components and encode a proteome with relatively less nitrogen content. Overall patterns of diversity and richness are significantly higher for bacterial communities collected on larger pore-size filters. Intra- and inter-relationships between sites based on diversity measures suggest that larger size class or particle-associated bacterial communities and those that inhabit more productive waters are significantly more variable.

This work was funded by the Office of Biological and Environmental Research in the DOE Office of Science (DE-FC02-02ER63453)

142

Influence of Nutrients and Currents on the Genomic Composition of Microbes across an Upwelling Mosaic

Lisa Zeigler Allen,[1,2] Andrew E. Allen[1]* (aallen@jcvi.org), Eric E. Allen,[2,3] Jonathan H. Badger,[1] John P. McCrow,[1] Ian T. Paulsen,[4] Liam D.H. Elbourne,[4] Mathangi Thiagarajan,[5] Doug B. Rusch,[5] Kenneth H. Nealson,[1] Shannon J. Williamson,[1] and **J. Craig Venter**[1]

[1]Microbial and Environmental Genomics, J. Craig Venter Institute, San Diego, Calif.; [2]Scripps Institution of Oceanography, University of California-San Diego, La Jolla; [3]Division of Biological Sciences, University of California-San Diego, La Jolla; [4]Macquarie University, Sydney, NSW, Australia; and [5]Informatics, J. Craig Venter Institute, Rockville, Md.

Project Goals: Our basic research program focused on acquisition and analysis of marine microbial metagenomic data and development of genomic analysis tools for broad, external community use. Our Marine Metagenomic Diversity effort generated and analyzed shotgun sequencing data from microbial communities sampled from >260 sites around the world. Starting in the summer 2007 and continuing until the winter of 2008, we initiated a series of sampling efforts focused on the Southern California Bight and California Current Ecosystem in collaboration with several groups. In the summer of 2007, this cruise was done in collaboration with the California Cooperative Oceanic Fisheries Investigations (CalCOFI) group to evaluate metagenomics at select areas characterized by coastal upwelling. These samples are a contrast to the primarily open ocean sites of the GOS expedition due to the upwelling of nutrient rich deep waters that enable an increase in primary production at the surface.

Metagenomic datasets were generated from samples collected along a coastal to open ocean transect between Southern California Bight and California Current waters during a seasonal upwelling event, providing an opportunity to examine the impact of episodic pulses of cold nutrient rich water into surface ocean microbial communities. The dataset consists of approximately 5.8 million predicted proteins across seven sites, from 3 different size classes: 0.1-0.8μm, 0.8-3.0μm, and 3.0-200.0μm. Taxonomic and metabolic analyses suggest that sequences from the 0.1-0.8μm size class correlated with their position along the upwelling mosaic. However, taxonomic profiles of bacteria from the larger size classes (0.8-200μm) were less constrained by habitat and characterized by an increase in Cyanobacteria, Bacteroidetes, Flavobacteria, and dsDNA viral sequences. Functional annotation of transmembrane proteins indicate that sites comprised of organisms with small genomes have an enrichment of transporters with substrate specificities for amino acids, iron and cadmium; whereas, organisms with larger genomes have a higher percentage of transporters for ammonium and potassium. Eukaryotic-type glutamine synthetase (GS) II proteins were identified and taxonomically classified as viral, most closely related to the GSII in Mimivirus, suggesting that marine Mimivirus-like particles may have played a role in the transfer of GSII gene functions. Additionally, a Planctomycete bloom was sampled from one upwelling site providing a rare opportunity to assess the genomic composition of a marine Planctomycete population. The significant correlations observed between genomic properties, community structure, and nutrient availability provide insights into habitat-driven dynamics among oligotrophic versus upwelled marine waters adjoining each other spatially.

Publication

1. In Press: Lisa Zeigler Allen, Eric E. Allen, Jonathan H. Badger, John P. McCrow, Ian T. Paulsen, Liam D. H. Elbourne, Mathangi Thiagarajan, Doug B. Rusch, Kenneth H. Nealson, Shannon J. Williamson, J. Craig Venter, Andrew E. Allen. Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. *ISME Journal.*

This work was funded by the Office of Biological and Environmental Research in the DOE Office of Science (DE-FC02-02ER63453)

143

Genomic Insights to SAR86, an Abundant and Uncultivated Marine Bacterial Lineage

Chris L. Dupont* (cdupont@jcvi.org), Douglas B. Rusch, Shibu Yooseph, Mary-Jane Lombardo, R. Alexander Richter, Ruben Valas, Mark Novotny, Joyclyn Yee-Greenbaum, Jeremy D. Selengut, Dan H. Haft, Aaron L. Halpern, Roger S. Lasken, Kenneth Nealson, Robert Friedman, and **J. Craig Venter**

J. Craig Venter Institute, San Diego Calif.

Project goals: Our basic research program focuses on acquisition and analysis of marine microbial metagenomic data and development of genomic analysis tools for broad, external community use. Our Marine Metagenomic Diversity effort generated and analyzed shotgun sequencing data from microbial communities sampled from >260 sites around the world. Here we used informatic and single cell techniques to generate genomes for uncultivated but abundant organisms, adding genomic context to the metagenomes.

Bacteria in the 16S rRNA clade SAR86 are among the most abundant uncultivated constituents of microbial assemblages in the surface ocean for which little genomic information is currently available. Bioinformatic techniques were used to assemble two nearly complete genomes from marine metagenomes and single-cell sequencing provided two more partial genomes. Recruitment of metagenomic data shows that these SAR86 genomes substantially increase our knowledge of non-photosynthetic bacteria in the surface ocean. Phylogenomic analyses establish SAR86 as a basal and divergent lineage of γ-proteobacteria, and the individual genomes display a temperature-dependent distribution. Modestly sized at 1.25–1.7 Mbp, the SAR86 genomes lack several pathways for amino-acid and vitamin synthesis as well as sulfate reduction, trends commonly observed in other abundant marine microbes. SAR86 appears to be an aerobic chemoheterotroph with the potential for proteorhodopsin-based ATP generation, though the apparent lack of a retinal biosynthesis pathway may require it to scavenge exogenously-derived pigments to utilize proteorhodopsin. The genomes contain an expanded capacity for the degradation of lipids and carbohydrates acquired using a wealth of tonB-dependent outer membrane receptors. Like the abundant planktonic marine bacterial clade SAR11, SAR86 exhibits metabolic streamlining, but also a distinct carbon compound specialization, possibly avoiding competition.

This work was funded by the Office of Biological and Environmental Research in the DOE Office of Science (DE-FC02-02ER63453)

144

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

Christopher L. Dupont[1]* (cdupont@jcvi.org), James McCarthy,[1] Adriano Nunes-Nesi,[2] Sabrina Kleeson,[2] Alisdair Fernie,[2] and **Andrew E. Allen**[1]

[1]Microbial and Environmental Genomics, J. Craig Venter Institute, San Diego, Calif. and [2]Max-Planck-Insitut für Molekulare Pfanzenphysiologie, Potsdam-Golm, Germany

Project Goals: While the complete genome sequence of a centric and pennate diatom, forward and reverse genetic techniques and *in silico* modeling have enabled our laboratory and others to begin characterizing unknown

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

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

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

References

1. Allen, A.E., Dupont,C.L., Oborník, M., Horák, A., Nunes-Nesi, A., McCrow, J.P., Zheng, H., Johnson, D.A., Hu, H., Fernie, A.F., Bowler, C. 2011. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473(7346):203-7
2. Lomas, M. W., and P. M. Gilbert. 1999. Temperature regulation of nitrate uptake: A novel hypothesis about nitrate uptake and reduction in cool-water diatoms. Limnology and Oceanography 44:556-572.
3. Parker, M. S., and E. V. Armbrust. 2005. Synergistic effects of light, temperature, and nitrogen source on transcription of genes for carbon and nitrogen metabolism in the centric diatom *Thalassiosira pseudonana* (Bacillariophyceae). Journal of Phycology 41:1142-1153.
4. Lomas, M. W., and P. M. Gilbert. 1999. Temperature regulation of nitrate uptake: A novel hypothesis about nitrate uptake and reduction in cool-water diatoms. Limnology and Oceanography 44:556-572.

This work was supported by the Office of Biological and Environmental Research in the DOE Office of Science.

145

Responses of and Interactions Between Nitrifying Bacteria to Environmental Changes: A Systems Level Approach

Peter Bottomley* (peter.bottomley@oregonstate.edu), Jeff Chang, Yannming Di, José Perez, Luis Sayavedra-Soto, and **Daniel Arp**

Oregon State University, Corvallis, Ore.

Project Goals: In nitrification, ammonia oxidation and nitrite oxidation are usually coupled; however it is not known to what extent ammonia-oxidizing bacteria (AOB) and nitrite-oxidizing bacteria (NOB) interact with each other. Two model organisms, *Nitrosomonas europaea* (AOB) and *Nitrobacter winogradskyi* (NOB) are being used to test the hypothesis that AOB and NOB interact with each other, affecting their growth rate, cell yield, and protein and gene expression, and that these effects are distinct from the individually cultured bacteria. Their genome inventories and physiological and gene expression responses to environmental changes, singly and in co-cultures, will be used to construct genome-scale metabolic pathway models. We will create predictive models of AOB and NOB that incorporate metabolism, the regulatory interactions that influence metabolism, and the signaling networks for interaction with the environment. The models will help predict how AOB and NOB will behave in response to changes in the environment.

This project started October 2011. During this time we established the protocols to culture *N. europaea* and *N. winograskyi*, singly and in coculture. The cells are being grown in chemostats and in pH controlled batch fermentors.

A viable chemostat with a coculture of *N. europaea* and *N. winograskyi* was obtained within a month and has been successfully maintained over four months under constant conditions. The hydraulic retention time of 8 days was established and cells were harvested for analysis, allowing for recovery between samplings to minimize variation. From these coculture samples total RNA was isolated and is being used to determine the whole-genome transcript levels. The batch cultures and cocultures were grown in similar medium and at a constant pH. The cells in the batch cultures were harvested at late logarithmic phase. Similarly, total RNA was isolated and is being used to determine the whole-genome gene expression levels.

We are using RNA-Seq to quantify gene expression levels, to detect all messages (including small RNAs) and to avoid cross-hybridization artifacts. An optimized protocol to discriminate the transcriptomes of *N. europaea* and *N. winogradskyi* cocultures is being applied. The computational and statistical methods allow discriminating of the gene expression levels of the two model organisms (Table 1).

Table 1: The large majority of RNA-Seq reads can be uniquely assigned in the genomes of *N. europaea* and *N. winogradskyi*.*

CDS	Possible 25mers	Unique to itself	% to itself	Unique to other	% to other
N. europaea	2,345,948	2,224,679	94.83%	2,224,640	94.83%
N. winogradskyi	2,785,084	2,638,289	94.73%	2,638,250	94.73%

*Unique 25mer sequences from either the concatenated coding sequences (CDSs) per genome were identified. Unique is defined as a sequence found only once (unambiguous). We attempted to align the unique 25mers against the concatenated coding sequences (CDSs) or genome of the other species with no allowance for mismatches. We selected a 25mer for the comparison since most errors in the ~36-bases reads of RNA-Seq are at either end. The reads are trimmed prior to the alignments.

We designed a method to monitor cell viability of *N. europaea* and *N. winogradskyi* while in coculture using allylthiourea (ATU). ATU is a selective inhibitor of ammonia oxidation but not of nitrite oxidation. In this method the ammonia- and nitrite-dependent oxygen uptake are monitored to assess the overall cell health and abundance of each population. The method also allows the monitoring of AOB and NOB activities in coculture through growth, presenting an advantage over monitoring nitrite or nitrate accumulations solely. With this method the relative cell abundances of *N. europaea* and *N. winogradskyi* during coculture growth can be determined within 10% accuracy.

We anticipate that analysis of transcriptomes, growth rates, and metabolites will provide insights to the responses of cells when in co-cultures. We anticipate that changes in gene expression will reflect how cells are sensing and responding to the environmental changes we impose on the co-cultures. The metabolic models will provide a basis to predict and analyze responses of the cell types in co-culture.

Funding for this work is through the U.S. Department of Energy and through the Experimental Station at Oregon State University.

146

Regulons, Uptake, and Salvage Pathways that Mediate Metabolite Exchange in Hot Spring Microbial Mats

Dmitry A. Rodionov[1]* (rodionov@burnham.org), Frank Collart,[2] Aaron T. Wright,[3] Margie Romine,[3] Jim Fredrickson,[3] and **Andrei L. Osterman**[1]

[1]Sanford-Burnham Medical Research Institute, La Jolla, Calif.; [2]Argonne National Laboratory, Argonne, Ill.; and [3]Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: The objective of the Pacific Northwest National Laboratory Foundational Scientific Focus Area (FSFA) is to develop a predictive, genome-enabled understanding of how microbial interactions impart stability, robustness, and functional efficiency to microbial communities. As part of this multi-institutional effort, we are utilizing genomic sequence to identify candidate systems that mediate and control the synthesis and secretion of chemical commodities by the autotroph and the cor-

responding systems that enable their uptake and salvage by interacting heterotrophs. By compiling this knowledge and employing various experimental methods to validate novel functional predictions, we intend to improve our ability to predict the nature of beneficial partnerships that occur between microorganisms that occupy the same niche.

Research is being conducted on three separate microbial communities; phototrophic mats that occur in alkaline siliceous hot springs in Yellowstone National Park (YNP) or a meromictic, hypersaline lake in northern Washington state and chemotrophic mats that occur in acidic hyperthermophilic springs in YNP. As sequence becomes available from organisms (or their near relatives) that occupy these communities we use integrative genomics-based reconstruction and experimental assessment of metabolic and regulatory networks to understand metabolic interactions and associated regulatory interactions in the target microbial communities. Our general workflow includes: (i) the subsystems-based genomic reconstruction of selected pathways and regulons; (ii) identification of currently unknown components of these pathways, primarily transporters, sensory and regulatory proteins using comparative genomics and high-throughput -*omics* data; (iii) experimental testing of selected functional predictions combining biochemical and genetic methods; and (iv) use model co-cultures to assess metabolic exchange in microbial communities.

Our current research focuses on members of the Cyanobacteria and Chloroflexi phyla that were isolated from or related to isolates detected in YNP phototrophic mats. In these mats, anoxygenic phototrophic bacteria (e.g., *Roseiflexus*, *Chloroflexus*) are thought to function predominantly as photoheterotrophs using metabolites produced by cyanobacteria (e.g., *Synechococcus*) as carbon sources. Whereas typical metabolic byproducts may provide a major flux of carbon and energy, vitamins (precursors of key cofactors) are required in relatively small amounts. We hypothesize that vitamin exchange may be rather widespread phenomena contributing to "opportunistic" relationships between species in mat communities. In a previous genomic survey, we observed a mosaic distribution of *de novo* and salvage pathways for biogenesis of major vitamins (such as niacin, pantothenate, biotin, thiamin, riboflavin) that leads to the presence in communities of the strict auxotrophs and strict prototrophs with respect to one or another vitamin. In this work, we focused on reconstruction of metabolic pathways and regulons involved in metabolite salvage and exchange in two groups of environmental bacteria, Chloroflexi and Cyanobacteria. Several co-cultures (e.g., *Synechococcus sp.* PCC 7002 and *Roseiflexus castenholzii*) have been established by combining isolates available in pure culture and thus provide tractable models for studying naturally occurring interactions.

In Cyanobacteria, we used the comparative genomic approach to reconstruct regulons for 18 transcription factors (including 14 previously known in the literature) and 10 riboswitches. This approach led to the discovery of a putative B_{12} transport uptake system (BtuFCD) in *Synechococcus sp.* PCC 7002. This strain lacks the ability to synthesize B_{12} and thus is dependent on acquiring it from other microbes that are able to produce it. The transporter prediction will be validated through knock-out construction (in collaboration with Don Bryant) and *in vitro* assay. In addition to using regulon/subsystem analysis to discover additional novel B_{12} transporters we will use chemical probe profiling to discover them experimentally.

In Chloroflexi, a repertoire of all transcription factors and environmental sensors encoded in 5 sequenced genomes from this phylum was identified and analyzed. Several bioinformatics workflows are used for identification of conserved DNA and RNA motifs and *ab initio* reconstruction of associated regulons. Currently this approach resulted in reconstruction of 20 regulons operated by either transcription factors or riboswitches, and the number of reconstructed regulons is growing.

The reconstructed regulons lead to the discovery of multiple novel uptake transport systems for essential vitamins (thiamin/B_1, riboflavin/B_2, B_{12}), metal ions (manganese, zinc, ferrous iron, molybdenum, cobalt), and carbohydrates (e.g., rhamnose). These novel transport systems fill the gaps in the respective metabolic pathways. Several examples of novel vitamin uptake systems were selected for further experimental validation. First, a putative B_2 transporter (RibXY) complements the absence of the *de novo* riboflavin biosynthesis pathway in *Roseiflexus* spp. The riboflavin transfer to *Roseiflexus* could be important in co-culture with B_2-prototrophic strains *Synechococcus* and/or *Thermochlorobacter*. Second, we identified putative B_1 precursor transporters (ThiXYZ, ThiW, CytX) that are required for B_1 salvage pathways in the auxotrophic bacteria from the Chloroflexi phylum. The ABC transporter predictions (ThiY, RibY) will be validated through *in vitro* assay. Third, we found a novel ECF-type transport system of yet unknown specificity that is involved in B_{12} metabolism in Chloroflexi. Finally, we found multiple predicted transport systems that are involved in the uptake of biometals (Fe, Zn, Mn, Co, Mo). They are regulated by committed transcription factors to maintain the cell homeostasis.

L-rhamnose is an essential component of the external polysaccharide layer in the cyanobacterial cells. Analysis of *Chloroflexus/Roseiflexus* genomes revealed a conserved gene cluster encoding enzymes from the rhamnose catabolic pathway. We hypothesize that transfer of carbon from YNP *Synechococcus* spp. to YNP *Chloroflexus* via L-rhamnose is an example of metabolite exchange that occurs in native YNP mats. We have identified a novel ABC-type transport system (RhaGHJF) and a DeoR-type transcription factor (RhaR) encoded in the rhamnose gene cluster and propose that they are involved in uptake and sensing of L-rhamnose to induce the catabolic pathway. The RhaR DNA binding sites were predicted and the respective rhamnose regulon was reconstructed by comparative genomics. Experimental validation of predicted functional roles of the novel rhamnose transporter and regulator in a selected model strain of *Roseiflexus castenholzii* is ongoing. In addition, we intend to test the hypothesis of L-rhamnose exchange between Cyanobacteria and Chloroflexi in a model co-culture combining gene expression, proteomics and metabolomics analyses.

This research is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

147

Methanotrophic-Mediated Metal Binding: Effects on *In Situ* Microbial Community Structure and the Sustainability of Subsurface Water Systems

Jeremy D. Semrau[1]* (jsemrau@umich.edu), Alan A. DiSpirito,[2] and Nathan Bandow[2]

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

http://sitemaker.umich.edu/methanobactin

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

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

This project, starting in September 2011, has as its immediate objective the characterization of metal binding by methanobactin produced by different methanotrophs. The molecule has been identified in a number of methanotrophs, but has only been structurally characterized from Methylosinus trichosporium OB3B (mb-OB3b) and Methylocystis strain SB2 (mb-SB2). mb-SB2 shows a significant similarity to mb-OB3b, including spectral and metal binding properties, and the ability to bind and reduce Cu(II) to Cu(I). Both forms of methanobactin contains similar five-member rings and associated enethiol groups, which together, form the metal ion binding site. Further, greatest amounts of methanobactin are found when both Methylosinus trichosporium OB3B and Methylocystis strain SB2 are grown in low (<1 μM) copper. Significant differences exist between the two known forms of methanobactin (Figure 1), including: (1) the number and types of amino acids used to complete the molecule, (2) mb-OB3b has two oxazolone rings for copper binding, while mb-SB2 has one imidazolone ring and one oxazolone ring, and; (3) a sulfate group is found in mb-SB2 but not in mb-OB3b. The sulfate in mb-SB2 is bonded to a threonine-like side chain and may represent the first example of this type of sulfate group in a bacterial-derived peptide. The gene sequence for a ribosomally produced precursor for mb-OB3b has also been identified in the genome of *Methylosinus trichosporium* OB3b. It indicates that the oxazolone rings in mb-OB3b are derived from glycine and cysteine residues. Taken together, the results reported here suggest methanobactins are a structurally diverse group of ribosomally-produced, peptide-derived molecules that share a common pair of five-member rings with associated enethiol groups that bind and reduce copper in aqueous environments.

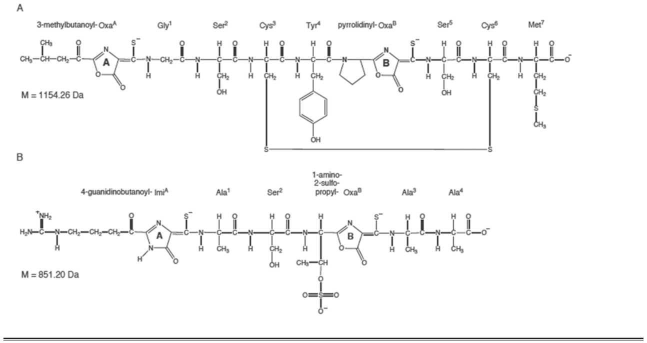


Figure 1. Primary structure of methanobactin from (A) *M. trichosporium* OB3b and (B) *Methylocystis* strain SB2.

* Presenting author

Comparison of the spectral and thermodynamic properties of mb-SB2 demonstrate that the spectral and basic copper binding properties of both methanobactins were similar and the metal binding capacity, particularly for copper, of both methanobactins lies primarily, if not solely, in the pair of five-membered rings and associated enethiol groups. To date, we have surveyed the range of metals bound by both forms of methanobactin, and have found that both forms bind many different metals, including mercury, chromium, lead, and uranium, and that metals are typically reduced after binding by methanobactin.

Next steps will characterize the effect of methanobactin-mediated metal speciation on activity and transcriptome of pure cultures of methanotrophs. In addition, we will collect groundwater and soil samples from the Integrated Demonstration Site of the DOE Savannah River Site where methanotrophs are known to exist and characterize how methanobactin affects metal mobility and bioavailability in the presence of soils from this site as well as methanobactin-mediated dissolution of soil-associated minerals using a set of spectroscopic and imaging techniques. The resultant effects on the broader microbial community structure and function will be also be determined using a "double-RNA" approach to characterize microbial community structure and function simultaneously from the meta-transcriptome.

148

Improved Understanding of Microbial Iron and Sulfate Reduction Through a Combination of Bottom-Up and Top-Down Functional Proteomics Assays

Ruth E. Richardson[1]* (rer26@cornell.edu), Annette R. Rowe,[1] Peter Kelly,[1] Hening Lin,[1] Sheng Zhang,[1] Robert Sherwood,[1] Lee Ann McCue,[2] Michael Wilkins,[2] and Stephen J. Callister[2]

[1]Cornell University, Ithaca, N.Y. and [2]Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: Our overall project objectives are to improve our understanding of the systems biology of the dissimilatory iron reducing bacteria (DIRB) and sulfate reducing bacteria (SRB) and to use this knowledge to aid in site-monitoring and management via informative biomarkers. The work will employ a wide range of proteomic and biochemical assays. These include both "bottom-up" and "top-down" approaches. In the bottom up approaches, gene targets are selected for heterologous expression based upon proteomic expression patterns and are then put through a range of *in silico*, proteomic and biochemical tests. In the top-down approach, we will develop a high-throughput functional protein assay that screens for function first before then identifying the responsible enzymes– thereby focusing mass spectrometry resources on the identification of functionally-relevant oxidoreductase enzymes and their interactors.

The advent of the genomics era has advanced our understanding not only of the genetic makeup of a wide range of microorganisms (and microbial communities) it has also enabled remarkably high throughput monitoring of gene expression for thousands of genes simultaneously. With current *in silico* bioinformatic tools, inferences can often be made as to the function of some of these gene targets simply from gene sequences or (in the case of "hypothetical" proteins) from expression patterns under different conditions. However, inferences are only the start with respect to learning the true function of poorly annotated gene targets. To turn these observations and inferences into an improved understanding of the systems biology of organisms of interest, functional annotation techniques and approaches must be improved.

We have created an experimental plan to tackle this functional annotation bottleneck in two groups of microorganisms of particular interest for heavy metals and radionuclide bioremediation: the dissimilatory iron reducing bacteria (DIRB) and the sulfate reducing bacteria (SRB). We anticipate that our approaches will both generate a better fundamental understanding of the systems biology of the DIRB and SRB, and lead to improved diagnostic tools for application at contaminated field sites—namely peptide biomarkers that can be used to document and enhance *in situ* activity. To this end we are particularly interested in improved functional characterization of oxidoreductase enzymes as they are key at the biotic/abiotic interfaces in aquifers.

The work is taking take place in three complementary phases. In Phase 1 ("bottom-up"), a combination of *in silico*, proteomic, and biochemical approaches are being applied to phylogenetically-diverse DIRB and SRB cultures (grown under different operating conditions of type of electron acceptor and rate of respiration). This phase will first identify the core proteomes specific to iron and sulfate reduction, then will focus on improved functional characterization of this core proteins via: follow-up *in silico* analysis of sequences, follow-up proteomics to identify protein-protein interaction partners, and biochemical assays focused on documentation of redox capabilities. In Phase 2 ("top-down"), we are developping a high-throughput functional redox protein assay consisting of three steps: non-denaturing separation of a (meta)proteome, parallel assaying of wells for desired redox activity, and tandem mass spectrometry identification of proteins in positive wells. In Phase 3, proteotypic peptides will be elucidated for a suite of functionally-informative enzymes (already known and newly discovered in Phases 1 and 2) and Mass Western assays will be developed for the simultaneous quantification of dozens of such peptide biomarkers in a single nanoLC-MS/MS injection. In Phases 2 and 3, assays will first be tested on lab culture proteomes and then used to quantify biomarkers in biomass samples from sites undergoing in situ bioremediation (e.g. the Rifle site). In our poster presentation we will highlight work performed in the first six months of the project.

This work is fully funded through the DOE Office of Biological and Environmental Research within the Office of Science.

149

In situ Spectroscopy on Intact *Leptospirillum ferrooxidans* Reveals that Reduced Cytochrome 579 is an Intermediate in the Iron Respiratory Chain

Robert C. Blake II* (rblake@xula.edu) and Megan N. Griff

Xavier University of Louisiana, New Orleans

Project Goals: The practical goal is to develop a new and powerful approach to quantitatively examine the dynamic behavior of bacterial electron transport systems at the microbe-mineral interface, an interaction that has heretofore been difficult to probe directly. We will accomplish this goal by refining and exploiting a novel integrating cavity absorption meter (ICAM) where the sample observation cell is also a spherical reflecting cavity. Light scattering losses due to sample turbidity will be eliminated or minimized by multiple reflections at the cavity wall that will prevent the scattered light from escaping the detector. The utility of this novel ICAM to study *in situ* bacterial electron transfer reactions will be demonstrated in real time equilibrium and kinetic measurements of electron transfer reactions in intact bacteria that respire by exchanging electrons with soluble and insoluble extracellular iron. The ability to conduct accurate real-time absorbance measurements in live organisms will permit any biological process that involves colored biomolecules to be studied in new ways.

This abstract introduces a new means to study respiratory electron transfer reactions *in situ* in intact bacteria under physiological solution conditions. The premise is that accurate UV-visible spectroscopy of electron transfer reactions among colored biomolecules can be conducted in highly turbid suspensions if the live bacteria are irradiated in an isotropic homogeneous field of incident measuring light. Under those conditions, the absorbed radiant power is independent of scattering effects. We conducted equilibrium and kinetic studies on the Fe(II)-dependent reduction and O_2-dependent oxidation of cytochromes in intact *Leptospirillum ferrooxidans* at pH 1.7. We used a novel integrating cavity absorption meter where the cuvette comprised a reflecting cavity completely filled with the absorbing suspension. *L. ferrooxidans* was selected because it is only known to respire on one substrate, reduced iron.

The aerobic iron respiratory chain of *Leptospirillum ferrooxidans* was dominated by the redox status of an abundant cellular cytochrome with an absorbance peak at 579 nanometers in the reduced state. Intracellular cytochrome 579 was reduced within the time that it took to mix a suspension of the bacteria with soluble ferrous iron at pH 1.7. Subsequent oxidation of the reduced cytochrome appeared to be the rate-limiting step in the overall aerobic respiratory process. Steady state turnover experiments were conducted where the concentration of ferrous iron was less than or equal to that of the oxygen concentration. The concentration of the reduced cytochrome 579 at any time point was directly proportional to the velocity of product ferric ion formation. Further, the integral of the area of the reduced cytochrome accumulated over time was also directly proportional to the total concentration of ferrous iron in each reaction mixture. These kinetic data obtained using whole cells were consistent with the hypothesis that reduced cytochrome 579 is an obligatory steady state intermediate in the iron respiratory chain of this bacterium.

The direct and accurate observation of absorbance changes *in situ* in intact organisms is a useful complement to traditional reductionist approaches and recent advances in proteomic and transcriptomic studies. The colored prosthetic groups of most electron transport proteins comprise intrinsic spectrophotometric probes whereby transient changes in the oxidation-reduction state of the proteins may be monitored with great sensitivity. There is no better means to establish physiological relevance in a metabolic function than to directly observe the function as it occurs in the intact bacterium. The movement of electrons through electron transfer complexes is central to energy production in all living cells. The ability to conduct direct spectrophotometric studies under noninvasive physiological conditions represents a new and powerful approach to examine the extents and rates of biological events *in situ* without disrupting the complexity of the live cellular environment. Studies such as these should increase our fundamental understanding of biological energy transduction.

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

150

DOE JGI Plant Genomics Program

Jeremy Schmutz[1,2]* (jschmutz@hudsonalpha.org), Jarrod Chapman,[2] David Goodstein,[2] Jane Grimwood,[1,2] Uffe Hellsten,[2] Jerry Jenkins,[1] Gerald A. Tuskan,[2,3] Jim Bristow,[2] Kerrie Barry,[2] Daniel S. Rokhsar,[2] and **Edward M. Rubin**[2]

[1]DOE JGI HudsonAlpha Genome Sequencing Center, Huntsville, Ala.; [2]DOE Joint Genome Institute, Walnut Creek, Calif.; and [3]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

Project Goals: The goal of the DOE JGI Plant Genome Program is to shed light on the fundamental biology of photosynthesis and transduction of solar to chemical energy. Other areas of interest include characterizing:

1. **Ecosystems and the role of terrestrial plants and oceanic phytoplankton in carbon sequestration.**
2. **The role of plants in coping with toxic pollutants in soils by hyper-accumulation and detoxification.**
3. **Feedstocks for biofuels, e.g., biodiesel from soybean; cellulosic ethanol from perennial grasses and trees.**
4. **The ability to respond to environmental change (e.g., loss of diversity associated with changes in tempera-**

ture or moisture availability; nitrogen fixing nodules in legumes reduce fertilizer need).

5. **The generation of useful secondary metabolites for positive/negative pest control in natural ecosystems with attendant influence on global carbon cycle.**

The Plant Genome Program contributes to and accomplishes these goals through the following activities:

Sequence. Produce genome sequences of key plant (and algal) species to accelerate biofuel development and understand response to climate change.

Function. Develop datasets (and synthetic biology tools) to elucidate functional elements in plant genomes, with special focus on handful of "flagship" genomes.

Variation. Characterize natural genomic variation in plants (and their associated microbiomes), and relate to biofuel sustainability and adaptation to climate change.

Integration. Provide a centralized hub for the retrieval and deep integrated analysis of plant genome datasets.

JGI plant projects are initiated directly from three major sources: DOE facilities such as the BioEnergy Research Centers; the DOE scientific research communities via the JGI Community Sequencing Program; and JGI Plant Program Projects directed at the improvement of community resources for DOE plant science. These projects fall into broad scientific categories including: plant *de novo* genome sequencing and improvement, diversity and population analysis, transcript profiling with RNA-seq, and mapping and recombination analysis of populations.

151

Genome Improvement of DOE JGI Flagship Plant Genomes

David Sims[1]* (dsims@hudsonalpha.org), Jerry Jenkins,[1] Wellington Muchero,[2] Gerald A. Tuskan,[2,3] Jim Bristow,[3] Kerrie Barry,[3] Daniel S. Rokhsar,[3] Jeremy Schmutz,[1,3] Jane Grimwood,[1,3] and **Richard M. Myers**[1]

[1]DOE JGI HudsonAlpha Genome Sequencing Center, Huntsville, Ala.; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [3]DOE Joint Genome Institute, Walnut Creek, Calif.

Project Goals: As part of the DOE JGI Plant Program, we are working to experimentally and computationally improve the genomic sequences of key DOE plant species which we have designated the Plant Flagships. These genomic sequences and related data support mission aims in producing cellulosic feedstocks and understanding carbon cycling and carbon sequestration.

Through discussion with the JGI Plant Genome Advisory Committee and DOE, a set of JGI plant genomes that are the most important to DOE mission and plant science have been designated as JGI Plant Flagship Genomes. This selected set of critical species allows us to focus our computational and experimental efforts to move beyond sequence to function and to provide the most direct benefit to mission science. The current Plant Flagships are:

- Poplar—the DOE tree, the basis for cellulosic research at ORNL
- Sorghum—widely planted grass crop for biomass, cellulose, and sugar
- *Brachypodium*—small grass model organism
- *Chlamydomonas*—the most studied algal species, model algal organism
- Soybean—the source of biodiesel and the number two US economic crop
- Foxtail millet—a grass model, recently evolutionary diverged from switchgrass
- *Physcomitrella*—moss model organism, basic comparator for land plants

Although the flagship designation was initially intended for genomes that had already been draft sequenced, we also consider several additional genomes as proto-flagships because of their importance to DOE biofuel mission; initial sequencing efforts are on-going along with the creation of genomic resources:

- *Panicum virgatum* (switchgrass)—a candidate biofuel feedstock that grows on marginal soil and is being used by all of the BioEnergy centers as a model crop species
- Miscanthus—a perennial grass species that produces large amounts of cellulosic material with low agricultural inputs, one of the top feedstock candidates
- *Panicum hallii* (Hall's panicgrass)—a small, evolutionary nearby diploid relative of switchgrass that may serve as laboratory model organism for switchgrass research

For genome improvement, the objective is to completely (or as near as possible) resolve each base pair within a localized region and then using mapping resources, position and incorporate the sequence into chromosome scale pseudomolecule assemblies. In some cases, we have needed to construct the mapping resources ourselves to push forward a chromosome scale release. For base pair level improvement, which we carry out for the entire gene space of an organism or directed regions, we collect experimental data in the form of primer walks with a variety of chemistries and templates, transposon mediated sequencing and shatter libraries from clones (the latter two techniques being essentially subprojects of the subprojects). We have ongoing efforts to explore practical uses of next-generation sequencing technologies for improving the extremely large and polyploid plant genomes. We release new versions of the Plant Flagship genomes periodically as we complete significant iterative improvements. The sequence and subsequent annotation is made public through the phytozome web site at www.phytozome.net.

We have recently completed major updates to the sequences of *Chlamydomonas, Physcomitrella* and poplar, all of which are currently in the annotation phase. We are proceeding with a directed improvement project of sorghum gene space and a complete improvement project for *Brachypodium.* With these improved sequences we seek to 1) provide near complete information for genomes of direct DOE mission importance, 2) support BRC and others efforts for cellulosic biofuel development and plant customization efforts, 3) foster communities to develop scientific research programs working with DOE organisms and attract new scientists to solve DOE problems by providing tool sets to work with these organisms, and 4) build a solid foundation for diversity, phenotype, resequencing, transcriptomic studies, and functional studies in these organisms.

152

Fidelity and Dynamics of DNA Methylation in Plants

Qin Yao[1]* (qyao@bnl.gov), Changjun Liu,[1] John Shanklin,[1] Chuan He,[2] and **John Dunn**[1]

[1]Department of Biology, Brookhaven National Laboratory, Upton, N.Y. and [2]Department of Chemistry and Institute for Biophysical Dynamics, University of Chicago, Chicago, Ill.

Project Goals: The main goal of our program is to identify fundamental pathways and protein-DNA interactions maintaining inheritance of epigenetic states of plant DNA and chromatin during cell divisions. During semi-conservative DNA replication, hemi-methylated sites are generated where only cytosines in the parental strands remain modified. The correct methylation patterns in the daughter strands need to be reestablished to ensure proper epigenetic homeostasis, i.e. expression of essential genes and silencing of parasitic transposons. Recently we found that Arabidopsis genomic DNA contains 5-hydroxymethylcytosine (5hmC) as well as 5-methylcytosine (5mC)[(1)] which raises several questions: 1) where are these residues located; 2) do proteins that help maintain proper DNA methylation patterns in replicated DNA also maintain these patterns after replication across sites with 5hmC; and 3) do these two modifications play distinct epigenetic roles? To help address these questions, we are using a sensitive chemical labeling method to capture DNA fragments with 5hmC to discover its distribution pattern in the *Arabidopsis* genome and *in vitro* DNA binding studies to determine whether proteins needed for high-fidelity maintenance of cytosine methylation lose their intrinsic preference for hemi-methylated DNA when 5mC is replaced by 5hmC. Ultimately, an efficient and single-base resolution method for 5hmC identification is highly desirable. To this end we will integrate J binding protein, which has been shown to recognize single 5hmC residues[(2)] with next generation sequencing techniques to develop a new high-resolution 5hmC mapping tool which can be easily operated in a standard molecular lab. This project will also attempt to use genome-wide association analysis to enhance our understanding that any observed changes in DNA methylation status are non-random, gene-specific alterations that can result in gene activation or inactivation, especially for genes involved in lignocellulose biosynthesis and oil formation. All the newly generated epigenetic data and tools will be deposited into the BESC Knowledge Base (KBase).

In *Arabidopsis*, several different mechanisms are utilized to reestablish appropriate epigenetic methylation marks in the daughter strand cytosines. One of these mechanisms involves the five member VIM (Variant in Methylation) family of proteins which help reestablish daughter strand cytosine methylation at symmetric CG dinucleotides. Using recombinant proteins and modified double-stranded deoxyoligonucleotides, we observed that full-length VIM1 binds preferentially to hemimethylated DNA with a single modified 5mCG site; a result consistent with its known role in preserving DNA methylation in vivo following DNA replication. However, when 5hmC replaces one or both cytosine residues at a palindromic CpG site, VIM1 binds with approximately 15-fold lower affinity. VIM3, another member of the five member VIM family, has a single amino acid change, S317A, relative to VIM1, 2, 4 and 5 in the region known to be important for binding to methylated DNA. Changing this residue in VIM1 did not significantly affect its DNA-protein binding properties. These results suggest that 5hmC may contribute to VIM-mediated passive loss of cytosine methylation in vivo during *Arabidopsis* DNA replication. Work is now in progress to extend these observations to other proteins involved in maintaining epigenetic fidelity to determine their ability to discriminate between substrates with 5mC vs. 5hmC using purified recombinant proteins as well as protein binding microarrays (PBMs)[(3)].

Previous studies of DNA hydroxymethylation suggest hydroxylation of 5mC may promote transcriptional de-repression by dissociation of mC-binding proteins and/or recruitment of effector proteins[(4)]. Our preliminary profiling of 5hmC residues in *Arabidopsis* reveals that 5hmC is enriched selectively in gene body regions rather than more distal regions and that there is a significantly different pattern around transcription start regions (TSS). This relative distribution is associated with gene expression levels. There is a highly negative correlation between enrichment in proximal regions to TSS's but no significant correlation between enrichment in the gene bodies and gene expression level. These results suggested that conversion of 5mC to 5hmC, possibly coupled to active or passive demethylation, might act as a switch to fine tune epigenetic homeostasis or perhaps have an active role in reprogramming the epigenome. Functional epigenomic studies will eventually address the mechanism from epigenotype to phenotype.

References

1. Yao, Q., Song, C.X., He, C., Dunn, J.J. Detection of 5-hydroxymethylcytosine in *Arabidopsis* and its potential impact on recognition of VIM1 (under review).
2. Robertson, A.B., Dahl, J.A., Vagbo, C.B., Tripathi, P., Krokan, H.E. and Klungland, A. (2011) A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res, 39, e55.

* Presenting author

3. M.L. Bulyk, Analysis of sequence specificities of DNA-binding proteins with protein binding microarrays. (2006) Methods Enzymol 410, 279-299.
4. 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.

153

PET Radiotracer Imaging in Plant Biology

Drew Weisenberger* (drew@jlab.org)

Thomas Jefferson National Accelerator Facility, Newport News, Virginia

The ability to detect the emissions of radioactive isotopes through radioactive decay (e.g. beta particles, x-rays and gamma-rays) has been used for over 80 years as a tracer method for studying natural phenomena. The Hungarian-born chemist George de Hevesy in 1913 while in Vienne published with Fritz Paneth the first account of the use a radioactive isotope as a tracer in a paper titled: “The Solubility of Lead Sulphide and Lead Chromate.” In 1923 Hevesy published his studies on the transport of the radioisotopes lead-210 and lead-212 in living plants. In 1948 in a University of California Radiation Laboratory report Melvin Calvin and Andrew Benson describe research into photosynthesis of brown algae using the radioactive tracer carbon-14. More recently a positron emitting radioisotope of carbon: carbon-11 has been utilized as a $^{11}CO_2$ tracer for plant ecophysiology research. Because of its ease of incorporation into the plant via photosynthesis, the $^{11}CO_2$ radiotracer is a powerful tool for use in plant biology research. Positron emission tomography (PET) imaging has been used to study carbon transport in live plants using $^{11}CO_2$. Presently there are several groups developing and using new PET instrumentation for plant based studies. Instrumentation originally developed for small animal PET has been modified for imaging ^{11}C in plant research as currently undertaken at Institute Phytosphere, Forschungszentrum Jülich in Germany. A two-headed planar PET plant imaging system for ^{11}C has also been described and used in plant studies by a research group at the Japan Atomic Energy Agency. At Brookhaven National Laboratory researchers have used a clinical human PET scanner to track the distribution of ^{11}C labeled metabolites in plants in response to environmental changes. The group I head at Thomas Jefferson National Accelerator Facility (Jefferson Lab) in collaboration with the Duke University Phytotron is also involved in PET detector development for plant imaging. In this presentation I will provide a brief overview of radiotracer imaging in plants and review some of the latest developments of the use of $^{11}CO_2$ tracer in plant studies of others as well as the developments of the Jefferson Lab/Duke University effort.

154

Pan-omics Informatics and Advanced Analysis Pipelines

Gordon A. Anderson* (gordo@pnnl.gov), Ronald J. Moore, Joshua N. Adkins, Kenneth J. Auberry, Mikhail E. Belov, Kristin E. Burnum, Kevin L. 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 J. Orton, Ljiljana Paša-Tolić, 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 T. Wright, Rui Zhang, Rui Zhao, and **Richard D. Smith** (rds@pnnl.gov)

Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Wash.

As part of our Proteomics and Pan-Omics project in support of GSP Systems Biology Research, we are developing an advanced informatics framework to enable application of pan-omics measurements that afford the comprehensive global molecular characterization necessary for understanding, modeling, and potentially manipulating biological systems. Pan-omics measurement capabilities are based on essentially identical separations-MS measurement platforms and similar data processing/informatics pipelines: metabolomics, lipidomics, and glycomics measurements in addition to expanded proteomics measurements. A key element of the pan-omics informatics framework are tools and methodologies for integrating data from various omics measurements and determining data quality and ambiguities (e.g., the confidence in peptide and protein identifications, modification sites, and abundance levels), as well as approaches for managing and communicating data.

We are leveraging the extensive experience and capabilities developed to date at PNNL to extend high throughput proteomics measurements to multiple omics measurements and provide a framework for 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. The Pan-omics Research Informatics Storage and Management (PRISM) system has been enhanced to support advanced pan-omics analysis pipelines that allow automated high level data analysis and integration of hundreds to thousands of individual datasets into one comprehensive research result report. The framework further supports integration of genomics data from public repositories and aims to provide the needed infrastructure to interoperate with the GTL Knowledgebase.

Current efforts are focused on developing a suite of data analysis tools, data consolidation applications, and statistical packages, as well as visualization software for data interpretation and integration. New and advanced tools support metabolomics, glycomics, and top-down proteomics pipelines, as well as pan-omics data integration. Each pipeline produces a self-documented, tested, and repeatable analysis, complete with visual inspection plots produced for each step in the pipeline. The advanced automation and critical infrastructure elements demonstrate high throughput and automated file processing, as well as data filtering and result aggregation.

Acknowledgements: This research is supported by the Office of Biological and Environmental Research (OBER) of the U.S. Department of Energy (DOE). Portions of this research were performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the DOE's OBER. Pacific Northwest National Laboratory is operated for the DOE by Battelle Memorial Institute through Contract No. DE-AC05-76RLO 1830.

155

Proteomics and Pan-omics Measurements for Comprehensive Systems Characterization of Biological Systems

Mary S. Lipton* (mary.lipton@pnnl.gov), **Thomas O. Metz*** (thomas.metz@pnl.gov), Kristin E. Burnum, Carrie D. Nicora, Qibin Zhang, Stephen J. Callister, Scott R. Kronewitter, Aaron T. Wright, Ronald J. Moore, Zeping Hu, Xiaoli Gao, Weijun Qian, Erin S. Baker, Therese R.W. Clauss, Karl K. Weitz, Daniel J. Orton, Matthew E. Monroe, Anil K. Shukla, Rui Zhao, Roslyn Brown, Joshua N. Adkins, **Gordon A. Anderson**, and **Richard D. Smith** (rds@pnnl.gov)

Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: This project is implementing and applying advanced capabilities for comprehensive molecular characterization of biological systems, including the extension of proteomics to cover post-translational protein modifications and the implementation of broad metabolomics, lipidomic and glycomic measurements, together with more widely available genomics and transcriptomics capabilities. Applications to microbes, plants and microbial communities enhance scientific understanding by elucidating phenotypic relationships between environmentally important microorganisms and higher organisms, as well as metabolic activities within microbial communities.

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. We are facilitating understandings of biological systems by applying pan-omics molecular measurement capabilities 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 measurement quality 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 elucidating protein isoforms (e.g., integrated top-down and bottom-up proteomics) and post-translational modifications (e.g., phosphoproteomics and characterization of protein glycosites).

The application to studies of fungus-growing ant–microbe symbiosis as a paradigmatic example of organic complexity generated through symbiotic association illustrates integration of these capabilities. Over the last decade, it has become a model system for studying symbiosis. We have demonstrated 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.

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.

156

Initial Application of a New LC-IMS-TOF MS Platform to Complex Environmental Samples

Erin Shammel Baker* (erin.baker@pnnl.gov), Yehia M. Ibrahim, William F. Danielson III, Kristin E. Burnum, Matthew E. Monroe, Kevin L. Crowell, Daniel J. Orton, Gordon W. Slysz, Brian J. Anderson, Carrie D. Nicora, Mary S. Lipton, Gordon A. Anderson, and **Richard D. Smith** (rds@pnnl.gov)

Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Wash.

As part of our Proteomics and Pan-Omics project in support of GSP Systems Biology Research, we are developing advanced pan-omics measurements capabilities for comprehensive global biomolecular characterization to enable the modeling, understanding, and potential manipulation of complex biological systems. Development and application of advanced separations-MS measurement platforms greatly increase measurement quality and throughput. The 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 to greatly expand biomolecule coverage and sensitivity, and the addition of the third dimension ion mobility separation allows for differentiation between molecule types. Pan-omics measurement capabilities are based on essentially identical separations-MS measurement platforms and similar data processing/informatics pipelines for metabolomics, lipidomics, and glycomics measurements, as well as for expanded proteomics measurements that cover a range of important modifications.

The extreme complexity of environmental samples presents many challenges for currently available MS-based analytical platforms. High concentrations of organic material (such as humic acid in soils and polyphenols in leaves) and sources of contamination from natural environments (such as high salts or other minerals) make environmental samples extremely difficult to analyze. Our new platform that utilizes advanced separations in conjunction with MS greatly increases measurement quality and throughput, even for “dirty” environmental samples. The new platform utilizes fast LC and an ion mobility separation (IMS) coupled with an ultra-fast and accurate mass measurement time-of-flight (TOF) MS to provide expanded proteome coverage and greater sensitivity when high concentration species are present in complex samples.

Previous studies of environmental ocean water, soil and leaf samples that have employed more conventional trapping MS instruments (e.g., LTQ Orbitrap MS) have presented a number of analytical challenges. For example, in peptide-centric analyses of complex samples, large collections of ionized species regardless of the type of molecule are essentially analyzed simultaneously in the ion trap that has a finite capacity. High concentrations of non-bio-molecular species eluted for extended periods of time rapidly fill the trap, making for very short accumulation times, which results in the inability to detect many lower concentration species that co-elute with the high concentration species. TOF mass spectrometers offer a promising solution to help with environmental samples in that they do not have an ion trap, so all ions are sampled equally at the detector instead of being limited by co-eluting high concentration species. These platforms also allow effective utilization of advanced, highly efficient ion sources and interface designs (e.g., using ion funnels) that increase dynamic range and detection of lower abundances species. Modern TOF mass spectrometers are also capable of both very high resolution and excellent (e.g., low ppm) mass measurement accuracies. The greatest advantage afforded by the new platform is that IMS is able to separate different classes of molecules onto specific ‘trend lines’ (e.g., peptides, lipids, oligonucleotides, etc.) because of distinctive backbones that make the molecules drift differently through the buffer gas and vastly different from small single ions or larger branched organic molecules. As a result, it is possible to separate and distinguish environmental contaminants from peptides or other biomolecule components in the same analysis.

Samples whose analyses have been known to be problematic with conventional MS-based platforms, such as ocean water isolates, soil extracts with humic acids, and plant extracts with poly phenols have recently been analyzed using the LC-IMS-TOF MS platform that provided coverage and measurement dynamic range much greater than with trapping-based MS platforms. Additionally, peptides and other biomolecules that have altered structures and different drift patterns can be distinguished, e.g., peptides with different post-translational modifications, adducts, or those that have been cross-linked. Modifications and adducts occurring in the same type of molecules also can be readily distinguished with the LC-IMS-TOF MS platform. For example, we recently found that mercury-modified peptides travel through the IMS buffer gas faster than peptides of the same mass-to-charge ratio, while phosphor- and heme-containing peptides travel slower because of the conformational changes induced by the modifications and adducts. By knowing these trends, specific modifications can be observed from complex mixtures.

This poster will highlight several complex environmental samples examined with commercially available instruments and with the IMS-TOF MS platform to illustrate the improvements in sensitivity and coverage observed with the new platform.

Acknowledgements: 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. DEAC05-76RLO 1830.

157

Integrating Transcriptomics and Proteomics Data for Identifying Mechanisms Regulating the Response of Microbes to Environmental Change

H. Steven Wiley[1]* (steven.wiley@pnnl.gov), Ronald C. Taylor,[2] Meng L. Markillie,[3] Mary S. Lipton,[3] Gretta Serres,[4] and Margie Romine[3]

[1]Environmental Molecular Sciences Laboratory, [2]Computational Biosciences, [3]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash.; and [4]Marine Biological Laboratory, Woods Hole, Mass.

Project Goals: The general goal was to identify global regulatory strategies used by microbes when responding to environmental changes. By integrating simultaneous proteomics and transcriptomics measurements, we sought to determine which cellular processes were regulated at the transcriptional versus post-transcriptional level. Using *Shewanella* as an initial test case, we developed approaches for accurately quantifying the level of protein and mRNA expression, integrating these data sets and visualizing their relationships on a global level. This approach was then evaluated for its ability to identify novel regulatory mechanisms and to obtain insights into regulatory strategies used by microbes in responding to environmental changes.

There are numerous mechanisms that regulate the response of microbial communities to changes in their environment. These include changes in the community composition as well as alterations in the complement of proteins that dictate the behavior of individual cells. Most investigations on how individual cells are regulated have focused on transcriptional control mechanisms, but there are other levels that could be important, such as translational control and protein turnover. The lack of suitable technologies to screen for alternate regulatory mechanisms on a global level has made it difficult to determine how common alternate mechanisms are and the role they might play in regulating microbial responses. To address this need, we developed a general and global approach for discriminating transcriptional from post-transcriptional regulatory mechanisms. The basic approach is to simultaneously quantify and compare the levels of proteins and mRNA transcripts on a global level. Proteins whose levels change in concordance with changes in their mRNA transcripts are considered transcriptionally regulated. Those that show discordance between changes in mRNA and protein levels are potentially regulated at a post-transcriptional level.

As a test case, we used *Shewanella* grown at steady state in chemostats under either high or low oxygen conditions. Samples were collected on three different days and subjected to quantitative proteomics and transcriptomics analysis. For proteomics, quadruplicate technical replicates were run for all three sets of biological replicates and peptides identified using the AMT approach (accurate mass and time). Transcriptional profiling was performed using SOLiD sequencing with triplicate technical replicates. The proteomics data were normalized and converted to relative protein levels by combining average ion intensity measurements and peptide counts followed by corrections for protein length. The mRNA levels of each measured protein was normalized and corrected for gene length. The average protein:mRNA ratio for each gene was then calculated for each condition. To visualize changes in the protein:mRNA ratios for these extremely large data sets, we created a "double ratio" plot where the mRNA ratio for each gene under the two treatment conditions was plotted against the ratios of their protein:mRNA ratios. These data were plotted together with information on peptide counts and ion intensity measurements using the exploratory data analysis program Aabel. This allowed us to readily identify proteins that displayed evidence of posttranscriptional regulation.

The absolute amounts of protein and mRNA expressed in *Shewanella* showed a general correlation at the level of individual genes (C.C. between 0.35-0.37). Binning the data into 25 equal groups ranked by absolute levels of gene expression showed much stronger correlation (C.C. ~0.98), suggesting that in general, the most abundant transcripts give rise to the most abundant proteins. Comparing changes of mRNA and protein expression in response to high and low oxygen conditions yielded similar results (C.C. ~0.43), thus supporting the hypothesis that the dominant mode of regulating protein levels is through transcriptional control. Nevertheless, there were several hundred proteins that showed evidence of posttranscriptional regulation. For example, the relative expression level of ribosomal proteins was essentially invariant despite changes in transcript abundance for several of them. Other proteins involved in regulating mRNA translation displayed a similar pattern. A clear case of post-transcriptional regulation was observed in the operon (*hypAEDCB*) which encodes the NiFe hydrogenase maturation complex. During high oxygen conditions, the mRNA of this operon goes down by about 2-3 fold. However, the corresponding protein levels go down 15-40 fold. Proteins from flanking operons or those that encode the NiFe hydrogenase showed no disparity between transcript and protein levels, suggesting that translation of the polycistronic message encoding HypAEDCB is specifically regulated. Preliminary pathway enrichment analysis of proteins displaying evidence for posttranscriptional regulation shows enrichment of leucine biosynthesis (p-value=1.8e-4), formate to TMAO electron transfer/anaerobic respiration (p-value=0.0023/0.013) and nitrogen metabolism (p-value=0.0085).

Our studies suggest that although most microbial proteins are regulated at the level of transcription, a significant fraction is regulated at the posttranscriptional level, particularly those involved in translation itself. This highlights the needs to include proteomics measurements in the analysis of microbial systems. The integrated approach we have developed for identifying the level at which proteins are regulated should also prove useful in the analysis of more complex microbial systems.

* Presenting author

This research was performed in the Environmental Molecular Sciences Laboratory at the Pacific Northwest National Laboratory (PNNL) and is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the PNNL Foundational Scientific Focus Area.

Early Career Award

158

Quantitative Proteomic Profiling of Site-Specific Redox Modifications, Proteolytic Processing, and N-glycosylation

Dian Su,[1] Jong-Seo Kim,[1] Lu Wang,[1] Uma K. Aryal,[1] Yi Qu,[1] Aaron T. Wright,[1] T.C. Squier,[1] Mary S. Lipton,[1] Richard D. Smith,[1] Ziyu Dai,[2] Scott E. Baker,[2] and **Wei-Jun Qian**[1]* **(weijun.qian@pnnl.gov)**

[1]Biological Sciences Division and [2]Energy and Efficiency Division, Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: One of the primary objectives of this early career research project is to develop novel proteomic approaches to enable quantitative measurements of site-specific regulatory protein posttranslational modifications (PTMs) in subcellular compartments. The ability to effectively and quantitatively characterize site-specific PTMs is essential for understanding the regulation of cellular signaling and protein functions at the post-translational level, and for enabling a systems biology approach to study organisms important for bioenergy or environmental applications. Our efforts have been primarily focused on three important classes of PTMs: (1) reversible redox modifications on cysteinyl thiols, (2) proteolytic processing and protein N-terminal modification, and (3) glycosylation. All three classes of modifications are ubiquitous in both prokaryotic and eukaryotic cells and their important roles in cellular regulation and signaling have been increasingly recognized.

<ins>Reversible redox modifications of protein thiols:</ins> Functional cysteinyl residues in proteins serve as "redox switches" through reversible oxidation, which is recognized as a fundamental mechanism of redox regulation in almost all organisms. We are developing approaches to enable parallel identification and quantification of several types of reversible thiol modifications such as S-nitrosylation (SNO) in response to different cellular reactive oxygen species. The strategy for achieving site-specific proteomic identification is illustrated in **Fig. 1A.** Briefly, thiol specific modifications can be reduced by specific reagents and the converted free thiols are subsequently captured and enriched by a thiol-specific resin and their dynamics can be quantified by isobaric labeling and LC-MS/MS. We have applied this approach to profile SNO in *Synechococcus* sp. PCC 7002. We identified 226 cysteine residues sensitive to SNO modification (**Fig. 1B**) and many of the sites were identified in proteins localized to the thylakoid, indicative of the role of redox regulation in photosynthesis. Similarly, the application of this approach to a fungal organism, *Aspergillus niger*, also resulted in the identification of ~700 SNO-modified sites with many of the proteins being metabolic enzymes.

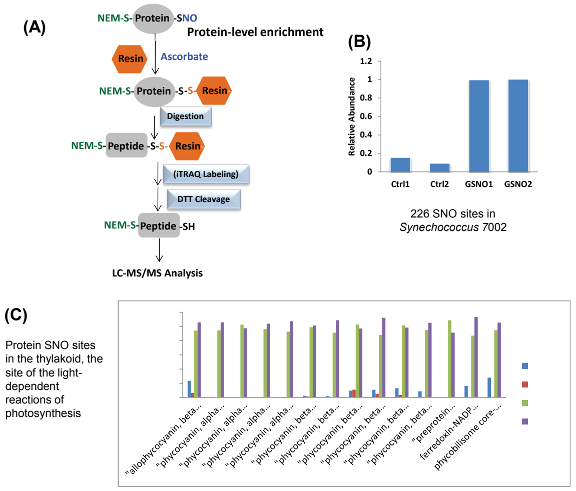


Figure 1. (A) The strategy for identification and quantification of S-nitrosylation (SNO); (B) SNO sites identified in 7002; (C) SNO sites identified in the thylakoid of 7002.

<ins>Proteolytic processing and N-terminal modifications:</ins> Proteolytic processing and *in vivo* N-terminal modifications are another common mechanism of regulation of protein functions. We have developed two complementary approaches to target specific protein N-termini and enrich N-terminal peptides for proteomic identifications. Applying these approaches to a whole cell lysate of *Aspergillus niger*, we were able to identity more than 2000 N-terminal sites with 228 *in vivo* modifications from 858 genes (**Fig. 2**). The number of N-terminal sites identified per gene reveals the complexity of the proteome due to proteolytic processing, which may reveal many novel functional proteins. Quantitative profiling will be performed to identify functional products.

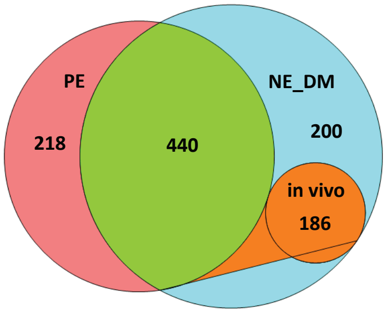


Figure 2. Number of genes identified with N-terminal sites. PE, positive enrichment; NE, Negative enrichment; *in vivo*, proteins with N-terminal sites modified *in vivo*.

<ins>Protein glycosylation</ins>: N- and O-linked glycosylation are known to play an essential role in cellular functions and secretory pathways. We have adapted and optimized of a hydrazide chemistry based enrichment approach for profiling N-glycosites. We have performed an extensive mapping

of N-glycosylated sites in *A. niger* by applying this approach using hydrazide-modified magnetic beads. The optimized protocol was applied to profile N-glycosylated sites from both the secretome and whole cell lysates of *A. niger*. A total of 847 N-glycosylated sites from 330 N-glycoproteins (156 proteins from the secretome and 279 proteins from whole cells)[1] were confidently identified by LC-MS/MS. The identified N-glycoproteins in the whole cell lysate were primarily localized in the plasma membrane, endoplasmic reticulum, Golgi apparatus, lysosome, and storage vacuoles (**Fig. 3**), supporting the important role of N-glycosylation in the secretory pathways. The extensive coverage of N-glycosylated sites and the observation of partial glycan occupancy on specific sites in a number of enzymes provide important initial information for functional studies of N-linked glycosylation and as a prelude to their biotechnological applications in *A. niger*.

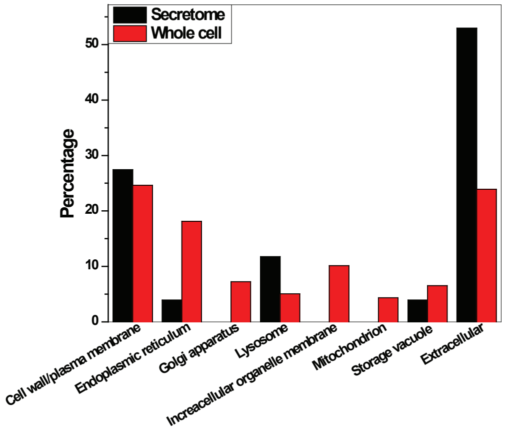


Figure 3. N-glycoproteins identified in cellular compartments based on gene ontology information.

Reference

1. Wang, L., Aryal, U. K., Dai, Z., Mason, A. C., *et al.*, Mapping N-Linked Glycosylation Sites in the Secretome and Whole Cells of Aspergillus niger Using Hydrazide Chemistry and Mass Spectrometry. *J. Proteome Res.* 2011.

This work is supported by a DOE Early Career Research Award under the Office of Biological and Environmental Research in the DOE Office of Science. Mass spectrometry-based proteomics measurements were performed in the William R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by OBER and located at PNNL.

159

Optimization of Microbial Community Proteomics

W.R. Cannon[1]* (william.cannon@pnl.gov), A. Hugo,[1] D.J. Baxter,[2] A. Kalyanaraman,[3] and B.-J. Webb-Robertson[1]

[1]Computational Biology and Bioinformatics Group, [2]Molecular Sciences Computing Facility, Pacific Northwest National Laboratory, Richland, Wash.; and [3]School of Electrical Engineering and Computer Science, Washington State University, Pullman

Project Goals: Analyzing proteomics data from environmental samples is extremely challenging. Generally, the analyses fall into one of two categories, either the complete lack of metagenome sequence information to match proteome data against, or the incomplete nature of metagenome sequence information. The goals of this 3-year SciDAC/GSP project were to alleviate this situation by leveraging high performance computing, advanced modeling and algorithms to develop more accurate and more complete analyses. Achievements 1-3 below address the accuracy issue, and achievements 4-6 address issues regarding complete analysis of metaproteome/metagenome data.

1. **Developed advanced models based on the mathematical overlap between statistical data analysis and statistical thermodynamics [1].**
2. **Implemented the advanced models to bear on problems relevant to DOE's mission in Bioenergy [2].**
3. **Released code for these tools that can be used on multiple computing platforms including workstations, HPC clusters and cloud resources [2-4].**
4. **Developed HPC solutions to the challenge of detecting protein sequence homology for expanding genome and metagenome datasets [5, 6].**
5. **Developed methods to analyze samples when no reference genomes or metagenomes exist [7].**
6. **Developed methods to maximally use metagenome sequences when uncertainty about the gene calls is not small [8].**

These last four achievements are discussed briefly below.

Proteotyping Environmental Samples without a Metagenome. When a metagenome sequence is not available, proteomic analysis is often not possible. This year we reported the development of a novel high performance computing method for proteotyping environmental samples. The method uses computational optimization to provide an effective way to control the false discovery rate. The method provides phylum/species information based on the expressed proteins in a microbial community, and thus complements DNA-based methods. Testing on blind samples demonstrates that the method provides 79-95% overlap with analogous results from searches involving only the correct genomes. Scaling and performance evaluations for the software demonstrates the ability to carry out large-scale optimizations on 1258 genomes containing 4.2M proteins [7].

Proteome Matching against a Metagenome: Bayesian Integration of Evidence. When a metagenome sequence is available, specific identifications of proteins are possible. However, due to the imperfect nature of metagenome sequence information, the protein identification rate is usually quite a bit lower than what is typical for single organism laboratory studies. This need not be the case if multiple levels of evidence are assessed when identifying proteins from an environmental sample. We report the development of a Bayesian framework that incorporates evidence based on peptide detectibilities, significance of MSMS matches, and prior probabilities of protein occurrence. The method more than doubles the number of spectra that are identified with proteins, and increases the number of proteins that are identified by 40-50%.

Proteomics Analysis Code: MSPolygraph. The analyses above were carried out using high performance computing. MapReduce, MPI and serial versions of *MSPolygraph* for peptide identification from mass spectrometry data have been developed. The MPI version runs on any MPI-capable cluster [2]. The MapReduce implementation can run on any Hadoop cluster environment [3]. Availability: The source code along with user documentation is available at omics.pnl.gov/software/MSPolygraph.php.

Remote Homology Detection. Large-scale projects generate millions of *new* sequences that need to be matched against themselves and against already available sequences. For example, the ocean microbiota survey project in 2007 analyzed a total of 28.6M sequences. The most time consuming step during analysis was homology detection, which accounted for 10^6 CPU hours despite the use of fast approximation heuristics such as BLAST.

We developed a novel parallel algorithm, *pGraph,* to efficiently parallelize the construction of sequence homology graphs from large-scale protein sequence data sets based on dynamic programming alignment computation. The parallel design is a hybrid of multiple-master/worker and producer-consumer models, which effectively addresses the unique set of irregular computation issues and input data availability issues. The implementation scales linearly up to 2,048 processors using up to 2.56×10^6 metagenomic protein sequences [6].

Computations were performed in the National Energy Research Scientific Computing Center (NERSC) in Berkeley, CA, and in the Molecular Sciences Computing Facility at the Environmental Molecular Sciences Laboratory (EMSL).

References

1. W. R. Cannon and M. M. Rawlins, "Physicochemical/Thermodynamic Framework for the Interpretation of Peptide Tandem Mass Spectra," *Journal of Physical Chemistry C,* vol. 114, pp. 5360-5366, Apr 1 2010.
2. W. R. Cannon, M. M. Rawlins, D. J. Baxter, S. J. Callister, M. S. Lipton, and D. A. Bryant, "Large Improvements in MS/MS-Based Peptide Identification Rates using a Hybrid Analysis," *J Proteome Res,* Mar 30 2011.
3. A. Kalyanaraman, W. R. Cannon, B. Latt, and D. J. Baxter, "MapReduce implementation of a hybrid spectral library-database search method for large-scale peptide identification," *Bioinformatics,* vol. 27, pp. 3072-3, Nov 1 2011.
4. G. Kulkarni, A. Kalyanaraman, W. R. Cannon, and D. Baxter, "A Scalable Parallel Approach for Peptide Identification from Large-Scale Mass Spectrometry Data," presented at the International Conference on Parallel Processing Workshops (ICPPW 09), 2009.
5. C. Wu, A. Kalyanaraman, and W. R. Cannon, "A scalable parallel algorithm for large-scale protein sequence homology detection," in *39th International Conference on Parallel Processing (ICPP)*, 2010, pp. 333-342.
6. C. Wu, A. Kalyanaraman, and W. R. Cannon, "pGraph: Efficient Parallel Construction of Large-Scale Protein Sequence Homology Graphs," *Revision submission- IEEE Transactions on Parallel and Distributed Systems,* 2012.
7. A. Hugo, D. J. Baxter, W. R. Cannon, A. Kalyanaraman, G. Kulkarni, and S. J. Callister, "Proteotyping of Microbial Communities using High Performance Optimization of Proteome-Spectra Matches," in *Pacific Symposium BioComputing*, In Press, 2012.
8. A. Hugo, B. J. Webb-Robertson, and W. R. Cannon, "Increased Proteome Coverage from Environmental Samples using Bayesian Analysis of Evidence," in *In Preparation*, 2012.

This research was supported by the Genomic Science Program, Office of Biological and Environmental Research, and the Scientific Discovery through Advanced Computing program, Office of Advanced Scientific Computing Research, U.S. Department of Energy.

160

PUNCS: Towards Predictive Understanding of Nitrogen Cycling in Soils

Robert Sanford[1]* (rsanford@illinois.edu), Allana Welsh,[1] Kostas Konstantinidis,[2] Joanne Chee-Sanford,[3] Kirsti Ritalahti,[4] Sukhwan Yoon,[4] and **Frank Löffler**[4,5]

[1]University of Illinois, Urbana-Champaign; [2]Georgia Institute of Technology, Atlanta; [3]USDA-ARS, Urbana, Ill.; [4]University of Tennessee, Knoxville; and [5]Oak Ridge National Laboratory, Oak Ridge

Project Goals: In anoxic environments, nitrate is largely consumed by dissimilatory nitrate reduction to ammonium (DNRA) and nitrate reduction to N_2O and N_2 (denitrification). DNRA oxidizes more C per mole of nitrate than denitrification and generates a cation (NH_4^+), which is retained in most soils, indicating that these processes have profoundly different impacts on N retention and greenhouse gas (CO_2, N_2O) emissions. Microbes capable of DNRA or denitrification coexist but the regulatory controls for these competing processes are unknown, and their relative contributions cannot be predicted faithfully with current models. This paucity of information limits the development of more accurate, predictive models of N-flux including the effects of N-retention on plant

growth and biomass yield for bioenergy production, as well as greenhouse gas (e.g., N_2O) emissions.

To elucidate the environmental factors controlling N-turnover in soils, the project team will conduct physiological studies, stable isotope probing (SIP) experiments, (meta)genomics and (meta)transcriptomics analyses, along with high resolution imaging (NanoSIMS) on systems of increasing complexity (i.e., pure cultures to mesocosms established with distinct soil types). We will quantitatively monitor genes and transcripts related to N-turnover in response to changes in pH, temperature, soil moisture, C- and N-content and assess the emitted gases using a mass balance approach based on (geo)chemical and stable isotope measurements. By comparing the gene-centric and genomic (*who is there?*) to the transcriptomic, SIP, and NanoSIMS (*who is how active?*) datasets gathered from different treatments, ***system-level insights into the pathway controls and the functional redundancy within microbial communities controlling N-flux in soils will be obtained***.

One project task determines the pathway controls for nitrate depletion in anoxic soil environments; denitrification versus DNRA. Previous observations suggested that the carbon to nitrogen (C:N) ratio controls the fate of nitrate to gaseous products (N_2O, N_2) or NH_4^+. *Shewanella loihica*, a unique bacterium with both complete denitrification and DNRA pathways, is used to explore the environmental factors (e.g., C:N ratio, type of C available, pH, temperature) that determine which N-oxide reduction pathway predominates. The results obtained to date suggest that the C:N ratio is one of several factors that influence the fate of nitrate. DNRA appears to be favored over denitrification under conditions of low nitrate:nitrite ratios and high C:N ratios. Further, nitrate reduction to ammonium predominated in mineral medium augmented with amino acids. Gene expression studies are being used to monitor the expression of DNRA and denitrification genes under these different growth conditions.

Another project goal is to better understand the diversity of *nosZ* genes involved in N_2O reduction to N_2. Genome analysis of the nitrite-to-NH_4^+-reducing, non-denitrifying bacterium *Anaeromyxobacter dehalogenans* strain 2CP-C revealed the presence of a complete *nosZ* gene cluster. Subsequent physiological studies corroborated that this organism uses N_2O as a growth-supporting electron acceptor. The denitrifier- and *Anaeromyxobacter*-types of *nosZ* genes share sequence similarity; however, the primers used for environmental surveys of denitrifier *nosZ* have failed to detect the *Anaeromyxobacter*-type *nosZ*. PCR primers specifically targeting the *Anaeromyxobacter*-type *nosZ* determined its distribution in different soil ecosystems, suggesting that an important, yet unrecognized N_2O sink exists. Recent isolation and sequencing efforts revealed an unexpected diversity of non-denitrifying N_2O reducers and *nosZ* genes in soil ecosystems suggesting that current N_2O emission models are missing a possible significant N_2O sink.

To elucidate the relative abundance of genes implicated in nitrate, nitrite and N_2O transformation in soils, we have embarked on surveys of existing metagenome datasets. The datasets incorporate millions of short-read sequences (e.g., shorter than 400 bp), obtained using either the Illumina or the Roche 454 sequencers. Therefore, the first objective of our work is to develop the bioinformatics pipeline that will allow us to reliably identify and align short fragments of the target genes recovered in the available metagenomes. Subsequently, we will evaluate in-silico the specificity of available primers for chosen target genes and design new primers for microbial groups that are not encompassed by the primers currently available. These approaches will enable the identification and enumeration of keystone microbial groups that respond to different incubation conditions and perturbations in mesocosm systems. We have also validated and applied Illumina Hi-Seq 2000 sequencing on soil samples, and have established the bioinformatics approaches to assemble such high-volume data (>50Gb per sample) and allow comparisons of different communities based on metagenomic datasets. We will report on recent developments including preliminary results from metagenome sequencing efforts.

This research is supported by DOE Genomic Science Grant DE-FG02-11ER65267.

161

Spatioelectrochemistry: The Molecular Basis for Electron Flow Within Metal-Reducing Biofilms

Daniel R. Bond[1]* (dbond@umn.edu), Ying Liu,[1] **Caleb Levar**,[1] and Janet Rollefson[2]

[1]Biotechnology Institute and Department of Microbiology and [2]Department of Biochemistry, Biophysics, and Molecular Biology, University of Minnesota, Twin Cities

Project Goals: Electrochemical, spectral, genetic, and biochemical techniques have provided evidence that multiple redox proteins and structural macromolecules outside the cell work together to move electrons long distances between *Geobacter* cells and to metals. This extracellular matrix contains many stable proteins not easily solubilized by standard methods, in addition to complex polysaccharides. Thus, much of these components were likely lost or ignored in previous proteomic and biochemical surveys. We aim to define this matrix and develop new tools to visualize it in action. The goals of this project are to 1) identify protein and polysaccharide elements crucial to both the assembly and function of the extracellular conductive matrix, 2) expand spectroelectrochemistry techniques to define the mechanism and route of electron transfer through the matrix, and 3) we will combine this knowledge of matrix proteins and their role in multicellular electron transfer to visualize redox and gene expression gradients in space over time.

When bacteria change the state of metals in the environment, they transport electrons unprecedented distances from intracellular metabolic reactions to distant mineral surfaces. This electron movement drives subsurface bioremediation, controls aquifer chemistry, and powers new microbial energy generation applications. Yet a molecular understanding of how this electron transfer is accomplished by *Geobacteraceae,* who are among of the most predominant bacteria in such systems, remains one of the grand challenges in microbial environmental processes.

Direct measurements of living biofilms using electrochemistry has revealed electron hopping between *Geobacter* redox proteins to be a rate-controlling step at all stages of growth. Direct spectral analysis of living biofilms has confirmed that *c*-type cytochromes are a major reservoir of charge in these films, and that these cytochromes experience a bottleneck to oxidation when electrons must be transferred longer distances. Fine-scale immunogold labeling has discovered gradients in cytochrome abundance throughout these films, further suggesting the presence of gradients within this biofilms. Genetics has discovered new polysaccharide biosynthesis operons, and secretion systems essential for the attachment of *Geobacter* to metals and other cells in the biofilm. Biochemistry has shown that the extracellular space acts anchors a diverse assemblage of essential *c*-type cytochromes and adhesion proteins outside the cell. We hypothesize that this data converges on a model of electron transfer mediated by multiple cooperating proteins, ultimately limited by hopping between a highly adaptable network of cytochromes, which are attached to polysaccharides and pili.

Our project hypothesizes that it is the components of this matrix, and its high adaptability, that explains why *Geobacter* can so easily interface with a wide range of mineral surfaces, grow as multicellular networks, and quickly adapt to disruptions in single cytochromes.

To accomplish these goals, we are screening mutant libraries for strains defective in different developmental stages of this matrix using a combination of traditional and high-throughput approaches, and developing biochemical extraction methods for separating and analyzing this matrix. Key to this phase is to identify proteins required for specific developmental stages such as surface recognition, vs. self-recognition, or cell-surface electron transfer vs. cell-cell electron exchange. In parallel, we are inventing spectroelectrochemistry cells able to monitor cytochrome redox states in living biofilms, and engineering proteins for spatiotemporal localization of activity. This multidisciplinary approach aims to link quantitative data for specific reactions that occur at the biotic-abiotic interface with genes and expression patterns that can be used in predictive modeling, environmental monitoring, and design of bacteria with altered conductive properties.

162

Metabolic Dynamics of Starvation Induced Dormancy in *Saccharomyces cerevisiae*

Niels Klitgord[1]* (njklitgord@lbl.gov), Fredrik Alsin,[1] Theodore Peters,[2] Matt Youngblood,[1] Robert E. Hughes,[2] and **Trent Northen**[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif. and [2]Buck Institute, Novato, Calif.

Project Goals: To understand the metabolic mechanism involved in the initiation and departure of starvation induced dormancy in yeast.

Many organisms have adapted evolutionary to dealing with various stresses by entering a state of dormancy. Organisms that enter this dormancy typically make physiological changes that allows them to shut down their metabolism, and become resistant to various environmental perturbations including dessication, pH, and heat stresses. However the specific environmental, metabolic and physiological changes an organism undergoes as it enters and leaves dormancy is not well understood.

Here we use the yeast *Saccharomyces cerevisiae* to address the question of what metabolic changes occur as they enter a starvation induced dormancy. We attempt to answer this question by using gas-chromatography mass-spectrometry (GC-MS) to profile the intracellular and extracellular metabolites present during both log and stationary phase of growth. By combing this data with the genome scale metabolic model of yeast, we are able to follow the extracellular dynamics of metabolic usage and production as they enter log-phase, and transition into stationary phase. Furthermore, by identifying changes in intracellular metabolite concentrations, we are able to identify which metabolic pathways must change as the yeast enter stationary phase.

Our results provide some some suggestions metabolic mechanisms of dominance, which allow for the development of specific hypothesis to test. Answering these questions might allow for the engineering solutions for enhanced dormancy in plants, animals or humans, as well the development of anti-dormancy agents to target deleterious organisms such disease causing bacteria.

This work 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.

163

Engineering Specialized Metabolism in a Single Cell Type

Anna Berim,[1] Adam Schmidt,[1] B. Markus Lange,[1] Michael Scanlon,[2] Carol A. Soderlund,[2] and **David R. Gang**[1]* (gangd@wsu.edu)

[1]Washington State University, Pullman; [2]Cornell University, Ithaca, New York; and [3]The University of Arizona, Tucson

Project Goals: The long term goal is to significantly increase our understanding of mechanisms that regulate and control specialized metabolism in single cell types of plants and to use that understanding to rationally engineer plants to produce desired metabolite profiles. The glandular trichome secretory cell is our model in vivo cell system. This project is divided into two focus areas. The first focus area involves development and testing of a model of metabolic partitioning between the MEP/terpenoid and shikimate/phenylpropanoid pathways in glandular trichome secretory cells. This model will be used to predict regulatory and control points in the large metabolic network, which then could be targets for modification to achieve desired alteration of metabolism in this single cell type. The second focus area will be the evaluation of the roles of specific protein modifications and of changes in their expression on carbon partitioning and metabolic flux in real world scenarios.

Mathematical Model Development and Data Collection

A number of advances were made to a previously generated kinetic mathematical model of peppermint essential oil biosynthesis. In particular, the Lange laboratory was able to demonstrate that the density and size distribution of glandular trichomes on peppermint leaves was the most important factor determining essential oil yield in various genotypes (wild-type and several transgenic plants) and under different adverse environmental conditions[1]. We considered using an analogous approach to model essential oil biosynthesis in basil, but there are critical gaps in our understanding of the physiology of trichome development and the organization of the terpene/phenylpropanoid pathways. Moreover, results over the course of the past year have supported the contention that significant amounts of non-volatile (ergo non-essential oil) metabolites accumulate in the glandular trichomes of basil. Based on these data, we are assembling a reaction graph representation of the pathways relevant to essential oil formation in sweet basil. We are using a Flux Balance Analysis (FBA) approach with one main difference (and key innovation) compared to existing FBA studies: we do not assume that biomass production is geared toward optimal growth (which works well in microorganisms but there are many problems with more complex organisms) but instead we are focusing our modeling on glandular trichomes. We are assuming the uptake of a transport sugar (stachyose/sucrose) and the maximizing of the production of essential oil and its removal from the network (accumulation in storage cavity of glandular trichome) as the objective function. These modeling activities will provide us with essential information regarding the flux distribution across multiple pathways in glandular trichomes. In addition, this approach does not require as much detailed kinetics data for the network elements as a kinetic model would require.

Advances in metabolite network determination

Work on characterizing the enzymes involved in production of the large array of methoxylated flavones in basil and peppermint has moved ahead very well. Because flavones, mostly highly methoxylated, constitute a considerable carbon sink both in basil and peppermint, which potentially competes with the isoprenoid pathway for carbon allocation, and the mechanisms of A-ring modifications have been barely studied hitherto, we have worked to characterize the late metabolic steps leading to the formation of these compounds. This information will be required for development and refinement of the model of metabolism in the secretory cells of basil. The functions of a number of enzymes involved in the formation of these compounds have been determined from basil and peppermint. For example, recombinant FOMTs display distinct substrate preferences and product specificities that can account for most detected 7-/6-/4'-methylated, 8-unsubstituted flavones in sweet basil. Apparent K_M values in the low micromolar range and specific gene expression profiles support the involvement of specific FOMTs in the biosynthesis of specific flavones in the different sweet basil lines. Structure homology modeling suggested the involvement of several amino acid residues in defining the proteins' stringent regioselectivities. The roles of these individual residues were confirmed by site-directed mutagenesis. A parallel study of flavone A-ring hydroxylases allowed us to delineate the network from apigenin to salvigenin, gardenin B and nevadensin, the major polymethoxylated flavones that accumulate in sweet basil.

Laser-microdissection of developing glandular trichome secretory cells for functional analysis and model refinement

The goal of this part of the project is to measure gene expression in glandular trichome secretory cells at specific developmental stages. This information is important in order to assign functions to the trichomes and not just leaf cells in general, and also for development and testing of the later mathematical models. We are using the technique of laser-microdissection coupled with RNA-sequencing to isolate specific cell types from sectioned tissue and to profile and quantify transcripts in these cells. Transcript profiles are being compared in peppermint and four basil chemotypes at four defined developmental stages from trichome initiation to fully expanded trichomes.

Reference

1. Rios-Estepa et al. (2010) Mathematical modeling-guided evaluation of biochemical, developmental, environmental and genotypic determinants of essential oil composition and yield in peppermint leaves. Plant Physiol. 152, 2105-2119

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

164

Characterization of Metabolites from Phototrophic Microbial Mats in Biogeochemically Unique Environments

Young-Mo Kim[1]* (young-mo.kim@pnnl.gov), Stephen R. Lindemann,[1] David M. Ward,[2] Sergey Stolyar,[1] Thomas O. Metz,[1] **Margaret F. Romine**,[1] and **James K. Fredrickson**[1]

[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash. and [2]Department of Land Resources and Environmental Science, Montana State University, Bozeman

Project Goals: The objective of the Pacific Northwest National Laboratory Foundational Scientific Focus Area is to develop a predictive, genome-enabled understanding of how microbial interactions impart stability, robustness, and functional efficiency to microbial communities. To achieve this goal, research is being conducted on three different model systems: iron- and sulfur-oxidizing communities associated with hyperthermophilic acidophilic springs at Yellowstone National Park (YNP) and phototrophic mat communities formed in high temperature springs at YNP or in a hyper-saline lake (Hot Lake) found in northern Washington State. As part of this effort, we are analyzing intracellular and extracellular metabolites to identify those that are secreted by microbes present in these phototrophic mat systems and to monitor their exchange among different community members.

The phototrophic microbial mats in YNP hot springs have been well-studied. However, the fundamental relationships between the primary producers (autotrophs) and consumers (heterotrophs) in these mats are not fully understood. Oxygenic phototrophic cyanobacteria, such as *Synechococcus* species, produce organic molecules by fixing inorganic carbon via photosynthesis or by fermenting stored molecules at night. For example, previous research has revealed that glycolate is produced during the day, while acetate is produced from the fermentation process at night. In turn, anoxygenic phototrophic bacteria, such as *Chloroflexus* and *Roseiflexus* species, and other heterotrophic organisms utilize these primary molecules for their carbon and energy needs. In this system, the metabolic processes of different community members contribute to the accumulation or consumption of storage molecules such as polyhydroxyalkanoic acids (PHAs) and wax esters for carbon storage and cyanophycin for nitrogen, which are thought to undergo diel cycling in *Roseiflexus* and *Chloroflexus*. PHAs are also possibly precursors for the synthesis of branched amino acids in certain bacterial species which have no enzymes to facilitate these processes. Using gas chromatography-mass spectrometry (GC-MS)-based metabolomics analyses, we have identified several key metabolites in this system. Some of these metabolites fluctuate during a diel cycle and may be exchanged between members of hot springs microbial mat communities. We will present results of metabolomics analyses of the mats, as well as the results of spent media analyses from the lab culture of *Thermosynechococcus* and *Chloroflexus* species which have been isolated from these systems.

The hyper-saline and ecologically unique Hot Lake in north-central Washington state has not been well-investigated by the scientific community. Recent studies have shown that several autotrophic and heterotrophic microorganisms comprise mats within this extreme environment, which contains almost 1 M magnesium sulfate and 0.5 M sodium chloride. Since the location of this lake is geographically isolated from any river or stream, organic carbon and nitrogen sources are not able to enter from outside sources. Therefore the majority of water input in this lake comes from precipitation, causing the lake to undergo fluctuations in salinity. The phototrophic mats in this system represent good models of energy and carbon fluxes. Importantly, the high levels of salts in this system present a challenge to the analysis of organic metabolites, which usually have similar size and polarity as those of salts. For example, high concentrations of salts may reduce extraction and derivatization efficiencies during GC-MS and generate excessive adduct ions in liquid chromatography-mass spectrometry (LC-MS) analyses. Therefore, we are developing and optimizing protocols to remove these salts to improve our measurements. We will present our on-going mass spectrometry-based metabolomics analysis of mat samples and secreted metabolites from cultures of autotrophic and heterotrophic organisms isolated from the Hot Lake community.

Results obtained from the above mentioned studies will be incorporated with data from transcriptomics and proteomics experiments and used for modeling the metabolism of the microbial mat communities.

This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area. Metabolite measurements were performed in the William R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by OBER and located at PNNL. Pacific Northwest National Laboratory is operated for the DOE by Battelle Memorial Institute through Contract No. DE-AC05-76RLO 1830.

165

Carbon Flow in Lithotrophic Acid Hot Springs Microbial Communities, Yellowstone National Park

Helen Kreuzer[1]* (Helen.kreuzer@pnnl.gov), James Moran,[1] Ryan Jennings,[2] Laura Whitmore,[2] Mark Kozubal,[2] Bill Inskeep,[2] and **Jim Fredrickson**[1]

[1]Pacific Northwest National Laboratory, Richland, Wash. and [2]Montana State University, Bozeman

Project Goals: The objective of the Pacific Northwest National Laboratory Foundational Scientific Focus Area is to develop a predictive, genome-enabled understanding

of how microbial interactions impart stability, robustness, and functional efficiency to microbial communities. To achieve this goal, research is being conducted on both natural mat communities and defined co-cultures designed to test hypotheses regarding processes naturally occurring in these communities. As part of this effort, we are elucidating carbon flow in the lithotrophic microbial communities in acid thermal springs in Yellowstone National Park, focusing particularly on the roles of autotrophy and heterotrophy in the communities.

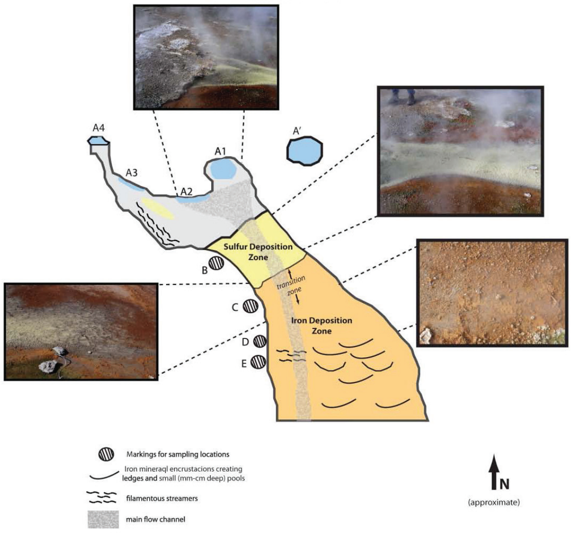


Figure 1. Beowulf spring in Norris Geyser Basin

Norris Geyser Basin is home to several acidic (pH ~3), sulfidic, hydrothermal (source temperature >65 ˚C) springs, of which Beowulf, Grendel, and OSP Springs are examples. Microbial communities inhabit these springs that are based on chemolithotrophy as evidenced by the oxidation of sulfide and ferrous Fe and subsequent deposition of oxidation products. What remains unclear, however, is how carbon is initially introduced into the systems; whether by autotrophy, heterotrophy, or a combination route. We are exploring carbon cycling in chemotrophic springs in the Norris Geyser Basin to better understand both the initial carbon entry into the system and how fixed C is exchanged in the microbial community. Our immediate goals include (1) determining whether autotrophy plays a major role in carbon incorporation, focusing initially on *Metallosphaera yellowstonensis*, a community member whose genome includes genes from the 3-hydroxypropionate 4-hydroxybutyrate carbon fixation pathway, and (2) compiling a stable isotope inventory of an example spring (Beowulf, Figure 1) and the surrounding landscape to help identify additional carbon sources.

By understanding how carbon is incorporated into and exchanged among microbial community members, we will gain insights into how microbial communities adapt to these extreme (high temperature, acidity, high sulfide and low biomass) environments.

Visual evidence reveals the two dominant metabolisms, shown in main Beowulf flow channel (Figure 1). Sulfide oxidation converts dissolved sulfide into elemental sulfur in zone B of the spring. The orange precipitates downstream of the sulfur deposition zone result from iron oxidizer activity.

M. yellowstonensis was isolated from Norris Geyser Basin and represents up to 20% of the total microbial nucleic acid associated with the iron oxide mats. *M. yellowstonensis* contains genes consistent with its presumed iron oxidizing metabolism.[1] Further, *M. yellowstonensis* has genes associated with the 3-hydroxypropionate, 4-hydroxybutyrate carbon fixation pathway. We sought to determine whether *M. yellowstonensis* actually does fix CO_2, and whether autotrophy occurs in the spring communities.

Pure cultures of *M. yellowstonensis* were incubated in mineral medium with or without yeast extract, and with un-labeled or ^{13}C-labeled CO_2 in the headspace. Ground pyrite was supplied as the electron donor, and O_2 was the electron acceptor. Cultures were incubated approximately two weeks, and the carbon isotope ratio of the acid-washed biomass was measured by isotope ratio mass spectrometry. When $^{13}CO_2$ was present, the isotope ratio of the biomass contained significantly more ^{13}C than when unlabeled CO_2 was present, demonstrating that *M. yellowstonensis* fixes CO_2. Much less ^{13}C was incorporated when yeast extract was present, suggesting that *M. yellowstonensis* grows heterotrophically in the presence of suitable organic substrates. Initial estimates of the CO_2 –> biomass fractionation factor (~3.5 ‰) are consistent with the hydroxybutyrate fixation pathway[2] and very similar to that measured for *M. sedula* (3.1 ‰).[3]

Samples of iron-oxidizing microbial mat from two similar Norris Geyser Basin springs, Grendel and OSP, were excised and placed into bottles containing pyrite and the mineral medium used in the laboratory experiments. Unlabeled CO_2 or $^{13}CO_2$ was injected into the headspace, and the cultures incubated fourteen days at 65 -70 ˚C. Following incubation, the C isotope ratio of the acid-washed mat material was measured by isotope ratio mass spectrometry. Data clearly demonstrate $^{13}CO_2$ fixation by the microbial communities in both springs.

Our data thus far indicate that there may be additional carbon inputs to the iron mat community. The bulk carbon isotope ratios we have measured in the iron oxide mats range from ~=18 to ~-21 ‰. Based on the approximate fractionation factor measured in the laboratory, autotrophic fixation of dissolved inorganic carbon by *M. yellowstonensis* would yield fixed carbon with a predicted isotope ratio of ~-7.5 ‰. A simple metabolic model for the iron mats, in which carbon is fixed by *M. yellowstonensis* and then passed through a heterotrophic food chain (which typically results in little isotopic fractionation) is not consistent with the measured bulk carbon isotope ratio in iron oxide mat samples. Alternative hypotheses to explain the bulk carbon isotope ratio of the mats include (1) the presence of significant landscape carbon (e.g., plant detritis) in the mats, or (2) direct heterotrophy of landscape carbon by members of the iron mat community. The average carbon isotope ratio of organic material in the landscape (plant material, animal dung) is ~ -26 to

-29 ‰, and that of dissolved organic carbon in Beowulf Spring is ~ -22 to -23 ‰. A mixture of passively present landscape carbon, or its heterotrophic uptake, in combination with carbon fixation by *M. yellowstonensis*, could explain the observed bulk isotope ratios. In the coming year, we will be working to associate carbon substrates with different community members by measuring isotope ratios of the different phylogenetic groups, and to further elucidate the flow of carbon through the iron mat community.

References

1. Kozubal, M.., et al, 2011. *Appl. Environ. Microbiol.* 77(5):1844.
2. Berg, I., et al., 2010. *Nature Rev. Microbiol.* 8:447.
3. House, C., et al., 2003. *Org Geochem*, 2003 34:345.

This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

166

Metabolic and Community Modeling of Phototroph-Heterotroph Community Interactions

William Cannon[1]* (william.cannon@pnl.gov), Gretta Serres,[2] Sebastian Jaramillo,[1] and Don Bryant[3,4]

[1]Computational Biology and Bioinformatics Group, Pacific Northwest National Laboratory, Richland, Wash.; [2]Marine Biological Laboratory, Woods Hole, Mass.; [3]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park; and [4]Department of Chemistry and Biochemistry, Montana State University, Bozeman

Project Goals: Understanding interactions between microbes in the environment is a daunting task. Because of the complexity of such interactions, it is necessary to use computational models to track the data and proposed interactions, and to develop intuition, hypotheses and predictions. With this in mind, the goals of the modeling component of the PNNL FSFA are two-fold. First, in the short-term we will evaluate the existing metabolic model of *Synechococcus* sp. PCC 7002 by incorporating metabolomic and proteomic data. We will initially utilize the model to evaluate whether the newly discovered complete but alternative cyanobacterial TCA cycle maximizes the utilization of 2-oxoglutarate for production of biosynthetic precursors for growth, and to evaluate the growth and production of extracellular metabolites. Second, in the long term, we will more generally identify the key metabolic interactions between the autotrophic and heterotrophic populations of Yellowstone hot spring mats and consolidate the quantitative data describing these interactions within a context of extended phototroph–heterotroph model, including metabolite predictions.

Building useful models requires several tasks. First, initial models are constructed based on genome sequences. These initial models only tell us about metabolic potential, rather than actual metabolism because genome annotation is not complete and regulatory information is missing. Enzymes and transporters that carry out reactions of interest may not be expressed under all conditions, predicted metabolites may not actually be produced, and more frequently unpredicted metabolites are observed. Thus a second critical aspect to building useful models is to have data analysis tools that analyze the data thoroughly. Multiple data types—proteomics, metabolomics and transcriptomics to name a few—may then be interpreted with respect to the model. Based on the data, the model is adjusted. This is the process frequently referred to as data integration. Finally, the model is curated to the point at which simulations can be carried out, the output of which can be directly compared to experimental data.

To enable process of model building, the PNNL SFA has developed 11 pathway genome databases of phototrophic and heterotrophic species being studied on this project. These databases are linked to external pathway databases of the same organisms for comparative studies. Using this platform, experimental data can be interpreted and used to refine the model.

Using this approach, we evaluated the protein complement of *Synechococcus* spp. 7002 grown under conditions of phototrophic growth under atmospheric CO_2 concentrations. Not surprisingly, pathways related to photosynthesis and carbon fixation were identified most frequently. However, initial analysis implied that a specific ICT family CO_2 transporter, SYNPCC7002_A0690, was not expressed, implying that either the annotation was incorrect or an alternative transporter is used. Since transporters set the boundary conditions that determine the energetic feasibility of metabolic processes, such adjustments to the metabolic model are critical. However, a new, highly accurate computational method that we applied found that the ICT transporter SYNPCC7002_A0690 was not only expressed but appears to be one of the most highly expressed proteins in the cell [1].

Pathway databases also serve as sources of hypotheses. Using pathway databases, we have begun to investigate pathways that have been inferred to be incomplete by computational methods. Investigation into the incomplete TCA cycle of cyanobacteria has demonstrated that, in fact, the cycle is complete and uses an alternative enzyme for the conversion of 2-oxoglutarate to succinate [2]. A principle difference between the standard pathway and the newly discovered alternative pathway is that the formation of succinate is no longer coupled to the highly favorable reaction involving coenzyme-A. Due to the removal of this highly favorable step, we hypothesize that the alternative pathway leads to an increased steady-state level of 2-oxoglutarate. This could be significant in that 2-oxoglutarate is the precusor used for assimilation of nitrogen, and the *Synechococcus* species in the Yellowstone mat system have been observed to dramatically increase nitrogen fixation during a brief 2-hour period in the morning. This burst of activity may serve as a nitrogen source for other members of the microbial community, and yet it is not clear how such a dramatic increase in nitrogen fixation can occur in such a short time period. Both laboratory and simulation based tests of these hypotheses are being conducted.

The simulation model starts with the framework of the previously determined genome-scale and flux-based model of *Synechococcus* sp. PCC 7002. We will model the TCA cycle in kinetic detail by incorporating aspects of the kinetic model by Singh and Ghosh. This will allow us to predict 2-oxoglutarate levels for both the standard and alternative TCA cycles under multiple conditions.

References

1. W. R. Cannon, M. M. Rawlins, D. J. Baxter, S. J. Callister, M. S. Lipton, and D. A. Bryant, "Large Improvements in MS/MS-Based Peptide Identification Rates using a Hybrid Analysis," *J Proteome Res,* Mar 30 2011.
2. S. Zhang and D. A. Bryant, "The Tricarboxylic Acid Cycle in Cyanobacteria," *Science,* vol. 334, pp. 1551-1553,2011. http://www.sciencemag.org/content/334/6062/1551.abstract

This research is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

167

Metabolic Coupling and Carbon Flux Regulation in Phototroph-Heterotroph Associations

Alex S. Beliaev* (alex.beliaev@pnnl.gov), Thomas C. Squier* (thomas.squier@pnnl.gov), Diana J. Bigelow, Grigoriy E. Pinchuk, Sergey M. Stolyar, Baowei Chen, Na Fu, Yijia Xiong, E.A. Hill, and **Jim K. Fredrickson**

Pacific Northwest National Laboratory, Richland, Wash.

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. As part of the FSFA cross-cutting science and technology theme, this research component elucidates pathways and regulatory mechanisms that govern metabolic exchange and interactions between organisms in phototrophic associations and contribute to their stability and efficient utilization of resources.

Scientific Concept: Secretion of low-molecular weight metabolites and biopolymers drives interactions between primary producers (autotrophs) and consumers (heterotrophs) and, at the organism level, is a function of environmental parameters such as light intensity, O_2 levels, and nutrient concentrations. Under photosynthetic conditions, O_2 is one of the key factors affecting carbon metabolism, growth, and survival of microorganisms. We hypothesize that photoautotroph-heterotroph associations have developed mechanisms to spatially and transiently decrease O_2 concentrations thereby lowering oxidative stress. Under dark conditions, the phototrophs ferment glycogen providing the heterotrophs with electron donor/carbon source to carry out anaerobic respiration which affects solubility of critical micronutrients such as Fe and Mn, thus making these elements more accessible for phototrophs. At a sub-cellular level, the interactions are mediated by multi-level regulatory networks that include redox-active proteins which sense environmental change through specific post-translational modifications and act as key control points to optimize metabolic efficiencies at both organism and community levels. To that end, we seek to identify regulatory proteins responsive to environmental change throughout the diel cycle, which includes redox-dependent disulfide bond formation and oxidative modifications to Met and Tyr (oxidative modifications are reversible through the action of repair enzymes). We hypothesize that light-responsive disulfide bond formation acts to control the functions of key proteins that regulate metabolic fluxes. Such adaptive regulation is suggested to diminish formation of reactive oxygen species (ROS), reduce protein oxidative damage, and enhance metabolic efficiencies. *In vivo* examples have been identified involving an inverse relationship between redox-dependent cystine formation and the abundance of site-specific oxidative modifications. As these post-translational modifications are reversible, they have the potential to rapidly reprogram metabolism through the control of specific branch points within metabolic pathways.

Our current line of research focuses on gaining an understanding of the pathways and regulatory mechanisms associated with the collective energy, carbon, and nutrient processing in axenic and mixed phototroph-heterotroph cultures. Systems under investigation include *Synechococcus* sp. PCC 7002 – *Shewanella putrefaciens* W3-18-1 co-culture, individual organisms (*e.g., Synechococcus* spp., *Thermosynechococcus* sp., *Roseiflexus castenholzii*, *Chloroflexus* spp.), and consortia derived from, hot spring communities (e.g., *Thermosynechococcus* sp. N55-*Roseiflexus castenholzii*). This research also utilizes advanced controlled cultivation capabilities to enable investigation of redox-dependent control mechanisms.

Baseline growth parameters and the presence of regulated metabolic coupling between *Synechococcus* and *Shewanella* spp. have been identified, as both organisms successfully grow as a binary co-culture, but not as separate monocultures, using either lactate or CO_2 as the sole source of carbon. During growth with lactate, O_2-dependent oxidation of lactate mediated by W3-18-1 produced sufficient amounts of CO_2 to maintain photosynthetic growth of 7002, which in turn, generates the necessary O_2 for growth of W3-18-1. These results establish that there is a tight metabolic coupling between the phototrophic and heterotrophic organisms using light as the only source of energy and lactate as the sole source of carbon. Growth of the co-culture under these conditions results in acetate accumulation by *Shewanella* W3-18-1 thus revealing an imbalance is caused by differences in growth rates where the O_2 consumption by the heterotroph exceeds the O_2 evolution rates by the cyanobacterium. When grown on CO_2 as the sole source of carbon, the metabolic coupling between 7002 and W3-18 is established through secretion of dissolved organic carbon compounds by the photoautotroph. NMR-based techniques identified formate, lactate, and acetate secreted by *Synechococcus* 7002 which may serve as the primary carbon and energy sources for *Shewanella* W3-18-1. Secretion of extracellular metabolites and/or biopolymers also appears to be common for some thermophilic cyanobacterial species, e.g., *Thermosynechococcus* sp. In hot spring phototrophic (HSP) mat communities major shifts in energy metabolism occur in response to increases in O_2 involving: *i*) alterations in metabolic exchange between community members through the release of acetate, and *ii*) diminished rates of carbon fixation due to photorespiration that may be associated with an uncoupling of RuBisCo and futile cycling involving glycolate production. These results indicate that enhanced metabolic efficiencies in communities of phototrophs and heterotrophs may arise, in part, due to increased efficiencies of carbon fixation as a result of reductions in photorespiration due to decrease in O_2 levels.

Additional cellular responses that impact community stability were found using chemical probes; they involve regulatory mechanisms that shift metabolic flux to minimize the oxidative damage to cellular proteins in a manner that will enhance metabolic efficiencies. Large decreases in the overall levels of protein oxidation are observed for cultures of 7002 grown in the presence of the W3-18-1, where the majority of oxidatively sensitive proteins in W3-18-1 are protected from oxidative modification in the co-culture. Reductions in oxidative stress are apparent despite substantially higher O_2 present in the co-culture (160% dissolved air saturation) in comparison to axenic cultures (44% dissolved air saturation) (both cultures are grown in caged photobioreactors using white light intensities of 240 μM photons m^{-2} s). These results support the hypothesis that opportunistic interactions between heterotrophic (*Shewanella*) and photosynthetic (*Synechococcus*) organisms permit metabolic coupling to enhance energy efficiencies and community stability. These measurements are consistent with prior indications that axenic isolates of the *Synechococcus* isolated from the HSP mats have a substantially enhanced sensitivity to light-induced oxidative stress in comparison to the natural mat community, indicating an importance of metabolic coupling between community members in the mat that promote stress resistance.

This research is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

168

Microbial Diveristy and Biogeochemical Function of the Phototrophic Microbial Mats of Epsomitic Hot Lake, Washington

Stephen R. Lindemann* (stephen.lindemann@pnnl.gov), James J. Moran* (james.moran@pnnl.gov), Alice Dohnalkova, David W. Kennedy, Sergey M. Stolyar, Paul D. Majors, H. Steven Wiley, Allan E. Konopka, and **James K. Fredrickson**

Pacific Northwest National Laboratory, Richland, Wash.

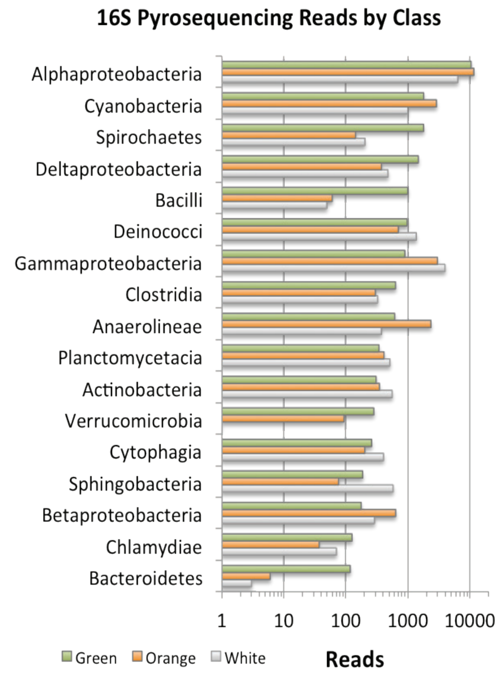
Project Goals: Determine the phylogenetic and functional diversity of members of a hypersaline microbial mat using (meta)genomics of whole mat, enrichment cultures, and sorted single cells; probe the dynamics of community function and metabolic interactions in a hypersaline microbial mat in response to changes in salinity, temperature, and photon flux using metatranscriptomics and stable isotope probing; ultimately, to generate a whole-mat model of energy and element cycling in the mat within its ecological context, especially with respect to carbon balance.

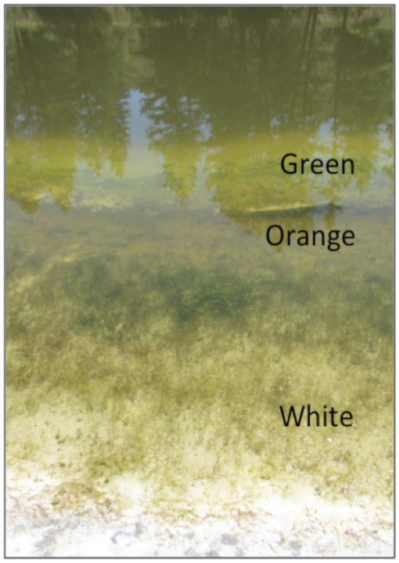
Hot Lake is a meromictic, hypersaline lake that occupies a small glacial endorrheic basin in north-central Washington and contains a benthic phototrophic microbial mat. Being highly organized, metabolically interactive, self-sustaining communities, phototrophic microbial mats are natural models for the study of energy and element cycling between phototrophs and heterotrophs and the effect these interactions have upon biogeochemical processes, especially with respect to carbon cycling. As there are no water inputs into Hot Lake besides meteoric water falling within the small basin it occupies, primary production within the lake is thought to be the major source of fixed carbon; this simplifies analysis of the carbon and energy budget of the phototrophic mat and the mat's influence on its broader geochemical context. Carbonate minerals precipitate within the mat beneath the cyanobacterial stratum by an as yet-unknown mechanism. As the metabolic activity of both sulfate reducers and oxygenic phototrophs have been independently implicated as the key driver of biogenic carbonate precipitation within microbial mats in disparate environments, study of the interactions between Hot Lake mat phototrophs and heterotrophs will shed light on the biological mechanisms of carbonate deposition common to epsomitic environments. Furthermore, Hot Lake is a dynamic system; besides variability in temperature and photon flux throughout the annum, Hot Lake also experiences a seasonal cycle in salinity, varying between observed extremes in epilimnion water of ~200 mM $MgSO_4$ after snow melt to greater than 2 M at the close of the dry season. This cycle permits analysis of the stability in composition and function of the same community under diverse chemiosmotic conditions and elucidation

* Presenting author

of the mechanisms that impart robustness. As changes in climate are likely to alter both the temperature and salinity of surface waters, the study of Hot Lake may also improve our prediction of the effects of climate change on microbial community function and biogeochemical cycling, as well as offer insight and bioprospecting for high-salt (especially epsomitic) industrial and bioenergy applications.

The first aim of this project is to understand the dynamics of community structure and function in response to the annual cycle of the major physicochemical parameters of Hot Lake, especially salinity, temperature, and photon flux. Initial analysis of community diversity by 16S rRNA pyrosequencing of mat obtained during near the seasonal maximum in photon flux revealed taxonomic units consistent with oxygenic and anoxygenic photosynthesis, sulfur oxidation and reduction, and halotolerance. Nearest-neighbor phylogeny inferred two main filamentous cyanobacteria, one of order *Oscillatoriales* and one of order *Nostocales* present within the mat. Halophilic purple sulfur bacteria of genus *Halochromatium* and *Thiohalocapsa* as well as purple nonsulfur bacteria such as *Roseobacter* and *Rhodovulum* compose the anoxygenic phototrophic guild. 16S sequences consistent with the presence of bacteria similar to the nitrate-reducing, obligately chemolithotrophic sulfur oxidizer *Thialkalivibrio nitratireducens* as well as the halophilic deltaproteobacterium *Desulfosalina* suggests active sulfur cycling within the mat. Samples were collected from three points along the depth/salinity gradient of the lake; while sequences from the major members of the expected major functional guilds were observed in all three samples, elevations in *Desulfosalina*, *Spirochaeta*, and *Bacteroidetes* in the region of the halocline suggest an increased role for anaerobic degradation of mat biomass in the transition from the oxic epilimnion to the anoxic hypolimnion. Seventeen unique strains have been isolated from the mat to date, including a previously uncultivated *Rhodobacteriaceae*, *Marinobacterium*, and *Salinibacterium* species. Both of the primary oxygenic phototrophs have been grown in unicyanobacterial consortia with mat heterotrophs with bicarbonate as the sole carbon source. These cultures are currently being analyzed by HR-NMR and LC-MS to determine the nature of the organic compounds that are being exchanged between the cyanobacterial phototrophs and associated heterotrophs. Furthermore, under a JGI Community Sequencing Program project, metagenomes of these co-cultures and whole mat will be generated and single cells from interesting taxa derived from whole mat will be isolated and sequenced. Paired samples will be utilized for metatranscriptomic analysis of *in situ* mat collected around the diel cycle. The resultant sequences, in concert with empirical data from naturally-derived simplified communities, will be used for predictive modeling of community membership and function, with an emphasis upon exchanged metabolites, in response to changes in physicochemical parameters.





The second major goal of this project is to generate a whole-mat model of energy and element cycling in the mat within its ecological context, especially with respect to carbon balance. Initial efforts towards this goal are focused on broad chemical classes; including dissolved organic carbon, sediment carbon, mat carbon, and surrounding landscape carbon. We are inventorying these pools to quantify where carbon is stored in the system and employing NMR and mass spectroscopy to identify key chemical species within each of these larger inventory pools. Stable isotope analysis is being used to link different pools and help identify carbon sources metabolized within the mat. Such surveys of total carbon and inventories of specific carbon compounds allow detection of connections between mat metabolic activity and the surrounding geochemical landscape. We are exploring metabolites produced in representative cultures and enrichments from Hot Lake to identify important potential targets for use in future SIP experimentation. This groundwork will provide a foundation for *in situ* stable isotope probing experiments to occur in the coming field season. A solid understanding of the carbon landscape within and surrounding the mats coupled with the SIP experiments should illuminate potential key carbon currencies in the mat and provide clues to carbon exchange between species within the mat community.

We are also working to understand how the system geochemistry influences the mat community. We are using a simple geochemical model to predict supersaturated species in the lake over very large seasonal fluctuations in water temperatures, salinity, and redox conditions. Initial measurements of microgradients within *ex vivo* mats have been measured with dissolved oxygen, pH and redox microelectrodes over a variety of light regimes, which is being coupled to the predictive model to refine our understanding of what geochemically-important processes are occurring within microenvironments inside the mat.

The combined genomic, enrichment incubation, and geochemical data collected to date will be compiled to guide sample collection during the summer of 2012, culminating in generation of a comprehensive dataset permitting multifaceted analysis of the activity of the mat and interactions of its members around the diel cycle near the seasonal photic maximum.

This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area. Electron microscopy and NMR measurements were performed in the William R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by OBER and located at PNNL.

169

Microbial Interactions in Hot Spring Phototrophic Mats and Isolates

David M. Ward[1]* (umbdw@montana.edu), Christian Klatt,[1] Donald A. Bryant,[1,2] Zhenfeng Liu,[2] and Sergey Stolyar[3]

[1]Montana State University, Bozeman; [2]The Pennsylvania State University, University Park; and [3]Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: Microbial mats in alkaline siliceous hot springs of Yellowstone National Park, which are constructed by a variety of oxygenic and anoxygenic phototrophic microorganisms, are model systems from which to learn principles that underlie a predictive understanding of microbial community ecology. Our goal is to use multiple "omics" and cultivation approaches to understand microbial interactions that govern the capture and recycling of energy and resources in this system.

Analyses of assembled metagenomic scaffolds have identified eight major functional guilds inhabiting the photic portion of the community. These include the oxygenic phototrophic *Synechococcus* spp., five anoxygenic phototrophic guilds (*Roseiflexus* spp., *Chloroflexus* spp. and Anaerolineae-like Chloroflexi; "*Candidatus* Chloracidobacterium thermophilum" and "*Ca.* Thermochlorobacter aerophilum") and two aerobic heterotrophic populations. Detailed annotations of genes in these assemblies have permitted reconstruction of the metabolic potential within each guild and this has guided us in developing a model of how the members of different guilds might interact metabolically. For instance, consistent with previous results, *Synechococcus* spp. possess genes that should enable them to produce glycolate through photorespiration and to produce glycogen and break it down via fermentation. *Roseiflexus* spp., "*Ca.* C. thermophilum", " *Ca.* T. aerophilum" and members of heterotrophic guilds possess genes that should enable them to metabolize glycolate and/or acetate. Surprisingly, "*Ca.* C. thermophilum" and "*Ca.* T. aerophilum" are both missing the same genes involved in the synthesis of branched amino acids (though not the transaminase genes). This observation suggests that they may obtain relevant biosynthetic precursors from another mat organism(s). One hypothesis is that polyhydroxyalkanoic acids (PHAs) of *Roseiflexus/Chloroflexus* spp. which are produced from related compounds can be a source of branched amino acid precursors for above mentioned organisms. Additionally, *Roseiflexus* spp. appear to be deficient in the synthesis of riboflavin, suggesting their dependence on populations that can produce it (e.g., *Synechococcus* spp., "*Ca.* T. aerophilum"). Finally, both *Synechococcus* spp. and *Roseiflexus* spp. may have potential to fix N_2, possibly suggesting that other community members may be dependent on them for a supply of fixed nitrogen.

These metagenomic assemblies have also been useful as references for interpreting metatranscriptomics and metaproteomics databases. While transcription patterns in general confirmed previous observations, they also revealed new ways to think about the metabolisms of mat community members. For instance, it was formerly thought that *Roseiflexus* spp. were primarily photoheterotrophic during daytime, aside from a brief period of light, but anoxic conditions in early morning, when they conduct carbon fixation via the 3-hydroxypropionate pathway. It was thus surprising to find that genes in this pathway were expressed during the day. Detailed examination of expression patterns for genes involved in central carbon metabolism and storage polymer synthesis and degradation now lead us to hypothesize that *Roseiflexus* spp. conducts a photomixotrophic metabolism throughout the day. We hypothesize that to do so they produce PHAs (and possibly wax esters) at night, which they use during the day to maintain an adequate supply of organic intermediates and reductant. If true, this could have important implications on the timing of transfer of intermediates derived from PHAs to other members of community.

These hypotheses are being tested in several ways. First, to demonstrate that *Roseiflexus/Chloroflexus* spp. incorporate CO_2 in the low-light, anoxic morning period and also in high-light oxic period, we investigated $^{13}CO_2$ incorporation into proteins of these organisms *in situ.* In a similar manner, ^{13}C-acetate has been used to assess the degree to which *Roseiflexus* spp., "*Ca*. C. thermophilum," and "*Ca*. T. aerophilum" may compete for this intermediate. Second, metabolomics analyses are being used to evaluate the hypothesis that PHAs (and possibly wax esters) undergo diel rhythms of synthesis and degradation in *Roseiflexus/Chloroflexus* spp. Third, we are developing experimental co-cultures consisting of cyanobacteria and Chloroflexi isolated from hot spring cyanobacterial mats, namely, *Thermosynechococcus sp. NAK55* (whose genome we obtained), *Roseiflexus castenholzii* and *Chloroflexus aurantiacus* J-10-fl. We found that *Thermosynechococcus* excreted a variety of metabolites such as formic, lactic, succinic, oxalic and isocaproic acids, dihydroxyacetone, sarcosine and glyceraldehyde, which have not yet been considered in our model of metabolic interactions inferred from metagenomic data. *Choloroflexus* J-10-fl was tested for its ability to utilize these metabolites and some of them were found to support growth. To investigate these potential interactions further, we are studying growth of the cyanobacterium in co-cultures with Chloroflexi isolates. Metabolites present in the growth medium and gene expression in the members of the co-cultures are being determined under light and dark conditions in two experimental settings: i) slowly mixed planktonic culture and ii) mat forming, benthic cultures growing on the surface of submerged agar. These investigations are enhancing our understanding of metabolic interactions among members of major guilds in natural hot spring cyanobacterial mats and providing new experimental approaches to evaluate the interactions.

This research is part of the PNNL Foundational Scientific Focus Area: Biological Systems Interactions.

170

Predicting Autotroph—Heterotroph Metabolic Interactions from Genome Sequence

Margrethe (Gretta) Serres[1]* (mserres@mbl.edu), Elizabeth Landorf,[2] Frank Collart,[2] Sebastian Jaramillo Riveri,[3] and **Margie Romine**[3]

[1]Marine Biological Laboratory, Woods Hole, Mass.; [2]Argonne National Laboratory, Lemont, Ill.; and [3]Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: The PNNL Foundational Scientific Focus Area, Biological Systems Interactions, investigates microbial interactions to understand how microbial communities work collectively to carry out complex biogeochemical processes. We seek to understand how the exchange of metabolites between organisms contributes to the stability of microbial communities. A longer-term goal is to develop predictive capabilities for the response of microbial communities to environmental change. A model system involving *Synechococcus* sp. PCC 7002 and *Shewanella* sp. W3-18-1 was selected to develop and test predictions of molecular exchange processes occurring during co-culture growth. To examine the molecular basis for this phototroph:heterotroph interaction we are applying comparative sequence analysis and experimental approaches to define and describe features in the genomes associated with their interactions and thereby identify functional traits important to microbial community structure. These investigations support lab-based investigations examining growth, physiology, metabolite exchange, and gene expression patterns in axenic and co-cultures.

The two model organisms chosen for our study, *Synechococcus* sp. PCC 7002 and *Shewanella* sp. W3-18-1**,** have been co-cultured without the supplement of nutrients supporting that their metabolisms can be coupled. *Synechococcus* sp. PCC 7002 is an oxygenic photoautotroph also capable of fermentative metabolism that produces different metabolic end products under variable environmental (i.e. light) conditions. *Shewanella* sp. W3-18-1 is a facultative anaerobe capable of utilizing a variety of carbon and energy sources.

Automated genome annotations of *Synechococcus* sp. PCC 7002 and *Shewanella* sp. W3-18-1 were curated to improve the specificity of functional assignments and to identify key knowledge gaps (e.g. missing steps in metabolic pathways, missing transport reactions) that require further investigation via computational and wet-lab methods. Gene models for both organisms have been adjusted using proteome data to validate protein starts and sub-cellular locations. Function predictions have included the use of experimental data from well-studied model genomes, such as *Synechocystis* sp. PCC 6803 (*Synechococcus*) and *S. oneidensis* MR-1 (*Shewanella*) as well as orthologs from phylogenetically related organisms. Pathway genome databases have been constructed for the *Synechococcus* and *Shewanella* pair. Regulatory interactions inferred in RegPrecise are included in the curation.

Comparative analysis of the *Synechococcus* sp. PCC 7002 and *Shewanella* sp. W3-18-1 genomes enables identification of the metabolic pathways that are unique to each organism. These pathways and their corresponding metabolites are a source of potentially exchanged nutrients that can contribute to the interspecies metabolic interactions. We also include pathways associated with compounds known to be present in autotrophic:heterotrophic consortia (i.e. glycolate, propionate, lactate, acetate). Based on our pathway predictions we are cataloging potential interaction points for the *Synechococcus* sp. PCC 7002 and *Shewanella* sp. W3-18-1 pair. To establish the molecular connection between these cellular metabolic pathways and exchange nutrients, we generated a profile for the transport capabilities of these organisms. Inferred functional attributes were identified by sequence comparisons using Blast and reconciled with predictions from the Transporter Classification Database. To validate inferred transport functions, we examine the ligand binding preference for the ABC transporters. This family of transporters is ubiquitous in bacterial systems and provides import and export capability for a wide spectrum of ligands. A typical ABC importer generally consists of a periplasmic solute-binding protein, two integral membrane subunits, and two cytoplasmic ATPases. It is the solute binding proteins (SBPs) of the ABC transport complex that recognize and bind specific substrates in the cell wall and transfer them to the membrane subunits, and are therefore responsible for the uptake of ligands from the environment. We screened and characterized a set of transporter proteins from *Synechococcus* sp. PCC 7002 and *Shewanella* sp. W3-18-1 using a fluorescence thermal shift assay (FTS) and isothermal calorimetry. These functional screens, that included metals, small ions, mono- and oligosaccharides, peptides, amino acids, polyamines and vitamins, generated specific binding ligand assignments for approximately 60% of the purified and screened proteins. The experimental studies provide new functional information for these transport complexes and also validate many of the assignments derived from comparative analysis. These transport capabilities support the identification and characterization of metabolic and regulatory pathways for these organisms and provide a basis for experimental validation of the potential exchange nutrients that contribute to the interspecies metabolic interactions.

This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

171

Chemical Profiling of Group B Vitamin Transport and Protein Interactions in Microbes and Microbial Communities

Aaron T. Wright[1]* (Aaron.Wright@pnnl.gov), Phillip K. Koech,[1] Andrei L. Osterman,[2] Dmitry A. Rodionov,[2] and **Margie Romine**[1]

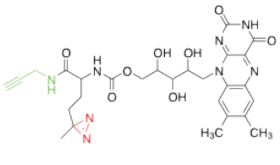
[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash. and [2]Sanford-Burnham Medical Research Institute, La Jolla, Calif.

Project Goals: The objective of the Pacific Northwest National Laboratory Foundational Scientific Focus Area is to develop a predictive, genome-enabled understanding of how microbial interactions impart stability, robustness, and functional efficiency to microbial communities. To achieve this goal, research is being conducted on both natural mat communities and defined co-cultures designed to test hypotheses regarding processes naturally occurring in these communities. As part of this effort, we have initiated a new line of investigation focused on identifying the transporters and sensors that mediate the exchange of materials between autotrophs and heterotrophs present in these communities. To complement ongoing genetic and *in situ* approaches aimed at testing substrate specificity, we propose to utilize chemical protein profiling to test specific substrates that are hypothesized to be exchanged in these communities, starting with B-type vitamins.

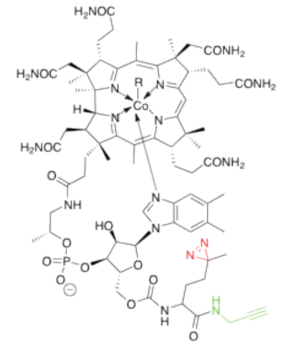
Transporters and sensor proteins are the cellular interface with the environment and the conduit for exchange of molecules and transfer of information in the ecosystem. However, their specificity cannot be readily accurately predicted from genome sequence information alone, making it difficult to predict the identity and nature of interacting microbial partners without further experimental investigation.

Chemical protein profiling employs synthesized probes that consist of three chemical elements: a binding group that biases the probe toward a protein family and/or is a substrate-mimic; a diazirine moiety for irreversibly labeling proteins; and a reporter tag for detection and isolation of probe-labeled proteins from biological systems. Probe synthesis will exploit the multimodal bio-compatible click chemistry (CC) reaction to create "tag-free" probes for profiling proteins in living systems. Probe-labeled proteins will be visualized by addition of a complementary azide-modified reporter tag using the CC cycloaddition reaction. Common reporter groups include fluorescent tags such as azido-tetramethylrhodamine for gel (SDS-PAGE)-based analysis, or an azido-biotin tag for enrichment and liquid chromatography-mass spectrometric analysis (LC-MS). The method permits a variety of characterization techniques including, but not limited to: live-cell fluorescent imaging, fluorescent SDS-PAGE, fluorescence polarimetry, FACS sorting, and LC-MS analysis. Anticipated applications include evaluation of molecular function, cellular distribution or uptake, characterization of probe-labeled proteins by MS-based proteomic analysis, and real-time monitoring of probe uptake and protein binding. We will integrate our experimentally identified metabolite-protein interactions with computational efforts to characterize novel sequence variants of transporter and sensor proteins.

As one of our first objectives we intend to characterize transporters and sensors that modulate group B vitamin exchange in microbial co-cultures and in natural communities. These vitamins are required, in relatively limited amounts, as intermediates/precursors in the biogenesis of key cofactors in central metabolism. Subsystems analysis revealed that the presence of *de novo* and/or salvage pathways is highly variable among representatives of microbial taxa in the natural communities under investigation and thus vitamin exchange is a likely driver of opportunistic interactions. Our initial analyses will involve the use of vitamin B_2 and vitamin B_{12} chemical probes to define group B vitamin transport/sensor interactions in a *Synechococcus* sp. PCC 7002 and *Shewanella putrefaciens W3-18-1* co-culture. Comparative analysis of subsystems and regulons suggest that *S. putrefaciens W3-18-1* is able to synthesize B_2 and B_{12} while *Synechococcus* sp. PCC 7002 can only synthesize B_2. The proteins required for uptake of vitamins B_2 and B_{12} have been predicted in each organism, however those involved in its export (if any exist) by *Shewanella* are unknown. Chemical probe profiling will be carried out *in vitro* and *in situ* to test functional predictions and potentially identify new proteins involved in exchange of this commodity. We believe this new chemical approach will reveal important vitamin-protein interactions in these systems, and importantly, be a technology universally applicable to more complex biological systems.



Vitamin B$_2$ Probe



Vitamin B$_{12}$ Probe

This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

172

Proteomic Characterization of Microbes and Microbial Communities

Mary S. Lipton[1]* (mary.lipton@pnnl.gov), Laurey Steinke,[2] Roslyn Brown,[1] Ronald J. Moore,[1] Therese R.W. Clauss,[1] Gordon Slysz,[1] Gordon A. Anderson,[1] David M. Ward,[3]* Christian Klatt,[3] Donald A. Bryant,[4] William P. Inskeep,[3] Mark Kozubal,[3] Zack Jay,[3] Jacob Beam,[5] and Ryan Jennings[3]

[1]Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Wash.; [2]Department of Biochemistry and Molecular Biology, University of Nebraska Medical Center, Omaha; [3]Montana State University, Bozeman; and [4]The Pennsylvania State University, University Park

Project Goals: The overarching goal of the PNNL Foundational Scientific Focus Area (FSFA) is to investigate fundamental scientific issues in microbial interactions—those that occur between different microbes as well as between microbes and their extracellular environment—to understand how microbial communities work collectively to carry out complex biogeochemical processes. This research component will develop and apply proteomic approaches to define these interactions, to track the flow of carbon from the autotrophic population to the heterotrophic population and to elucidate the metabolic control of energy flow and nutrient cycling in thermophilic microbial mats.

The challenges associated with the proteomic characterization of natural microbial mat communities growing in extreme environments are as varied as the types of environments. These challenges range from extremely small sample size and the presence of contaminating inorganic and organic materials in the mats to the difficulties in delineating which proteins and which organisms are responsible for the function that is present. To address these challenges, we are developing and applying methods for the extraction and analysis of proteins from the mats that include in situ stable isotope probing, advanced protein extraction methods, and subcellular fractionation methods in mixed populations.

Microbial autotroph- heterotroph associations are metabolically interactive, self-sustaining biological systems that are widespread in terrestrial and aquatic environments. Of specific interest to the FSFA program are phototrophic microbial mats found in alkaline siliceous hot springs in Yellowstone National Park (YNP), and chemoautotrophic iron and sulfur-oxidizing mats found in acidophilic, hyperthermophilic springs in YNP.

While the characterization of the microbial mats at Mushroom springs at YNP is limited by small samples sizes, the well annotated metagenome and relative ease of sample processing allow advanced analysis using *in situ* Protein Stable Isotope Probing (Pro-SIP) labeling strategies. Specifically, Pro-SIP has been applied to samples taken directly from the field with the immediate short-term exposure to ^{13}C-CO$_2$. Advanced analytics and informatics developed at PNNL have allowed for identification of taxa and associated proteins that are expressed and are actively taking up CO$_2$ during the incubation at the tested time point. A preliminary analysis of the first 10% of the data yielded 24 proteins validated by two or more unique labeled peptides. Since 14 were from *Roseiflexus RS-1*, 9 were from *Synechococcus* and only 1 was from *Candidatus Thermochlorobacter* aerophilum, it was determined that *Roseiflexus RS-1* is active in reducing CO$_2$ in the morning period of sunrise through the beginning of high light, and *Synechococcus* is active to a lesser extent in this time frame. During this time frame, these organisms are mainly synthesizing chaperones and transcription associated proteins.

In contrast to the phototrophic mats, the chemoautotrophic mats from YNP present different challenges for proteomic characterization. The relative high level of archaeal components and the high concentration of iron and sulfur in the mats confound cell lysis and protein extraction. To overcome these challenges, we have developed novel protein extraction techniques and have successfully detected a suite of iron, sulfur and arsenic oxidation proteins from the organisms present in the these communities. The meta-genome analysis suggests that the organism *Thermoproteales* (strain WP30) is present and further that this heterotrophic population respires on elemental sulfur and/or arsenate during growth on complex carbon sources. This genomic hypothesis is supported by the identification of the Mo-pterin proteins responsible for reduction of elemental sulfur and arsenate in these communities.

We will present the new and developing technologies for protein extraction, protein labeling and subcellular fractionation for organisms inhabiting natural microbial mats in extreme environments along with information on the interactions among these organisms and their relationship to that environment.

This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

173

Integrating Geochemical, Microscopic, and –Omics Analyses to Understand Microbial Interactions in High-Temperature Chemotrophic Communities

William P. Inskeep[1]* (binskeep@montana.edu), Mark Kozubal,[1] Zack Jay,[1] Jacob Beam,[1] Ryan Jennings,[1] Hans Bernstein,[1] Laura Whitmore,[1] Ross Carlson,[1] Roslyn Brown,[2] Mary Lipton,[2] Helen Kreuzer,[2] James Moran,[2] Alice Dohnalkova,[2] Margie Romine,[2] James Fredrickson,[2] Susannah Tringe,[3] and Tanja Woyke[3]

[1]Montana State University, Bozeman; [2]Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Wash.; and [3]DOE Joint Genome Institute, Walnut Creek, Calif.

Project Goals: To understand metabolic control of energy flow and nutrient cycling, and interactions among microbial community members in thermophilic chemotrophic microbial mats.

The advent of molecular tools and -omics technologies has provided opportunities for assessing the predominant and relevant indigenous organisms present in natural microbial communities, as well as their function within a connected network of different populations. High-temperature microbial communities are often considerably less diverse than mesophilic environments and are constrained by geochemical attributes such as pH, dissolved oxygen, Fe, sulfide, and or trace elements including arsenic and mercury. Consequently, one of the project goals is to understand microbial interactions among chemoautotrophic and heterotrophic members of high-temperature acidic Fe-oxidizing communities and sub-oxic elemental sulfur systems in Yellowstone National Park (YNP). Specific objectives of this work are to identify the predominant transcripts, proteins, metabolites and isotopic signatures associated with high-temperature microbial communities and to establish metabolic network models for different habitat types using consensus sequence assemblies of major phylotypes present in Fe(III)-oxide mats and sulfur sediments. Microbial community metabolic models derived from metagenome assemblies of individual populations are being used to test hypotheses regarding the role of specific phylotypes within each habitat type as well as microbial interactions occurring *in situ*.

Prior metagenome sequencing of high-temperature Fe-oxide and elemental sulfur systems from Yellowstone National Park (YNP) reveal communities dominated by thermophilic archaea and/or members of the deeply-rooted bacterial order Aquificales. Phylogenetic and functional analysis of metagenome sequence has provided 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, the predominant microbial populations present in acidic Fe-oxide microbial mats of Norris Geyser Basin (YNP) include chemoautotrophs such as *Metallosphaera yellowstonensis*, a member of the crenarchaeal order Sulfolobales. Pure-culture laboratory experiments have recently confirmed that this chemoautotroph can fix inorganic carbon (i.e. CO_2) using the 3-hydroxypropionate/4-hydroxybutyrate pathway while obtaining energy from the oxidation of ferrous Fe. Gene expression studies under different treatment conditions showed that *M. yellowstonensis* utilizes a novel terminal oxidase complex to oxidize Fe(II) (*fox* gene complex) while fixing inorganic carbon. Consequently, *M. yellowstonensis* is an important primary producer in high-temperature acidic Fe-oxide mats, and may provide a source of organic carbon for other heterotrophs present in the community.

Metagenome sequence assemblies show that the Fe-oxide mats also contain potential heterotrophs including Desulfurococcales and Thermoproteales-like populations. Functional analyses of genes belonging to these phylotypes suggest that these organisms degrade complex carbon sources, and that specific proteins may serve as a primary carbon and energy source. Recent isolation and characterization of a representative Thermoproteales organism (strain WP30) from YNP shows that this heterotrophic population respires on elemental sulfur and/or arsenate during growth on complex carbon sources. The Mo-pterin proteins responsible for reduction of elemental sulfur and arsenate have been identified and gene expression studies are underway to confirm the role of these novel proteins in community function. A deeply-rooted archaeal population has also been identified as a major community member in high-temperature Fe mats (referred to here as 'Novel Archaea Group 1'), and *de novo* sequence assemblies suggest that this organism is heterotrophic, potentially utilizing complex carbon sources produced by *M. yellowstonensis*. Iron depositional studies have been conducted to correlate Fe-oxidation rates with O_2 flux rates measured at the Fe-mat interface (using O_2-microelectrodes) as well as 16S rRNA gene copy-numbers of *M. yellowstonensis* (measured using quantitative PCR). Results show an excellent correlation between *M. yellowstonensis* copy-number and Fe-oxide deposition rate, especially during early stages of Fe-oxide mat development.

Early results from proteomic analyses of Fe-oxide mat samples confirm the importance of *M. yellowstonensis* and Novel Archaeal Group 1 populations in high-temperature Fe-oxidizing communities. Proteins identified in Fe-mat communities are being used to refine and improve metabolic models constructed using genome sequence. Pathway specific processes are also being elucidated using isotope measurements focused on ^{13}C and ^{34}S of different chemical fractions. Additional proteomic and future transcriptomic results will be used to assess the importance of specific pathways, and will provide detailed information necessary to test hypotheses regarding the major microbial interactions that define community structure and function. Integration of genomic, proteomic, and metabolic information to understand autotroph-heterotroph interactions is tractable within high-temperature geothermal systems, in part due to the relative simplicity of the communities and the stability of key geochemical variables including pH, Fe, O_2 and dissolved sulfide/elemental sulfur.

This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

174

In Situ Correlated Molecular Imaging of Chemically Communicating Microbial Communities

R. Masyuko,[1] C.M. Driscoll,[1] **P.W. Bohn**[1]* (pbohn@nd.edu), J. Shrout,[1] and J.V. Sweedler[2]

[1]University of Notre Dame, Notre Dame, Ind. and [2]University of Illinois, Urbana-Champaign

Project Goals: (see below)

This project is exploring the potential of heterocorrelated mass spectrometric (MS) and confocal Raman microscopy (CRM) chemical imaging, as targeted to the problem of microbial/environmental processes. Specifically we are: (1) enhancing the functionality and performance of secondary ion mass spectrometry (SIMS) and laser desorption ionization (LDI) via tandem MS and improving the spatial resolution and analyte range by introducing a high-flux C_{60} source along with improved stage control and automated collection routines; (2) developing nanoparticle-enhanced correlated imaging with heightened spatial/temporal resolution and increased sensitivity; (3) developing a laboratory testbed for the *in situ* (natural) system to enable the simultaneous investigation of all components (bacteria-root-fungus) of a three-component rhizosphere model. Initially, we are examining the bacterium *Pseudomonas aeruginosa*, starting with the "relatively" simple case of *P. aeruginosa* group motility on idealized surfaces, imaging of homoserine lactones and surface remodeling with rhamnolipid and then transitioning to a more complex three-component system, composed of a bacterium (*P. aeruginosa*), a fungus (*Laccaria bicolor*) and model root derived from *Populus deltoides*. This poster will report on initial results based on chemometric approaches to Raman spectral devolution in complex matrices and enhancements to sensitivity in both LDI-MS and SIMS imaging.

175

From Genomes to Metabolomes: Studying Mechanisms of Interspecies Interaction Using the Archaeal System *Ignicoccus-Nanoarchaeum*

Richard Giannone,[1] Brian Bothner,[2] Valérie Copié,[2] Robert Hettich,[1] and **Mircea Podar**[1,3]* (mpodar@utk.edu)

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

Few, if any, microbes live in functional or spatial isolation. The nature of the various types of inter-species interactions can be complex, ranging from competition to syntrophy and mutualism. Such relationships can impact keystone species and play a major role in energy and element cycles at scales that extend past ecosystem boundaries. There is limited understanding of fundamental mechanisms of interspecies recognition and communication, how they impact genome evolution, what genetic regulatory mechanisms control metabolic/energetic coupling between species in response to environmental factors. To address such questions we are using the archaeal system *Ignicoccus hospitalis-Nanoarchaeum equitans*. With a combined genomic complement of less than 2000 genes and an obligate chemolithoautotrophic metabolism, this system represents one of the simplest specific microbial association and allows fundamental system level investigations and modeling of symbiosis. The integrated approach and the results of this research will be applied to investigations of more complex environmental systems.

A laboratory cultivation system for *Ignicoccus-Nanoarchaeum* has been established. The genomes of the two organisms are available and in addition we have sequenced two additional *Ignicoccus* species that do not interact with *N. equitans*. We have also performed an in depth proteomic analysis of *Ignicoccus-Nanoarchaeum* (1). Differences in the relative abundance of >75% of predicted protein-coding genes from both Archaea were measured to identify the specific response of *I. hospitalis* to the presence of *N. equitans* on its surface. A purified *N. equitans* sample was also analyzed for evidence of interspecies protein transfer. The depth of cellular proteome coverage achieved is amongst the highest reported for any organism. Based on changes in the proteome, *I. hospitalis* reacts to *N. equitans* by curtailing genetic information processing (replication, transcription) in lieu of intensifying its energetic, protein processing and cellular membrane functions. Using the information from initial studies we are now combining cultivation of these archaea under various settings with parallel transcriptomic, proteomic and metabolomic analyses. We are testing the specific hypothesis that the physical interaction between *Ignicoccus hospitalis* and *Nanoarchaeum* in laboratory cultures is induced and controlled by specific temporal gene expression, metabolic events and surface protein-protein interactions. We will identify candidate genes, proteins, and small molecules regulating the metabolic/energetic coupling network shared by the two organisms, which will allow us to establish a model of symbiosis at the genomic and metabolic level. Using comparative and functional genomics in multi-species cultures with archaea that do not serve as hosts for *Nanoarchaeum*, we will independently test the mechanism of interaction between *I. hospitalis* and *Nanoarchaeum* and will derive an evolutionary genomics model for the development of a microbial symbiotic relationship.

Reference

1. Giannone RJ, Huber H, Karpinets T, Heimerl T, Küper U, Rachel R, Keller M, Hettich RL, Podar M. (2011). Proteomic characterization of cellular and molecular processes that enable the *Nanoarchaeum equitans-Ignicoccus hospitalis* relationship. PLoS One. 2011;6(8):e22942.

This research is funded by the U.S. Department of Energy, Office of Biological and Environmental Research under the Genomic Science Program (DOE-DE-SC0006654).

176

System-Scale Modeling of Mycorrhizal Symbiosis

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

[1]Biosciences Division, Argonne National Laboratory, Lemont, Ill. and [2]Department of Biological Sciences, University of Alabama, Huntsville

http://www.bio.anl.gov/molecular_and_systems_biology/proteins.html

Project Goals: The Argonne "Environment Sensing and Response" Scientific Focus Area (ESR-SFA) program addresses the hypothesis that cellular behavior can be modeled through an understanding of the biological interface with the environment and the cellular responses that originate from the cell/environment interaction. The overall objective of the ESR-SFA program is to identify the molecular basis of cellular transport and sensory pathways that mediate the response to environmental nutrients. This project furthers the objectives of this ESR-SFA by developing approaches that evaluate plant/fungal symbiosis by inferring molecular mechanisms from transcriptomic analysis of a laboratory *Laccaria bicolor*:aspen mycorrhizal system. Our approach generates testable biological hypotheses for the specific molecular mechanisms that drive nutrient exchange in ectomycorrhizal symbiosis. Increased understanding of how partners in plant/fungal symbiosis sense and respond to environmental parameters will identify the specific molecular metabolic and environmental response pathways for eukaryotic organisms in soil ecosystems.

The roots of many plant species have the ability to form symbiotic relationships with the community of soil fungi. In these relationships, mycorrhizal fungi provide mobilized nutrients from the soil, and plants provide photosynthetically-derived sugars. The plant:fungal symbiosis can provide growth and survival benefits, but the many details of the molecular exchanges are not completely understood. Deep RNA sequence analysis has revolutionized our access to the molecular activity of these communities, providing a system-scale perspective of symbiotic interactions at unprecedented resolution. However, no single method of transcriptomic analysis alone is sufficient to enable the inference of metabolite-space from transcriptomic-space. Identification of significantly expressed genes and mapping those genes onto known metabolic pathways generates a scaffold on which analysis of differential expression and predicted metabolic turnover can be accomplished. The results of this analysis are predictions of specific metabolic compounds and expressed protein activities that can be used to generate hypothesis-driven molecular biological experiments.

We generated a model of mycorrhizal metabolome that confirms prior biological knowledge and provides insight into the nutrient exchange process associated with mycorrhizal interaction. When *L. bicolor*:aspen mycorrhizae are cultured on media with inorganic nitrogen sources, *L. bicolor* is predicted to take up inorganic nitrogen from the media and synthesize complex nitrogenous compounds for its symbiotic plant partner. These predictions indicate *L. bicolor* is an active metabolic partner in mycorrhizal interactions in addition to passively extending the absorptive surface of aspen roots. Proposed transcriptomic experiments, examining mycorrhizal systems in the context of different nutrient environments, will uncover additional mechanisms of mycorrhizal metabolic interactions. Application of further transcriptomic analysis methods such as identifying gene splice variants expressed during different stages of mycorrhizal interaction, have the ability to provide even greater richness to the transcriptomic-generated models of the mycorrhizal metabalome. The approach applied to this ectomycorrhizal system can be generalized to other interacting systems in the rhizosphere.

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).

177

Plant-Microbe Interfaces: Mining Genomic Signatures of Species-Specific Microbe Interactions in *Populus* Using High Density SNP Arrays and Whole Genome Resequencing

Wellington Muchero* (mucherow@ornl.gov), Jessy L. Labbé, Ranjan Priya, Lee Gunter, **Gerald A. Tuskan**, 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 experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we 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.

Highly specialized plant-microbe interactions such as mycorrhizal symbiosis exhibit remarkable species specificity in economically important plants such as *Populus*. As such, molecular genomic signatures of these specialized interactions are expected to closely mirror the molecular genomic signatures of species differentiation, offering a manageable entry point in efforts to identify and characterize genetic elements essential for establishment of such interactions. In this study, we characterized genome divergence on Chromosome 1 of the *Populus* genome known to harbor major genetic determinants of *Populus*-*Laccaria* symbiosis as a test case for identifying candidate genes based on species-level genome divergence. We used a high-density single nucleotide polymorphism (SNP) array with a genome-wide coverage based on 34,130 probes to genotype three *Populus* species. Out of 3,716 SNPs on chromosome 1, we identified 229 SNPs that were not transferable between *P. trichocarpa* and *P. deltoides* or *P. fremontii*. These non-transferrable SNPs were over-represented in three regions on chromosome 1 representing putative zones of genome divergence during *Populus* speciation. In these regions, 25 genes with putative functions in plant-microbe interaction were identified. These included a phosphate transporter-related gene previously implicated in symbiotic phosphate transportation in arbuscular mycorrhizae, as well as a Vapyrin gene that was reported to be an essential factor in intracellular progression of arbuscular mycorrhizal symbiosis. Resequencing data for these genes will be used to characterize differences in gene structure between *P. trichocarpa* and *P. deltoides* that could potentially explain species-related *Populus*-*Laccaria* interactions.

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

178

Plant-Microbe Interfaces: Emerging Research Directions

Jay Chen* (chenj@ornl.gov), Jessy Labbé, Jennifer Morrell-Falvey, Sarah Melton, Jim Guo, Udaya C. Kalluri, Dave Weston, Dale Pelletier, **Gerald A. Tuskan**, 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 experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we 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 molecular events that lead to recognition and colonization of a host plant by beneficial bacteria or fungi are poorly understood. Ongoing research in this project is aimed at identifying molecular, genetic and cellular events in *Populus* involved in recognition and establishment of beneficial microbial interactions, and at identifying and isolating microbes associated with natural *Populus* ecosystems. Results from these ongoing studies have led to initiation of new research directions to enhance our understanding of plant-microbe interactions. New efforts are focused on elucidating the molecular mechanisms of microbial biofilm formation on plant roots, the role of helper bacteria in promoting beneficial plant-fungal interactions, and on development of a *Populus* protoplast transient expression system for the molecular and biochemical characterization of regulatory proteins.

The mutualistic association between plant roots and microbes form an environment that is ideal for biofilm formation, including sufficient moisture and nutrients, which are supplied by the plant host. This project is focused on identifying the specific conditions in the plant rhizosphere that trigger biofilm formation in selected microbes. To this end, we have tested the ability of selected microbes to form biofilms under various growth conditions using a well-described biofilm formation assay. Data from multiple *Pseudomonas* strains indicate differential responses to additions in the growth media. For example, sucrose, a common constituent of plant root exudates, appears to induce biofilm formation in GM49. Likewise, the presence of high levels of phosphate induces biofilm formation in GM60 and GM67. These and other phenotypes will form the basis for selecting microbes on which to focus in-depth studies to elucidate molecular pathways involved in root colonization.

The soil is probably one of the most complex ecosystems in which plant–fungal–bacterial interactions operate. Mycorrhizal fungi are surrounded by complex microbial communities, which modulate the mycorrhizal symbiosis that impacts biomass production, defense against pathogens, and tree nutrition. Among this diverse microbial community, the so-called mycorrhiza helper bacteria (MHB) are thought to assist mycorrhiza formation and symbiosis. Since very little is known about the role of MHB in *Populus*-fungi interactions, this project is aimed at dissecting the signaling mechanisms underlying *Populus*–fungal–bacterial interactions. To this end, we have analyzed the morphological changes induced in *Laccaria bicolor* by several bacterial strains isolated from the *Populus* rhizosphere, and performed *Populus*–*L. bicolor*–bacteria co-cultures under controlled greenhouse conditions. We demonstrate that some bacterial strains influence *Populus*–*L. bicolor* colonization and have a clear beneficial effect. This study provides new insights into the mechanism of interaction between *Populus* and its complex microbial communities.

Transcriptomcis, proteomics and metabolomics have uncovered a suite of genes involved in *Populus*-microbe interactions. However, it remains unclear how these components interact with each other and are coordinated in the same signaling cascades to regulate some specific processes of plant-microbe interactions. Molecular and biochemical characterization of these genes is essential for elucidating the molecular mechanism underlying plant-microbe interactions. We have established a protocol for efficient isolation of protoplasts from *Populus* leaf mesophyll cells. Subsequently, we have established a protoplast transient expression system using *Populus* protoplasts. We demonstrated that *Populus* protoplasts respond to biotic and abiotic stimuli in a similar manner as that in intact plants. Furthermore, by using a series of fluorescent fusion proteins, we have established subcellular localization of various organelle markers. The *Populus* protoplast transient expression system will be used to study protein-protein interactions, protein-DNA interactions, post-translational modifications, protein degradation, and artificial microRNA-mediated gene silencing. Furthermore, together with gene reporters, this system will be used to examine the specific role of candidate genes in regulating plant-microbe interactions. The *Populus* protoplast transient expression system represents a much-needed system for research in the post-omics era.

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

179

Plant-Microbe Interfaces: Revealing Regularities in Large Datasets with Biological Annotations Using Networks and Rules of their Association

Tatiana Karpinets[1]* (karpinetstv@ornl.gov), Byung Park,[2] Guruprasad Kora,[2] Mustafa Syed,[1] Christopher Schadt,[1] Gerald Tuskan,[1] Dale Pelletier,[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: 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 experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing 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 expanse of experimental data produced by the Plant-Microbe Interfaces (PMI) project requires the development and application of novel computational tools for their processing. The data, collected from thousands of samples from diverse geographical locations in multiple PMI experimental campaigns, include a complex set of biological annotations, like phenotypic and genotypic characteristics of *Populus*, structure of fungal and bacterial communities in the tree rhizosphere and endosphere, and inherently heterogeneous soil properties. Combined in one table, this mash up of qualitative and quantitative data is challenging to analyze and interpret. To address this challenge we developed a computational framework that expands current approaches to viewing, visualizing, searching, and analyzing information in large databases with biological annotations. The approach is based on two novel concepts, the type-value format and the association network (Anet) that supplement the idea of association rules (Arules) produced by 'Apriori'. Type-value format simplifies computational processing, filtering and grouping of biological annotations by preserving their two-level structural organization in the transaction records and further in Arules and Anets. For annotation types presented by quantitative data, such as genome size or GC content, quantities are replaced with their quality levels based on distributions of the quantities in the dataset records. Association network provides a way to link a diverse set of annotations directly, by the number of transactions where each pair of annotations co-occurred, and indirectly, by considering similarity between profiles of their co-occurrences with other annotations. Monte Carlo simulation is used to assess the significance of similarity by p-value. The resulting Anet is further analyzed and visualized at different levels of resolution, or p-value thresholds, using clustering and visualization techniques. In combination, Anets and Arules provide researchers a powerful tool to create a bird's-eye view of the collected information, to extract hidden biological regularities and to generate hypotheses for further experimental validation.

To test this framework we applied it to the analysis of metadata of sequenced prokaryotic genomes from the GenomeOnLine Database (GOLD). The overlapping structure of the data provides a good case study for the proposed framework. The generated Anet revealed a hidden structure in metadata of the prokaryotic organisms with three major clusters representing metadata of pathogens, environmental isolates and plant symbionts. The annotations clustered in each group represented a distinct signature profile of metadata for each group and showed a strong link between phenotypic, genomic and environmental features of the organisms.

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

180

Plant-Microbe Interfaces: Dynamic Data Analytics Through Integrated Knowledgebase and LIMS

Guruprasad Kora[2]* (koragh@ornl.gov), Susan Holladay,[1] Tatiana Karpinets,[1] Mustafa Syed,[1] Christopher Schadt,[1] **Edward Uberbacher**,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn. and [2]Computer Science and Mathematics, 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 experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we 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.

Data management, efficient tracking and effective visualization are essential aspects of any field-study related biological process. Each of these aspects pose numerous challenges as size and complexity of the data and field-study campaigns grow. Establishing a disciplined data management infrastructure at all levels of field-study campaigns enables effective life-cycle (data collection to analysis) management of any biological processes.

The Plant-Microbe Interfaces (PMI) project has developed an easy to use, web-based data management and tracking portal as part of their Knowledgebase for PMI field-studies that enables users to track and manage the entire life-cycle of the collected data. PMI LIMS and Knowledgebases are coupled together to form an integrated data analytics platform to perform statistical as well as data mining operations dynamically. The Knowledgebase augments data analytics infrastructure with scientific visualization techniques (visual data analysis). It enables users to use graphical representation of data as a means of gaining understanding and insight into the data. It helps researchers to comprehend spatial and temporal relationships between collected data. It provides efficient interactive techniques for researchers to focuses on exploratory, comparative analytics and visualization.

The PMI Knowledgebase has an established data acquisition workflow that enabled field-studies carried out in 2009-2011 to be effectively managed. Currently, the system supports various types of data from raw soil readings to tree specification to post-processed data like 454 sequences, isolates etc. The system supports various datatypes from image files, ascii-based text files, binary blobs etc. The workflow 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 Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

* Presenting author

181

Plant-Microbe Interfaces: Genotype-by-Environment Interactions Drive Root-Associated Microbiome Composition in Natural Populations of *Populus deltoides*

Christopher W. Schadt[1,2]* (schadtcw@ornl.gov), Migun Shakya,[1,2] Neil Gottel,[1] Hector Castro,[1] Zamin Yang,[1] Marilyn Kerley,[1] Gregory Bonito,[3] Dale Pelletier,[1,2] Tatiana Karpinets,[1] Guruprasad Kora,[1] Jessy Labbé,[1] Edward Uberbacher,[1,2] Wellington Muchero,[1] Gerald Tuskan,[1] Mircea Podar,[1,2] Rytas Vilgalys,[3] and **Mitchel J. Doktycz**[1,2]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Genome Science and Technology Program, University of Tennessee, Knoxville; and [3]Department of Biology, Duke University, Durham, N.C.

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 experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we 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.

Members of the genus *Populus* represent genetically diverse, ecologically widespread riparian species that are potential cellulosic feedstocks for biofuels, and the first woody plant species to have a genome sequence. The trees are also 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 few published efforts to comprehensively describe microbial communities of mature natural populations of *Populus*. We have compared root endophyte and rhizosphere samples collected from two dozen sites within watersheds/populations of *Populus deltoides* in Tennessee and North Carolina over multiple seasons. 454 pyrosequencing has been applied to survey and quantify the microbial community associated with *P. deltoides*, using primers targeting the V4 and V7-8 regions of the bacterial 16S rRNA gene and the D1 region of the fungal 28S rRNA gene. 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. Soil physical, chemical and nutrient status, as well as tree growth and age characteristics were also evaluated. Root endosphere and rhizosphere communities have been found to be composed of distinct assemblages of bacteria and fungi with largely non-overlapping OTU distributions. Within these distinct endophyte and rhizosphere habitats, community structure is also influenced by soil characteristics, watershed origin and/or plant genotype, while observed seasonal influences have been minimal. We have also isolated over a thousand bacteria and fungi from these environments representing dominant community members *in situ.* Many of these isolates show distinct growth phenotypes with *Populus.* These findings indicate that the characteristics of the *Populus* root/soil environment may represent a relatively strong selective force in shaping endophyte and rhizosphere microbial communities and their functions.

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

182

Plant-Microbe Interfaces: Characterization of IAA7.1 in *Populus* and its Role in Plant-Microbe Interactions

Poornima Sukumar[1]* (sukumarp@ornl.gov), Tatiana Karpinets,[1] Karuna Chourey,[2] Gregory Hurst,[2] Jay Chen,[1] **Gerald A. Tuskan**,[1] **Udaya C. Kalluri**,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division and [2]Chemical Sciences 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 experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we 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.

Auxins are a class of plant hormones regulating many physiological and developmental processes. Auxin signaling involve a group of transcription repressor proteins known as AUX/IAA (Auxin/IAA induced), and another family of transcription factors that induce auxin induced gene expression called ARFs (Auxin Response Factor). Both AUX/IAAs and ARFs have conserved protein family domains. Many studies in *Arabidopsis* have explored the individual roles of AUX/IAAs and ARFs. Our comparative bioinformatics analysis of the gene families in *Populus* and *Arabidopsis* shows that *Populus* typically has multiple orthologs (co-orthologs) for a given *Arabidopsis* gene. It is intriguing to understand whether the co-orthologs differ functionally in a more complex plant system such as *Populus*, with respect to the auxin signaling pathway and/or in establishing beneficial plant-microbe relationships. The present study examines the effect of knocking down two *Populus* co-orthologs (PtrIAA7.1 and PtrIAA7.2) of the *Arabidopsis* gene, IAA7- PtrIAA7.1 has a unique protein family domain that has thus far not been reported in any other member of the gene family, in any species. The domain is a tandem duplication of domain II, which is known to cause protein instability in other plant species. Physiological comparisons show that RNAi lines of PtrIAA7.1 have more severe phenotypes than those of PtrIAA7.2. PtrIAA7.1 mutants display reduced plant height, increased lateral branching, and reduced of apical dominance. Additionally, co-culture with *Piriformospora indica*, *Laccaria bicolor* strain S238N, and *Pseudomonas spp* strain GM30, resulted in decreased responses in the PtrIAA7.1 mutant compared to control plants, including reduced sensitivity to microbe induced changes in shoot and root growth and fresh weights. Protein profiling of shoot and root samples of PtrIAA7.1 revealed that shoots of PtrIAA7.1 lack a predicted importin protein belonging to a class of proteins involved in import of substrates into the nucleus. Additionally, PtrIAA7.1 mutant has lower abundance of ADP Ribosylation Factor proteins. Further experiments are underway to investigate the potential role of these differential expressed proteins in the auxin signaling pathway and their regulation during plant-microbe interactions.

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

183

Plant-Microbe Interfaces: Sequencing of Twenty-One *Pseudomonas* Genomes and Twenty-Three Genomes from Diverse Bacteria Isolated from *Populus deltoides* Rhizospheres and Endospheres

Sagar M. Utturkar[1]* (sutturka@utk.edu), **Steven D. Brown**,[1,2] Dawn M. Klingeman,[2] Courtney M. Johnson,[2] Stanton L. Martin,[2] Miriam L. Land,[2] Tse-Yuan S. Lu,[2] Christopher W. Schadt,[1,2] **Dale A. Pelletier**,[1,2] and **Mitchel J. Doktycz**[1,2]

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

http://pmi.ornl.gov
http://genome.ornl.gov/microbial/PMI/

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 experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we 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 interactions within the rhizosphere have been shown to have important roles in plant health and productivity (*1, 4*). As part of an ongoing effort to better understand the microbial communities associated with native *Populus deltoides* (Eastern cottonwood), we have under taken both cultivation independent and cultivation dependent assessments of microbial communities from *P. deltoides* rhizospheres and endospheres (*3*). 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.

We sampled *P. deltoides* at sites along the Caney Fork River in central Tennessee and at Yadkin River in North Carolina, USA. These sites represent ecotypes and soil conditions that are common to this region. Approximately 1,100 diverse bacterial strains were isolated from these sites over multiple seasons. The isolates comprise 7 classes and 85 genera of bacteria including, *Actinobacteria* (14%), *Bacilli* (17%), *Flavobacteria* (6%), *Sphingobacteria* (3%), and α- (22%) β- (16%) and γ- (22%) Proteobacteria. A number of our isolates were *Pseudomonas* species, which have considerable genetic and phenotypic variability and different members are pathogenic, biocontrol and plant growth promoting bacteria. Representatives were chosen for genome sequencing based on phenotypic traits and phylogeny. A total of 24 γ-Proteobacteria, 6- β-Proteobacteria, 10 α-Proteobacteria, 2-Bacteroidetes and 2-Bacilli were sequenced.

Paired-end DNA libraries with an average insert size of 500bp were created and draft genome data was generated using the Illumina (*2*) HiSeq2000 technology. CLC Genomics Workbench (version 4.7.1) and FASTQC (*6*) were applied to trim reads for quality sequence data. We analyzed the effect of quality based trimming and filtering of reads on genome assemblies and compared assembly outputs generated by the Velvet assembler (Version 1.1.04) (*7*) and the CLC Genomics Workbench software. In each case found the selection of high quality reads was a key step for successful assemblies. Quality trimming dramatically improved the assembly i.e. reduced the number of contigs while maintaining the expected genome size. Velvet assemblies were further optimized by selecting different Kmer values and employing a scaffolding algorithm within the program. In the case of *Rhizobium* sp. PDO-076 a GS FLX shotgun dataset was also available, which was used to generate a hybrid assembly by combining 454 reads and shredded Velvet scaffolds using Newbler (Version 2.6). Initial assembly validation was performed based on parameters such as number of contigs, genome size, N50, contig length etc.

Draft genome sequences were annotated at Oak Ridge National Laboratory using the Microbial Genome Annotation Pipeline automated annotation pipeline at ORNL, which is based on the Prodigal gene prediction algorithm (*5*). Final validation was performed by annotating the assembled genome with ORNL genome annotation pipeline and comparing with known conserved sequences. We assembled diverse genomes that varied in size from 4.4MB to 10.7MB and had G+C contents in the range of 33% to 69%. The 4.9 Mb of *Phyllobacterium* genome was assembled in as few as 39 contigs with largest contig size of 1.4 Mb, while maximum number of contigs generated was the 778 for the 5.4 Mb *Caulobacter* genome. The N50 statistics varies from 14,172 to 506,356 bp with average value of 154,391 bp. These new genome sequences will allow more comprehensive comparisons for bacteria involved in plant-microbe interactions.

References

1. Bais, H. P., T. L. Weir, L. G. Perry, S. Gilroy, and J. M. Vivanco. 2006. The role of root exudates in rhizosphere interations with plants and other organisms. Annu. Rev. Plant Biol. 57:233-266.
2. Bennett, S. 2004. Solexa Ltd. Pharmacogenomics 5:433-438.
3. Gottel, N. R., et al. 2011. Distinct microbial communities within the endosphere and rhizosphere of *Populus deltoides* roots across contrasting soil types. Appl. Environ. Microbiol. 77:5934-5944.
4. Hodge, A., G. Berta, C. Doussan, F. Merchan, and M. Crespi. 2009. Plant root growth, architecture and function. Plant Soil 321:153-187.
5. Hyatt, D., G. L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer, and L. J. Hauser. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.
6. Kircher, M., P. Heyn, and J. Kelso. 2011. Addressing challenges in the production and analysis of illumina sequencing data. BMC Genomics 12:382.
7. Zerbino, D. R., and E. Birney. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821-829.

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

184

Plant-Microbe Interfaces: Whole Genome Sequence Analysis of Bacterial Strains Isolated From the *Populus* Microbiome

Dale A. Pelletier[1,2]* (pelletierda@ornl.gov), Tse-Yuan Lu,[1] Christopher W. Schadt,[1,2] Tatiana Karpinets,[1] Sagar Utturkar,[1,2] Dawn Klingeman,[1] Steven D. Brown,[1,2] Patricia Lankford,[1] Marilyn Kerley,[1] Amy Schaefer,[3] Caroline S. Harwood,[3] and **Mitchel J. Doktycz**[1,2]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Genome Science and Technology Program, University of Tennessee, Knoxville; and [3]Department of Microbiology, University of Washington, Seattle

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 experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we 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 microorganisms within their endosphere and rhizosphere that can have positive effects on the host. Our goal is to understand the phylogenetic and functional diversity within the *Populus* microbiome and to elucidate the metabolic and molecular mechanisms responsible for shaping the *Populus*-microbial interface. To begin to untangle this complex ecosystem, we isolated bacterial strains native *Populus deltoides* roots collected in central Tennessee and North Carolina. A diverse array of bacterial strains (>1000) comprising some 7 class and 89 genera of bacteria were isolated from the rhizosphere (529) and root endosphere (558). The isolates comprise Actinobacteria (15%), Bacilli (17%), Flavobacteria (6%), Sphingobacteria (3%), and a- (22%) b- (15%) and g- (22%) proteobacteria. In order to explore potential metabolic and physiological diversity present within *Populus* microbiome isolates, we performed whole genome sequencing on a subset of 43 bacterial isolates. Isolates where chosen for sequencing based on a variety of factors including abundance in native ecosystems, ability to colonize plants, both microbial and plant phenotyping and physiological properties. Sequence data was generated by Illumina HiSeq2000 paired-end sequencing of 500 bp insert libraries. Sequence reads were assembled using Velvet and annotated by the ORNL automated genome annotation pipeline. The initial metabolic analyses of the sequenced strains targeted a comparison of KEGG orthologous groups. Metabolic activities of the organisms were predicted using the same KEGG annotation pipeline and then analyzed using network tools to identify clusters of KEGG orthologous groups that are specific for each isolate or group of isolates. The analysis revealed metabolic diversity of the isolates. Metabolic reconstructions generated for isolates provided further insight into the nature of metabolic interactions of the isolates with the plant host.

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

185

Plant-Microbe Interfaces: Dynamics of Bacterial Microbiome of *Populus deltoides*

Migun Shakya,[1,2]* (shakyam@ornl.gov), Neil R. Gottel,[3] Hector F. Castro,[1] Zamin Yang,[1] Marilyn Kerley,[1] **Mircea Podar**,[1,2] Mitchel J. Doktycz,[1,2] and **Christopher W. Schadt**[1,2]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Genome Science and Technology Program, University of Tennessee, Knoxville; and [3]Department of Microbiology, University of Texas, Austin

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 experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we 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 root-rhizosphere interface of *Populus* is an ideal model to study the interaction between plants and microorganisms. In our recent paper (Gottel et al. 2011), we characterized microbial communities from root endophytic and rhizospheric habitats of *P. deltoides* in mature, natural trees. However the relative effects of drivers of community composition such as soil and environmental properties, seasonal variation, and host genotype that shape these communities remains unclear. To address these issues, we carried out 454 pyrosequencing using primers targeting V7-V9 region of 16S bacterial SSU rRNA from samples collected from the rhizosphere and endosphere of two-dozen sites distributed across watersheds in North Carolina and Tennessee over two seasons. From a total of 686,384 high-quality, denoised, non-chimeric sequences, we identified key bacterial taxa associated with the endosphere and rhizosphere. Proteobacteria consistently dominated both environments, while Actinobacteria were found across all samples and in some endosphere samples they replaced Proteobacteria as the dominant taxa. More than 24,000 unique bacterial OTUs were detected however 21,487 OTUs were exclusive to rhizosphere, while only 2,598 OTUs were found in endophyte samples. We did not observe a clear distinction in the community composition of samples collected over multiple seasons, but variations associated with location, soil properties, and plant genotypes were observed. Our data indicate that the endophyte associated bacterial community has low diversity, but is highly variable from tree to tree. While rhizosphere associated bacterial communities are highly diverse community makeup is very similar from tree to tree. This study is one of the more comprehensive analyses of bacterial communities in the endosphere and rhizosphere of *P. deltoides* to date and highlights the complexity of the bacterial diversity associated with mature trees in natural systems.

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

186

Plant-Microbe Interfaces: Quorum Sensing Systems are Prevalent in the *Populus* Microbiome

Amy L. Schaefer*[1] (amyschae@uw.edu), **Dale A. Pelletier**,[2,3] Sagar Utturkar,[2,3] Steven Brown,[2,3] Colin R. Lappala,[1] Ryan Morlen,[1] Daniel B. Mueller,[1] 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]Genome Science and Technology Program, University of Tennessee, Knoxville

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 experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we 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

* Presenting author

these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

As part of the 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 rRNA sequences indicates the *Populus* bacterial communities are dominated by Acidobacteria, Alphaproteobacteria, and Gammaproteobacteria and the Proteobacteria are the predominant group isolated from *Populus* endophyte and rhizosphere samples (Gottel et al, Appl Environ. 77:5934). 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 novel AHL-type signals (*p*-coumaroyl-HSL, cinnamoyl-HSL, and isovaleryl-HSL) made by several soil- and plant-associated Alphaproteobacteria. This suggests that there may be additional novel HSL-type signals made by bacteria.

We initially screened ~120 Proteobacteria isolated from *P. deltoides* for AHL production and found >80% Alphaproteobacteria and >20% of the Gammaproteobacteria isolates to be positive. This suggested AHL signaling is prevalent in *Populus* microbial communities. When we examined the entire genome sequence of 40 of the Proteobacterial isolates, AHL signaling genes were even more prevalent than we predicted from AHL screening results as at least one *luxI*- or *luxR*-type gene was present in 10/10, 3/6, and 24/24 of the sequenced Alphaproteobacteria, Betaproteobacteria, and Gammaproteobacteria genomes, respectively. Interestingly, sequence analysis of some of the *luxI*-type genes suggests the potential for novel AHL-type signals. Sequence analysis of the *luxR*-type genes suggests some receptors proteins may be responsive to an unknown plant-derived compound, as has reently been reporter for bacteria associated with certain crop plants (Subramoni et al, Appl Environ. 77:4579).

AHL quorum sensing often controls the production of "public goods" such as antimicrobials and exoezymes, 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. Using this enzyme we are defining the AHL-regulons of *Populus*-associated bacteria, by using RNAseq (Hirakawa et al, J Bacteriol. 193:2598).

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

187

Plant-Microbe Interfaces: *Populus deltoides* Supports Distinct Fungal and Bacterial Root Associates From Other Ectomycorrhizal Tree Hosts

Gregory M. Bonito[1]* (gmb2@duke.edu), **Christopher W. Schadt**,[2] Gerald A. Tuskan,[2] Mitchel J. Doktycz,[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: 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 experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we 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 and is largely dependent on flooding for seedling recruitment. This species is unique from co-occuring tree species in that it forms root associations with both arbuscular and ectomycorrhizal fungal species. *Populus deltoides* also harbors both bacterial and fungal endophyte communities within its roots. To address the influence of edaphic or genotypic factors on the structuring of rhizospheric assemblages, we carried out a series of trap-plant experiments in growth chamber environment using rooted cuttings of *P. deltoides* and two other tree species (*Quercus alba* and *Pinus taeda*) grown in soils from our field sites. We used 454 multiplex amplicon pyrosequencing to characterize rhizospheric fungal and bacterial diversity for each plant in the experimental treatments. Specific objectives of this research were to:

1. Determine the effect of host species (*P. deltoides*, *Q. alba*, *P. taeda*) on the structuring of rhizosphere bacterial and fungal communities

2. Determine the influence of *P. deltoides* genotype on the structuring of rhizosphere bacterial and fungal communities
3. Determine the effect of different field soil inoculum on bacterial and fungal rhizosphere communities of individual host species and *P. deltoides* genotypes

Field soils from our ORNL *P. deltoides* research sites in NC and TN were used as the source of microbial inoculum in these trap-plant studies. Cuttings from different *P. deltoides* genotypes were planted into a 50% mixture of sterile sand and field soil. For the host treatment we grew *P. deltoides*, a *P. deltoides x P. trichocarpa* hybrid, oak (*Quercus alba*) and pine (*Pinus taeda*) in the same soil. For the soil treatment a single *P. deltoides* genotype was grown in multiple soil types. Plants were harvested after five-months of growth. Soils were washed off the root systems, roots visually assessed for ectomycorrhizas and fungal infection, and samples of bulk roots used for DNA extraction. The fungal community from each plant was sequenced at the ITS and LSU rDNA regions using the fungal specific primers ITS1f/ITS4 and LROR/LR3. Arbuscular mycorrhizae were preferentially amplified selectively with the primer set AML1 and a modified AML2 primer. Bacterial 16S rDNA primers that amplify across the V4 region and discriminate against plastid DNA were used to compare rhizosphere bacterial communities in selected samples. Pine and oak seedlings showed high ectomycorrhizal colonization (>80%), while most of the *P. deltoides* genotypes had low ectomycorrhizal colonization (<30%). This observation was verified through sequence-based assessments, with more than 75% of the sequences from *Populus* belonging to endophytic fungi and over half of those from oak and pine belonging to ectomycorrhizal fungi. The most abundant ectomycorrhizal fungi on *P. deltoides* belonged to the genera *Peziza, Inocybe* and *Hebeloma*. The *Peziza* and *Inocybe* taxa also co-occurred on oaks and pines. Ectomycorrhizal species of *Tuber* and *Laccaria* were recovered from both *Populus* and oak roots. Although *Populus* hosted fewer ectomycorrhizal taxa than either oak or pine, a more species-rich assemblage of endophytic fungi was detected in *Populus* roots. The total richness of root associated fungi (238) and bacteria (283) taxa was significantly greater for *Populus* than oak (175-f;184-b) or pine (157-f;185-b). For the bacteria, *Populus* was characterized by a higher relative abundance of *Actinobacteriales* and *Sphingobacteriales* than were oak or pine, and a lower abundance of *Rhizobiales* and *Berkholderiales.* Nonetheless, a core set of fungal (43) and bacteria (103) taxa were shared between the three hosts. There were some minor responses of bacterial and fungal communities to *Populus* genotype. In particular, the *P. deltoides x P. trichocarpa* hybrid had a wider spread across species space (based on PCA ordination) and was represented by a greater frequency and relative abundance of ectomycorrhizal taxa (e.g. *Inocybe, Tomentella, Hebeloma*). Arbuscular mycorrhizal fungi belonging to the Glomerales (*Glomus* sp.) and Paraglomales (*Paraglomus* sp.) were present in all genotypes, but sequences belonging to the Diversisporales were only detected in the hybrid genotype. Soils differed significantly in their effect on microbial communities. For example, a novel species of *Atractiellales* was the most abundant species in some assayed soils, but was absent in other soils. Similar patterns were observed for other microbial taxa. While we did detect fungi and bacteria in our negative controls, these clustered tightly in ordination space and showed little overlap with taxa from the experimental treatments. In conclusion, the microbiota associated with roots of *P. deltoides* is diverse and unique from oak and *pinus*, and appears to be structured both by the microbial inoculum available in soils and (to a lesser extent) plant genotype.

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

188

Plant-Microbe Interfaces: Genome-Wide Identification of *Populus* Small Proteins Responsive to Mycorrhizal Symbiosis

Ting Li,[1] Jonathan Plett,[2] Chu-Yu Ye,[1] Gregory B. Hurst,[3] Jessy L. Labbé,[1] Wellington Muchero,[1] Priya Ranjan,[1] Hengfu Yin,[1] Gerald A. Tuskan,[1] Timothy J. Tschaplinski,[1] **Francis Martin**,[2] and **Xiaohan Yang**[1]* (yangx@ornl.gov)

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Ecogenomics of Interactions Laboratory, UMR ‘Interactions Arbres/Micro-Organismes’, INRA-Nancy, France; and [3]Chemical Sciences 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 experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we 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 mycorrhizal symbiosis, representing the most widespread plant-microbe association, offers various benefits including 1) enhancing carbon sequestration in terrestrial ecosystems, 2) increasing nutrient availability, 3) remediating degraded soils and 4) improving water use efficiency. All of these beneficial aspects make plant-mycorrhizal association an excellent strategy for improving the sustainability of bioenergy feedstock production. However, our knowledge about the molecular mechanism underlying mycorrhizal symbiosis is still very limited. To address this limitation, we performed a genome-wide analysis of *Populus* genes in response to *Laccaria* inoculation (Fig. 1). We identified 1,282 transcripts differentially expressed during mycorrhizal development, among which 764 novel transcripts have not been documented in the current version of *Populus* genome annotation. About 11% of the 1,282 differentially-expressed transcripts encode small proteins of <100 amino acids in length. Gene ontology analysis revealed that mycorrhizal symbiosis between *Populus* and *Laccaria* involves different sets of genes over the time course of symbiosis development (Table 1). A large number of protein sequences encoded by these transcripts were predicted to be located in the nuclei, suggesting that they may play roles in gene expression regulation. Interestingly, many protein sequences were predicted to be located in both nuclei and extracellular space, implying that they are putative signaling proteins responsible for communication between *Populus* and *Laccaria.* Also, we found that 11 transcripts were up-regulated through the time course of mycorrhizal development, suggesting that they are important to the maintenance of plant-fungus interaction. Since *P. trichocarpa* better associates with *Laccaria* when compared to *P. deltoides*, we examined the differences in the DNA-seq coverage of differentially expressed genes (based on genome resequencing data) between the two *Populus* species. We identified 10 genes that are present in *P. trichocarpa* but absent in *P. deltoides*. Lastly, we performed homology search in 24 other sequenced plant genomes ranging from algae to moss to angiosperm and found that only about 40% of the *Laccaria*-responsive genes in *Populus* have homologs in other plant species, suggesting that lineage-specific molecular mechanism may play an important role in regulating plant response to mycorrhizal fungi.

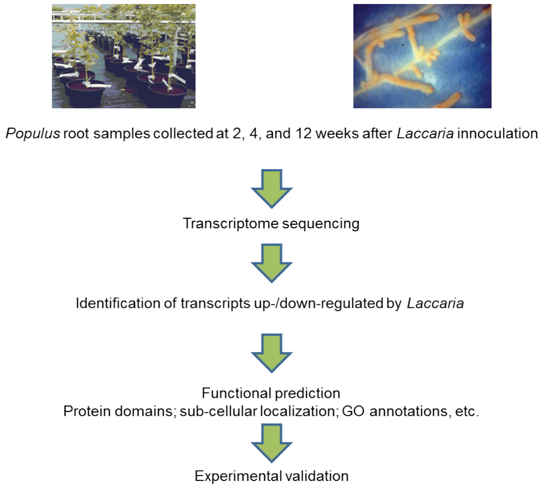

Populus root samples collected at 2, 4, and 12 weeks after *Laccaria* inoculation
Transcriptome sequencing
Identification of transcripts up-/down-regulated by *Laccaria*
Functional prediction
Protein domains; sub-cellular localization; GO annotations, etc.
Experimental validation

Figure 1. A pipeline for discovery of *Populus* small proteins mediating plant-fungus interaction.

Table 1: Biological processes enriched in up-/down-regulated *Populus* genes at 2, 4, and 12 weeks after *Laccaria* inoculation.

	2 weeks	**4 weeks**	**12 weeks**
Up regulation	Response to stimulus	Macromolecule metabolic process	Nitrogen compound metabolic process
Down regulation	Nitrogen compound metabolic process	Nitrogen compound metabolic process	Biosynthetic process

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

189

Plant-Microbe Interfaces: Extending Single Plant-Microbe Co-Expression and Metabolic Networks to Community Scales

David J. Weston (westondj@ornl.gov), Dale A. Pelletier, Abhijit A. Karve,* Jennifer L. Morrell-Falvey, Timothy J. Tschaplinski, Sara S. Jawdy, Lee E. Gunter, Nancy L. Engle, Jessy L. Labbé, Tse-Yuan S. Lu, Chris W. Schadt, **Gerald A. Tuskan**, 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 experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we 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.

Colonization of plants by nonpathogenic *Pseudomonas fluorescens* strains can confer enhanced defense capacity against a broad spectrum of pathogens. Few studies, however, have linked defense pathway regulation to primary metabolism and physiology. In this study, physiological data, metabolites and transcript profiles are integrated to elucidate how molecular networks initiated at the root – microbe interface influence shoot metabolism and whole-plant performance. Experiments with *Arabidopsis thaliana* were performed using the newly identified *Pseudomonas fluorescens* GM30 or *P. fluorescens* Pf-5 strains. Co-expression networks indicated that Pf-5 and GM30 induced a subnetwork specific to roots enriched for genes participating in RNA regulation, protein degradation and hormonal metabolism. In contrast, only GM30 induced a subnetwork enriched for calcium signaling, sugar and nutrient signaling and auxin metabolism, suggesting strain-dependence in network architecture. In addition, one subnetwork present in shoots was enriched for genes in secondary metabolism, photosynthetic light reactions and hormone metabolism. Metabolite analysis indicated that this network initiated changes in carbohydrate and amino acid metabolism. Consistent with this, we observed strain-specific responses in tryptophan and phenylalanine abundance. Both strains reduced host plant carbon gain as estimated by net photosynthesis, yet provided a clear fitness benefit when plants were challenged with the pathogen *Pseudomonas syringae* DC3000.

The trade-off between host carbon cost under optimal conditions and fitness benefit under pathogen attack brings to question how plant-microbe interactions are perceived and initiated within natural systems harboring complex microbial communities. Does the host, for example, favor carbon partitioning to microbial associates that in turn confer a fitness advantage? Alternatively, carbon acquisition from the host may simply be driven by the ability for the microbe to evade host plant defense. To begin to address these questions, we have developed a microcosm system providing axenic conditions in a realistic soil environment that is conducive to whole plant phenotype imaging and multi-omic sample collections. We are exploiting the use of whole-genome sequences from 43 microbial isolates collected from the *Populus* microbiome (see Pelletier poster) to create synthetic communities. Within this system, we are currently testing whether host plant genotype and nutritional state feedback to influence microbial community structure. In addition, we are testing whether the selected microbial community confers a beneficial host phenotype through community decomposition studies.

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

190

Plant-Microbe Interfaces: Proteomics at the Plant-Microbe Interface

Gregory B. Hurst[1]* (hurstgb@ornl.gov), Karuna Chourey,[1] Xiaohan Yang,[2] Ting Li,[2] Chongle Pan,[1] Zhou Li,[4] Rachel M. Adams,[4] Robert L. Hettich,[1] Patricia K. Lankford,[2] Keiji G. Asano,[1] Tatiana V. Karpinets,[2] Udaya C. Kalluri,[2] Poornima Sukumar,[2] David J. Weston,[2] Sara M. Allen,[2] Dale A. Pelletier,[2,4] Jonathan Plett,[3] Francis Martin,[3] Timothy J. Tschaplinski,[2] **Gerald A. Tuskan**,[2] and **Mitchel J. Doktycz**[2,4]

[1]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]Ecogenomics of Interactions Laboratory, INRA-Nancy, France; and [4]Genome Science and Technology Program, Knoxville, 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 experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we 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 proteomes and metaproteomes of plants, bacteria, and fungi offer insights into the mechanisms that these organisms employ at plant-microbe interfaces. We have initiated proteomics studies of several organisms involved in these interfaces, with an emphasis on those involving *Populus* trees, to identify processes in which proteins are important for establishment and maintenance of the interface.

-Proteomics of *Laccaria bicolor* mycelia We have performed "shotgun" proteomics of mycelia from the ectomycorrhizal fungus *Laccaria bicolor* grown under a variety of conditions, including different fungal strains, media compositions, cold and heat stresses, and proximity to roots of *Populus* plants. The average number of *L. bicolor* proteins identified per sample was ~1200. Approximately 440 proteins were detected in at least one replicate of each condition, establishing a "core" proteome for *L. bicolor*.

-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 developing proteomics methods to target small proteins from plant tissues. After evaluating several methods for enriching small proteins in an *E. coli* model system, we applied the most promising technique (in-gel digestion of low molecular weight regions from 1D SDS-PAGE separations) to studying small proteins in *Arabidopsis* roots and shoots. Of a total of >5000 proteins identified from unfractionated extracts and small protein fractions, a few tens of proteins were identified exclusively in gel fractions corresponding to molecular masses below 20 kilodaltons. 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 an important first step towards studying plant-microbe interactions in the rhizosphere. Extraction of proteins from plants is technically challenging, and further complications arise when the samples are obtained from mature tree roots. We have identified an approach that reliably yields proteins from roots of naturally occurring *Populus* trees, sampled during PMI collection trips to the Yadkin River in North Carolina. We typically identify >1000 proteins from each root sample, despite heterogeneity in morphology, differences in location, soil type, etc.; from 4 individual trees, we have evidence for ~2600 proteins to date (with some redundancies due to gene duplications, etc.). A "core" proteome that is common to all these root samples contains several hundred proteins. Analyses of functional categories and subcellular locations are ongoing.

-Comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics Using a model soil bacterial species, *Pseudomonas putida* F1, we have compared figures of merit such as depth of proteome coverage, quantification accuracy, precision, and reproducibility for several quantitative proteomics methods using a high-performance hybrid mass spectrometer, the LTQ Orbitrap Velos. Each approach has particular merits, and the final choice of approach depends on the requirements of the experiment at hand. Our results indicate that isobaric 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.

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

191

Nutrient Cycling for Biomass: Interactive Proteomic/Transcriptomic Networks for Global Carbon Management Processes within Poplar-Mycorrhizal Interactions

Leland J. Cseke[1]* (csekel@uah.edu), Frank M. Collart,[2] Peter E. Larsen,[2] Stephen Barnes,[3] and Helen Kim[3]

[1]University of Alabama, Huntsville; [2]Argonne National Laboratory, Lemont, Ill.; and [3]University of Alabama, Birmingham

Project Goals: This project will facilitate the development of system-scale models of the symbiotic interaction between ectomycorrhizal fungi (such as *Laccaria bicolor*) and tree species (such as poplar) in response to environmental nutrient availability/biochemistry.

The experimental plan will integrate multiple "omics" approaches to model ectomycorrhizal regulatory networks and metabolic pathways that are predictive of atmospheric carbon sequestration in the form of plant and/or subsurface fungal biomass. The project will test the hypothesis that essential regulatory and metabolic mechanisms can be inferred from transcriptomic and proteomic changes that occur at the mycorrhizal interface in response to environmental nutrient availability. Guided by abundant genome sequence and ongoing transcriptomic input, this hypothesis will be addressed using modern protein analytic approaches to fill the gap in our understanding of how mycorrhizal metabolic and regulatory processes at the transcript-level translate to nutrient uptake, carbon management and ultimate net primary productivity of plants in the environment. Specifically, we make use of targeted as well as discovery-based proteomics, biochemical assays and ChIP-seq characterization of carbon, nitrogen and phosphorous regulators to identify symbiosis-specific molecular mechanisms that control plant carbon management and allocation. These molecular mechanisms regulate nutrient cycling, accumulation of plant and fungal biomass, and ultimately are important to forest management and atmospheric carbon sequestration.

192

Global Gene Expression Profiling of Switchgrass Following Inoculation with *Burkholderia phytofirmans* Strain PsJN

Alejandra Lara-Chavez[1]* (alejandra.lara@ialr.org), Seonhwa Kim,[1] Scott Lowman,[1,2] Yuhong Tang,[4] Barry Flinn,[1,2],[3] John Seiler,[3] Jerzy Nowak,[2] and **Chuansheng Mei**[1,2,3]

[1]Institute for Sustainable and Renewable Resources, Institute for Advanced Learning and Research, Danville, Va.; [2]Departments of Horticulture and [3]Forest Resources and Environmental Conservation, Virginia Polytechnic Institute and State University, Blacksburg; and [4]Genomics Core Facility, The Samuel Roberts Noble Foundation, Inc., Ardmore, Okla.

http://www.ialr.org/images/stories/research/isrr/meiprojectreport9-28-11.pdf

Project Goals: 1) Global gene expression profiling induced by switchgrass bacterization with the beneficial bacterial endophyte *Burkholderia phytofirmans* strain PsJN and 2) Identification of key genes from global gene expression profiling and a study of their functions.

Switchgrass is one of the promising bioenergy 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. The overall goal of the project is to develop a low input and sustainable switchgrass feedstock production system utilizing beneficial bacteria endophytes. Previous results on the inoculation of switchgrass lowland cultivar Alamo with *Burkholderia phytofirmans* strain PsJN indicated a significant increase of growth under *in vitro*, growth chamber and greenhouse conditions. However, no beneficial responses were recorded with the upland cultivar Cave-in-Rock. In order to explore this genotype effect further, a comparative global gene expression profiling was conducted both cultivars following PsJN inoculation using the DOE-funded switchgrass cDNA microarrays. Ten-day old seedlings were inoculated with PsJN (0.5 at OD_{600}) and tissue samples collected at 0 (prior to inoculation), 0.5, 2, 4 and 8 days after inoculation, with three biological replicates. Non-inoculated controls were treated with PBS buffer and sampled as above.

Based on the microarray data analyses, 50 candidate genes that exhibited significant differences in the expression levels between PsJN-inoculated Alamo and Cave-in-Rock were chosen for further study. These genes are being subjected to secondary verification using qPCR. So far, 20 out of the 50 have been verified. From these 20 genes, five key genes representing glutathione S-transferase, calmodulin-related calcium sensor protein, an EF-hand transcription factor, histidine-containing phosphotransfer protein and a zinc-finger protein have been chosen for further functional studies using overexpression and RNAi knockout/knockdown techniques. Overexpression constructs for glutathione S-transferase and calmodulin-related calcium sensor genes were also introduced into switchgrass embryogenic callus, and plants will be regenerated and tested for endophyte PsJN responses.

Further analysis of the microarray data focused on groups of genes that are up-regulated in Alamo and down-regulated in Cave-in-Rock at each sampling point. We identified approximately 1947, 877, 402 and 1140 genes that displayed this pattern at 0.5, 2, 4 and 8 days after inoculation, respectively. The majority of genes showing differences are annotated as "expressed proteins" and "unknown proteins". We are currently focused on transcription factor genes, such as AP2 domain, MYB family, F-box domain, and zinc finger protein. Further studies of these transcription factor gene functions are listed in Table 1.

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

Table1. Expression level changes of transcription factor genes of interest in Alamo and Cave-in-Rock at 0.5, 2, 4 and 8 days following inoculation with PsJN, compared with expression level at 0 day, respectively.

ID probe	**Annotation**	**Alamo**				**Cave-in-Rock**			
		0.5	**2**	**4**	**8**	**0.5**	**2**	**4**	**8**
AP13ITG55712_at	AP2 domain	1.71	1.48	2.14	2.80	0.05	0.05	0.07	0.07
AP13ITG63524RC_s_at		2.27	1.75	2.59	2.29	0.89	0.68	0.79	1.14
AP13CTG22494_at	bZIP	1.88	3.58	3.03	1.80	1.27	1.48	1.51	0.95
AP13ITG54829_at		2.62	2.05	2.79	1.71	1.45	1.40	1.39	1.68
AP13CTG24092_at	MYB family	1.52	1.68	2.06	1.24	0.95	0.94	0.91	0.98
KanlowCTG34263_at		1.24	2.09	5.46	4.58	0.71	0.93	1.42	2.12
KanlowCTG22073_s_at		2.25	0.94	0.57	0.52	1.36	1.37	1.26	1.24
AP13ITG65291_at	F-box domain	1.53	2.03	2.26	2.88	1.15	1.06	0.79	0.98
KanlowCTG42852_s_at		1.20	1.70	2.13	2.15	0.75	0.77	0.66	0.68
AP13ITG41289_at		1.18	1.60	2.07	1.83	0.32	0.27	0.29	0.33
AP13ITG57608_s_at	RING-H2 finger	1.09	2.28	2.49	2.81	0.77	0.95	0.88	0.96
AP13ITG69131RC_at	zinc finger, C3HC4 type	1.56	1.76	2.07	2.28	0.74	0.62	0.67	0.71
AlamoCTG04292_s_at		2.26	1.22	1.16	1.11	1.65	1.95	1.89	1.90
AP13CTG19863_at	TFs having WRKY and zinc finger domain	3.13	1.94	1.70	1.69	0.12	0.11	0.17	0.17
AP13CTG44559_s_at		1.68	2.53	4.58	4.20	0.03	0.04	0.05	0.06
AP13.12336.m00003_s_at	No apical meristem	3.60	1.55	0.83	0.90	3.80	4.31	4.85	3.94
KanlowCTG46205_s_at	Transcription elongation factor	3.71	2.02	1.88	1.17	0.77	0.70	1.29	1.67
AP13CTG09371_s_at	zinc finger	2.53	1.30	0.89	0.73	1.15	1.82	1.96	1.95
AP13ITG48832_s_at	AT hook motif	2.73	1.38	0.67	0.45	1.20	1.33	1.35	1.19

193

ENIGMA Biotechnology: Systems Approaches to Studying Microbial Communities

Richard Baran,[1] Benjamin Bowen,[1] Morgan Price,[1] Christopher Petzold,[1] Simon Allen,[3] Susan J. Fisher,[3] Haichuan Liu,[3] Steve Hall,[3] Evelin D. Szakal,[3] H. Ewa Witkowska,[3] Adam A. Arkin,[1] Gareth P. Butland,[1] Swapnil Chhabra,[1] John Tainer,[1] Greg Hura,[1] Mark Biggin,[1] Kenneth Downing,[1] Robert Glaeser,[1] Manfred Auer,[1] Gary Siuzdak,[2] Adam M. Deutschbauer,[1] Paul Adams,[1] and **Trent Northen**[1]* (trnorthen@lbl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]The Scripps Center for Mass Spectrometry, La Jolla, Calif.; and [3]UCSF Sandler-Moore Mass Spectrometry Core Facility, University of California, San Francisco

http://enigma.lbl.gov/

Project Goals: Microorganism-based approaches 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, at a molecular systems level, the bacterial soil communities at DOE sites contaminated with heavy metals or radionuclides with sufficient detail to predictively model interactions within microbial and community processes that drive complex geochemistry in key environments. In doing so we expect to define biological principles governing selection of microbial community function and composition in given environments.

To ascertain the key processes contributed by an organism to a community and to characterize that process within the organism in molecular detail, it is necessary not only to identify the components of the pathways involved, but to also understand their functional interactions with other processes in the organism of interest. The Biotechnology Component of ENIGMA is developing and applying a suite of technological approaches, from genetics, protein abundance, structure, localization, and metabolism to enable systems-level insights into microbial activity. We have established a flexible experimental pipeline in metal-reducing and sulfate-reducing bacteria (SRB) for (1) high-throughput strain/construct generation, (2) evidence-based annotation of gene function using mutagenesis and extensive phenotyping, (3) evidence-based annotation of transcripts using tiling microarrays and RNAseq (4) protein and protein complex isolation (TEM and x-ray techniques), (5) mass spectrometric based proteomic analysis, (6) Protein and protein complex structural analysis, (7) mass spectrometry based metabolomics analysis (GC-TOF, LC-MS/MS, NIMS), and (8) high resolution imaging (FIB/SEM, SBF/SEM, PALM, STORM). Further technology development will enable us to apply our approaches to environmental isolates rapidly and cost effectively. Additionally, we are exploring the integration of diverse data types including metabolomics and high-throughput genetics to elucidate gene function. Long term, these diverse data will form the foundation for predictive models for a number of key microorganisms from a single environment thereby providing a rich resource for assessing ecological questions relevant to microbial community structure and function.

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.

194

ENIGMA Biotechnology: Metabolic Profiling of Bacterial Mutant Libraries to Link Metabolite Utilization to Genotype

Richard Baran* (RBaran@lbl.gov), Benjamin Bowen, Morgan Price, Adam A. Arkin, Adam Deutschbauer, and **Trent Northen**

Lawrence Berkeley National Laboratory, Berkeley, Calif.

http://enigma.lbl.gov/

Project Goals: Microorganism-based approaches 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, at a molecular systems level, the bacterial soil communities at DOE sites contaminated with heavy metals or radionuclides with sufficient detail to predictively model interactions within microbial and community processes that drive complex geochemistry in key environments. In doing so we expect to define biological principles governing selection of microbial community function and composition in given environments.

The discrepancy between the pace of sequencing and functional characterization of genomes has been recognized as one of the major challenges in microbial genomics. The ENIGMA program at LBNL is addressing this by integrating high-throughput genetics and mutant libraries with metabolic observations to establish direct functional associations between genes and metabolic processes. Mass spectrometry (MS) based metabolomics allows the profiling of metabolites in complex biological samples with high sensitivity and is well suited for interrogating the metabolic capabilities of microorganisms. We have used untargeted metabolomics to identify unexpected and novel metabolites as well as to map the uptake and release of a broad range of metabolites by different microorganisms. Large, genome-wide libraries of mutant strains, developed for multiple ENIGMA microorganisms, enable investigations into the role of specific genes in various metabolic processes using high-throughput metabolomics. However, metabolite profiling of complex samples is usually performed using time-consuming chromatography to separate the metabolites prior MS analysis. This makes it largely incompatible with screening large mutant libraries. To meet this challenge, we developed a workflow combining untargeted and high-throughput metabolomics to identify genes of *Escherichia coli* and *Shewanella oneidensis* MR-1 related to the utilization of specific metabolites. The ability to uptake specific metabolites from different complex media was analyzed using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). A subset of metabolites found to be utilized were supplemented to minimal media. A limited complexity of the media facilitated high-throughput analysis of spent media extracts using MS. A total of 8000 mutant strains were screened for defects in metabolite utilization. The presence of one of the tested metabolites in the spent media extracts of specific mutants directly links genes to a metabolic defect. Intracellular metabolites of these mutants were then profiled by LC-MS to identify potential accumulation of intermediates related to the utilization of specific metabolites. Using this approach, we identified genes of known function as well as putative transport proteins and enzymes with previously ambiguous annotations.

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.

195

ENIGMA – Biotechnology: High-Throughput Chemogenomic Fitness Profiling of ENIGMA Relevant Microbes Using a Next Generation Barcode Sequencing Assay

Adam M. Deutschbauer,[1] Jennifer He,[1] Morgan Price,[1] Jennifer V. Kuehl,[1] Kelly Wetmore,[1] Nancy L. Liu,[1] Sonia Reveco,[1] **Swapnil R. Chhabra**,[1] Grant M. Zane,[2] Tom R. Juba,[2] **Judy D. Wall**,[2] **John-Marc Chandonia**,[1] **Trent R. Northen**,[1] **Adam P. Arkin**,[1] and **Gareth P. Butland**[1]* (GPButland@lbl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif. and [2]University of Missouri, Columbia

Project Goals: ENIGMA's goal is to understand, at a molecular systems level, the bacterial soil communities at DOE sites contaminated with heavy metals or radionuclides. 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. It is therefore critical to understand how microbes respond to environmental changes which may invoke a cellular stress response and inhibit growth or alternatively provide an opportunity to thrive within a community. A comprehensive understanding of the pathways impacted by specific stresses or changes in environment, together with data from other ENIGMA groups detailing changes in community composition as a whole under similar conditions will allow us to link the viability of specific organisms of interest to their ability to contribute or perform specific functions for or within the community.

To ascertain the key processes contributed by an organism to a community and to characterize that process within the organism in molecular detail, it is necessary not only to identify the components of the pathways involved, but to also understand their functional interactions with other processes in the organism of interest. To address these questions, we have utilized recently created comprehensive genome wide collections of single mutant knockout strains and performed chemogenomic fitness profiling to assess the change in viability (growth fitness) of thousands of mutant strains under a variety of growth conditions (i.e. presence of stress agent, change in growth medium, etc.) in a high throughput format. Chemogenomic interactions are identified when a specific mutant or collection of mutant strains have a statistically significant defect or growth advantage under a specific condition when compared to other mutants, and are a powerful indicator of functionally related genes which may be present in the same or related pathways and processes.

Previously used methods to assay the fitness of mutant strains in model systems have relied on the imaging of colony growth on agar plates. This method is not an option for a potentially diverse range of microbes of interest to ENIGMA which may not form morphologically uniform colonies or may excrete chromophoric substances into the media, precluding accurate imaging. However, due to the successful development and application of a TagModule based approach to create mutant strains of *Shewanella oneidensis* MR-1, *Desulfovibrio alaskensis* G20 and *Desulfovibrio vulgaris* Hildenborough with dual unique 20nt molecular barcodes, we are leveraging these strain collections to perform pooled fitness assays, containing thousands of mutants in a single sample. Microarray hybridization or next generation sequencing can be used as a readout for the change in abundance of specific mutant barcodes within the pool under specific stress conditions.

To this end, we have recently designed and implemented an assay, based on pioneering work in yeast, which utilizes the sequencing of unique molecular barcodes on an Illumina platform as a quantitative readout for mutant strain fitness. This assay, as with a microarray readout, is species independent but will allow for massively increased throughput and dynamic range compared to existing approaches at a lower per experiment cost. Initial development of this assay utilized a collection of *S. oneidensis* mutants, each containing a different pair of molecular barcodes, due to the large quantity of microarray based chemogenomic fitness data available from this collection for comparative analysis. Preliminary assay data comprised over 173 million usable barcode reads for a single Illumina lane. This massive sequencing capacity has allowed us to utilize experimental indexes to analyze multiple conditions on the same Illumina lane. Using an initial 18-index multiplex, for an ~4000 mutant complexity pool, we were able to obtain an average of ~2000 barcode

* Presenting author

reads for each mutant present (typical range 700 to 3000). These data and previous work in yeast suggests that at minimum, 48-index multiplexing (24-48 stress conditions per lane) is feasible. As our Illumina library preparation protocol optimization is complete, we will apply this technique to perform chemogenomic fitness profiling of single gene mutants of *Desulfovibrio alaskensis* G20 and *Desulfovibrio vulgaris* Hildenborough to highlight the hypersensitivity and resistance of strains to stress conditions, chemical agents and diverse metabolic conditions, thereby providing critical insight into gene function and pathway composition.

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.

196

ENIGMA Biotechnology: Identification of Differentially Expressed Metabolites in *Desulfovibrio* and Communities Using Meta-Analysis and the METLIN Database

Richard Baran,[1] Kevin Cho, [2] Benjamin Bowen,[1] Paul Adams,[1] Adam A. Arkin,[1] Adam M. Deutschbauer,[1] Linh Hoang,[2] Ewa Kalisiak,[2] Nitin S. Baliga,[3] Zhengjiang Zhu,[2] Ralf Tautenhahn,[2] Trent R. Northen,[1] and **Gary Siuzdak**[2]* (siuzdak@scripps.edu)

[1]Lawrence Berkeley National Laboratory, Berkeley Calif.; [2]The Scripps Center for Mass Spectrometry, La Jolla Calif.; and [3]Institute for Systems Biology, Seattle Wash.

http://enigma.lbl.gov
http://masspec.scripps.edu/

Project Goals: Our objectives are to examine microorganisms and their communities and measure both their chemical input and output as a read out of specific biochemical activity. This effort will build on our previous experience of complex microbial systems by examining how bacteria communities rely on symbiotic relationships for survival and reproductive success, and how these relationships consequently affect their biochemical capabilities.

Metabolomics has emerged as a powerful tool to interrogate cellular biochemistry at the systems level by tracking alterations in the levels of small molecules. The success of metabolomics over the past decade has largely relied on advances in mass spectrometry instrumentation, coupled with developments in bioinformatic tools such as XCMS-Online, it has now become relatively routine to comprehensively compare the levels of metabolite peaks. To facilitate metabolite identification within the ENIGMA program we have developed a freely accessible metabolite database called METLIN (http://metlin.scripps.edu) which contains tandem mass spectral data from thousands of metabolites. This repository allow investigators to compare MS2 data from their research samples to MS2 data from model compounds catalogued in the database and thereby improve the speed, efficiency, and cost effectiveness of untargeted studies. One approach to define cellular dynamics with respect to alterations in of small molecules has been to consider metabolic flux. While flux measurements have proven effective for model organisms, acquiring multiple time points at appropriate temporal intervals for many sample types is challenging. As an alternative, meta-analysis provides another strategy for delineating metabolic cause and effect perturbations. This combination of metabolomic data from multiple genotypes or environmental conditions enables the association of specific changes in small molecules with unique phenotypic alterations. We recently developed metabolomic software called metaXCMS to automate these types of higher order comparisons. Here we discuss the utility of metaXCMS for analyzing proteomic datasets and highlight the biological value of combining meta-results from both metabolomic and proteomic analyses. The combined meta-analysis has the potential to facilitate efforts in functional genomics and the identification of metabolic disruptions related to specific phenotypes. In particular, we present this approach to characterize strains of *Desulfovibrio alaskensis* and microbial communities. The output of these experiments include the identification of novel endogenous metabolites as well as proteins uniquely associated to a range of pathways including glycolysis, the citric acid cycle, the urea cycle, and select amino acid metabolism.

Informatics Development. Our **XCMS/Metlin** platform consists of a continuously evolving technology that has become the world standard for metabolomics data analysis. **Metlin (http://metlin.scripps.edu/)** the largest tandem mass spectrometry metabolite database, has had over 10 million hits containing over 45,000 structures, and MS/MS data on over 5000 metabolites. We are currently developing Metlin with the addition of ENIGMA bacterial metabolites as well as extending its functionality within **XCMS**. XCMS is the most cited metabolomics software tool with over 50,000 downloads. **XCMS ONLINE** is being developed (**https://xcmsonline.scripps.edu/**) to facilitate the analysis of microbial studies and already has over 1500 users in 36 countries. **XCMS/Metlin** is being designed with an easy to use command driven interface with direct connection to **Metlin** for molecule identification. XCMS ONLINE is also being developed for bacterial analysis to allow for second-order ("meta") analysis (**Figure 1**). MetaXCMS (Anal. Chem. 2011, **Nature Protocols** in press) allows for data reduction when analyzing multiple sample groups to provide information on key proteins/metabolites related to a phenotype (such as *Desulfovibrio desulfuricans* metal reduction).

Proteomics/Metabolomics Analytical Development. Our mass spectrometry based technology developments include quantitative **QqQ MRM** metabolomics methods, **LC/MS/MS** shotgun high throughput protein analysis (**Nature 2010**) as well as novel **CESI-MS** (capillary electrophoresis mass spectrometry). In particular, we are applying these methods to genetically characterized strains of *Desulfovibrio desulfuricans* including the G100 wild type and the G20 mutant. The output of these experiments include the identification of novel endogenous metabolites (agmatine, cytosine, and 7-hexadecenoic acid) and 66 proteins uniquely present in the G100 wild type organism. These approaches are being developed for ENIGMA are allowing us to examine glycolysis, the citric acid cycle, the urea cycle, and select amino acid metabolism among other pathways.

Community Flux by Pulse labeling with a Biological Event. Although very biologically informative, isotopic pulse labeling flux experiments are extremely difficult in the context of mass spectrometry. Especially when examining systems that are largely unknown, stable isotopes cause a shift in mass readout that complicates databases and limits analysis by current bioinformatic software. Therefore, as an alternative to stable isotopes, we are pulse labeling with a biological event to examine temporal changes with respect to that event. These experiments are being performed with different populations of the community where the primary purpose is to identify the key biochemical perturbations that occur as a response to that particular event.

References

1. Patti G.J., Yanes O., Siuzdak G. Metabolomics: leading from behind *Nature Reviews, in press*, 2012.
2. Patti G.J., Yanes O., Shriver L., Courade J-P., Tautenhahn R., Manchester M., Siuzdak G. Metabolomics Implicates Altered Sphingolipid Metabolites *Nature Chemical Biology, in press*, 2012.
3. Patti G.J., Tautenhahn R., Siuzdak G. Meta-Analysis of Untargeted Metabolomic Data from Multiple Profiling Experiments *Nature Protocols, in press*, 2012.
4. Patti G.J, Tautenhahn R., Fonslow B.R., Cho Y., Deutschbauer A., Arkin A., Northen T., Siuzdak G. Meta-Analysis of Global Metabolomics and Proteomics Data to Link Alterations with Phenotype *Spectroscopy, in press (cover article)*, 2012.
5. Tautenhahn R., Patti G., Kalisiak E., Miyamoto T., Schmidt M., Fan Y.L., McBee J., Baliga N.S., Siuzdak G. metaXCMS: Second-Order Analysis of Untargeted Metabolomics Data *Analytical Chemistry*, 2011, 83, 696-700.
6. Yanes O., Tautenhahn R., Patti G., Siuzdak G. Expanding Coverage of the Metabolome for Global Metabolite Profiling *Analytical Chemistry*, 2011, 83, 2152-2161.
7. Greving M.P., Patti G.J., Siuzdak G. Nanostructure-Initiator Mass Spectrometry (NIMS) Metabolite Analysis and Imaging *Analytical Chemistry*, 2011, 83(1), 2-7.
8. Long J.Z., Cisar J.S., Miliken D., Niessen S., Wang C., Trauger S.A., Siuzdak G., Cravatt B.F. Metabolomics annotates ABHD3 as a regulator of medium-chain phospholipids *Nature Chemical Biology, in press*, 2011.

This work conducted by ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) 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.

197

ENIGMA Biotechnology: Membrane Protein Complexes—Their Roles in *Desulfovibrio vulgaris* Stress Response and in the Establishment and Maintenance of Communities

Peter J. Walian[1]* (PJWalian@lbl.gov), Simon Allen,[2] Max Shatsky,[1] Lucy Zeng,[1] Evelin D. Szakal,[2] Haichuan Liu,[2] Bonita Lam,[1] Jil T. Geller,[1] Kristina L. Hillesland,[3] Steven C. Hall,[2] Susan J. Fisher,[1,2] **Matthew W. Fields**,[4] **David A. Stahl**,[3] **Terry C. Hazen**,[5] **Steven E. Brenner**,[6] **Adam M. Deutschbauer**,[1] **Trent R. Northen**,[1] **John-Marc Chandonia**,[1,6] **H. Ewa Witkowska**,[2] **Mark D. Biggin**,[1] **Bing K. Jap**,[1] and **Paul D. Adams**[1]

[1]Lawrence Berkeley National Laboratory; [2]University of California, San Francisco; [3]University of Washington; [4]Montana State University; [5]University of Tennessee, Knoxville; and [6]University of California, Berkeley

http://enigma.lbl.gov

Project Goals: Key aims of this project are—to develop a system for the high-throughput isolation and identification of membrane protein complexes, optimizing this process for effectiveness across a range of sample types including planktonic cultures and biofilms; to apply this system in the study of DOE relevant microbes such as *Desulfovibrio vulgaris* in order to detect and characterize changes in their membrane protein complexes brought about by environmental stressors, and through the role of these proteins in the establishment and maintenance of communities.

A central goal of the ENIGMA consortium is to develop robust molecular-level models capable of predicting how target microbes respond to a range of environmental conditions. In support of this goal, our interests within ENIGMA have centered on the dynamic role of membrane protein complexes in this process. Cell membranes represent the "front-line" of cellular defense and the interface between a cell and its environment. Significant changes in response to environmental conditions are expected to take place through the proteins situated within these membranes. Membrane protein-associated changes may occur in the form of abundance level, protein-protein interactions, post-translational modifications and even mutations. To understand some of the earliest and perhaps most critical responses to stress, characterization of these changes on a molecular level is needed.

The study of membrane proteins presents a major challenge in protein biochemistry; to address this we have developed a unique high-throughput process for the isolation and identification of untagged membrane protein complexes that features mild, but effective, detergent solubilization, liquid chromatography and native electrophoresis methods. We have been applying this system in two main areas of

* Presenting author

investigation, one of which has been our work on developing a *D. vulgaris* membrane protein complex database covering standard and stressed growth conditions, and the second to characterize the roles of selected membrane proteins in the establishment and maintenance of communities.

Our study of membrane protein complexes in the outer-membrane of *D. vulgaris* grown under standard conditions is complete and we are at an advanced stage with the inner-membrane component. An interactome of proteins identified in *D. vulgaris* outer-membrane preparations is in the final stages of refinement. These preparations have yielded 69 outer-membrane protein identifications (which is over 80% of the number expected); 90% of these proteins were found to be in complexes. The most prevalent categories of proteins detected were the lipoproteins, and proteins with non-specific annotations (hypothetical and conserved hypothetical). This compendium of *D. vulgaris* outer-membrane protein complexes will serve as an essential reference for the detection and characterization of environment-driven changes in these proteins. Processing of outer-membrane proteins from stressed *D. vulgaris* cultures (including growth to stationary phase, and growth under elevated levels of nitrate or NaCl) has recently been completed. Initial analysis of stress-associated changes in outer-membrane protein abundance suggests that for many proteins there are significant differences between these changes and the changes in expression level inferred from mRNA experiments. Efforts on the preparation of the inner-membrane protein interactome and completion of the analysis of stress-induced changes occurring in the outer-membrane proteins of *D. vulgaris* are on-going.

In addition to our studies on large-volume planktonic monocultures, we have refined the methods employed in the pipeline so that they may be used to process samples derived from more native-like sources. Recent improvements made in pipeline sensitivity and resolution are now allowing us to work successfully with relatively small cell samples such as biofilms. We anticipate that through additional optimization of the system, we will be able to process yet smaller samples, not only cultured in the lab but obtained directly from field sites. To assess the potential for discovery from such sources, we have begun pilot studies on biofilm samples. Early results contain evidence of protein changes occurring during the transition from stationary phase to biofilms, suggesting that this will be a productive direction for future studies of microbial communities. Recent work by the Stahl group on adaptive evolution experiments with co-cultures of *Methanococcus maripaludis* and *D. vulgaris,* has identified mutations of soluble and membrane protein genes likely to be important in establishing the syntrophic mutualism between these species. These results suggest a large influence of acquired membrane protein mutations leading to improved growth rates within this co-culture community. We have begun to process membranes from clonal isolates to characterize changes in the *D. vulgaris* membrane protein population. Interestingly, the most abundant protein of the *D. vulgaris* outer-membrane (DVU_0799) is also the most consistently mutated protein in these experiments. Therefore, a key goal of ours will be to purify and functionally characterize those mutated membrane proteins found to play a role in facilitating improved rates of growth.

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.

198

ENIGMA Biotechnology: Biofilm Imaging: From Protein Complexes to Intact Microbial Communities

M. Auer,[1] Amita Gorur[1]* (agorur@lbl.gov), P. Arbeleaz,[2] N. Baliga,[3] D.A. Ball,[1] M. Biggin,[1] J.M. Chandonia,[1] S. Chhabra,[1] R. Csencsits,[1] K.H. Downing,[1] M.W. Fields,[4] J.T. Geller,[1] R. Glaeser,[1], T.C. Hazen,[1] G.L. Hura,[1] T. Juba,[5] B. Lam,[1] C.M. Leung,[1] J. Liphardt,[2] J. Malik,[2] J.P. Remis,[1] S. Reveco,[1] J. Tainer,[1] A. Tauscher,[1] J. Wall,[5] A. Deutschbauer,[1] T. Northen,[1] A. Arkin,[1] and **P.D. Adams**[1]

[1]Lawrence Berkeley National Laboratory, Berkeley Calif.; [2]University of California, Berkeley; [3]Institute for Systems Biology, Seattle, Wash.; [4]Montana State University, Bozeman; and [5]University of Missouri, Columbia

Project Goals: The ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) project aims to elucidate the mechanisms and key processes that enable microorganisms and their communities to function in metal-contaminated soil sites. The ENIGMA Biotechnology Component focuses its efforts on providing cross-cutting technologies that will support all other ENIGMA components with a particular focus on biological imaging at different scales of bacteria and microbial communities. These data will help to develop models of microbial community activity and principles of community organization in an effort to predict the role that microbial species and their interactions play in the dynamics of geochemical transformations in a changing environment.

Microbial physiology is inherently a multiscale biological process that coordinates complex processes such as extracellular metal reduction and response to environmental stresses and competing species. Bacteria often assemble into sustainable communities that allow individual bacteria to coordinate their respective behavior and thus optimizing the efficiency of biological processes, which may enhance the chances for species survival.

ENIGMA is addressing the complexity of multiscale spatiotemporal biofilm organization through a combination of expertise in traditional structural biology and modern multimodal imaging. SAXS (Rambo *et al.* 2010) and single particle cryo-EM are proven technologies to determine protein complex stoichiometry and shape, allowing the fitting of high-resolution structures into the intermediate resolution density envelope (Han *et al.* 2009). Cryo-electron tomography of bacterial whole mount samples can detect intra- and extracelluar specializations e.g. those important for metal reduction. Cryo-EM analysis is complemented by widefield 2D section TEM and advanced 3D SEM imaging approaches (FIB/SEM and SBF/SEM) of cryo-preserved, freeze substituted and resin-embedded samples. With these novel EM imaging approaches, we have begun to examine large areas and volumes of biofilms in DvH and other soil bacteria. We have found outer membrane vesicles, vesicle chains and cell-cell connections (Palsdottir *et al.* 2009, Remis *et al*. 2010, Remis *et al.* submitted), as well as compartmentalization of metal precipitation (Auer, unpublished observation). These observations suggest a an intricate set of interactions and possibly coordination of function between community members. X-ray and EM-based imaging approaches are complemented by tag-based labeling of proteins both at the light and electron microscopy level, and allow the study of cell-to-cell variations in protein abundance and protein localization (Chabra *et al.* 2010). Advanced optical super-resolution imaging methods (including PALM and STORM) allow high precision localization and counting (Betzig *et al* 2006). Further integration of small molecule mass spectrometry imaging, while at a somewhat larger size scale, promises to link structural observation and protein localization with metabolic activity of biofilm regions. Through the integrated application of these imaging modalities ENIGMA is deconstructing a mechanistic understanding of biofilm function.

Acknowledgement: This work conducted by ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) 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. Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, Davidson MW, Lippincott-Schwartz J, Hess HF. Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. Science 2006 313 (5793): 1642–1645
2. Chhabra SR, Butland G, Elias DA, Chandonia JM, Fok OY, Juba TR, Gorur A, Allen S, Leung CM, Keller KL, Reveco S, Zane GM, Semkiw E, Prathapam R, Gold B, Singer M, Ouellet M, Szakal ED, Jorgens D, Price MN, Witkowska HE, Beller HR, Arkin AP, Hazen TC, Biggin MD, Auer M, Wall JD, Keasling JD. Generalized schemes for high-throughput manipulation of the *Desulfovibrio vulgaris* genome. Appl Environ Microbiol. 2011 Nov;77(21):7595-604
3. Han BG, Dong M, Liu H, Camp L, Geller J, Singer M, Hazen TC, Choi M, Witkowska HE, Ball DA, Typke D, Downing KH, Shatsky M, Brenner SE, Chandonia JM, Biggin MD, Glaeser RM. Survey of large protein complexes in *D. vulgaris* reveals great structural diversity. Proc Natl Acad Sci U S A. 2009 Sep 29;106(39):16580-5.
4. Palsdottir H, Remis JP, Schaudinn C, O'Toole E, Lux R, Shi W, McDonald KL,Costerton JW, Auer M. Three-dimensional macromolecular organization of cryofixed Myxococcus xanthus biofilms as revealed by electron microscopic tomography. J Bacteriol. 2009 Apr;191(7):2077-82
5. Rambo RP, Tainer JA. Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. Curr Opin Struct Biol. 2010 Feb;20(1):128-37.
6. Remis JP, Costerton JW, Auer M. Biofilms: structures that may facilitate cell-cell interactions. ISME J. 2010 Sep;4(9):1085-7
7. Remis JP, Wei W, Gorur A, Allen S, Witkowski HE, Costerton JW, Auer M Bacterial Social Networks: Targeted Delivery of Outer Membrane Proteins via Membrane Vesicle Chains, submitted

199

ENIGMA Computation: An Integrated Framework of Databases and Software for Exploring Genes, Genomes, Proteomes, and Networks in Microbes and Communities

J. Christopher Bare[1]* (Christopher.Bare@systemsbiology.org), Paramvir S. Dehal,[2] Karen Foley,[1] Keith Keller,[2] Morgan N. Price,[2] Wei-Ju Wu,[1] Serdar Turkarslan,[1] Jason Baumohl,[2] Marcin P. Joachimiak,[2] John T. Bates,[2] David J. Reiss,[1] Pavel S. Novichkov,[2] **Adam P. Arkin**,[2] **Nitin S. Baliga**,[1] and **Paul D. Adams**[2]

[1]Institute for Systems Biology, Seattle, Wash. and
[2]Lawrence Berkeley National Laboratory, Berkeley Calif.

http://microbesonline.org

Project Goals: Currently, one of the grand challenges in biological discovery is not the production of data but it is the data processing and analysis. Often, algorithms or computational tools exist tailored to the analysis of specific datasets yet they lack modularity to handle new or non-standard data formats, automation for non-expert usage or connectivity with downstream applications. Therefore it is critical to develop biology driven accessible computational frameworks that will promote, support and validate biological discovery. In order to tackle this grand challenge, the goals of this project is to i) integrate existing tools and ongoing efforts for algorithm, software and database development within ENIGMA, ii) develop an automated and modular framework for storing, analyzing and visualizing biological data for microbes or communities iii) promote knowledge transfer within ENIGMA and in the scientific community by providing a suite of computational resources.

While current technological advances enabled production of massive amounts of biological data, the real challenge is to connect these data sources to biological discovery. Here we describe our efforts to develop a modular and integrated framework for storage, automated analysis and visualization of biological datasets. Using Microbes Online as a source of high-quality curated data, we infer regulatory networks for several organisms. Feeding gene expression data and experimental metadata through the cMonkey and Inferelator, network inference algorithms produce, for a given organism, a set of co-expressed and putatively co-regulated modules. These networks are archived in the Network Portal, which provides a web interface for searching, visualizing and annotating the networks. The Network Portal will be tightly inte-

* Presenting author

grated with Microbes Online, which already provides a rich set of features for comparative microbial genomics as well as RegPrecise for curation of regulatory inferences. Additional experimental validation of network inferences can be performed using the MicrobesOnline Fitness Experiment data. Automation of the network inference pipeline will enable us to capture phylogenetic diversity to drive comparative studies and availability of high quality fitness data will help us to asses and curate network inferences. In addition, the Gaggle integration framework provides interoperability with several popular bioinformatics tools including Cytoscape, MeV, and R. Firegoose provides Gaggle connectivity to web resources such as KEGG, DAVID and EMBL STRING. This suite of integrated analysis and visualization tools provides a powerful and easily extended environment for the study of the regulatory systems of microbial biology.

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.

200

ENIGMA Computation: Inferring Correlation Networks from Genomic Survey Data

Yonatan Friedman and **Eric J. Alm*** (ejalm@mit.edu)

Computational and Systems Biology Initiative, Massachusetts Institute of Technology, Cambridge

High-throughput sequencing based techniques, such as 16S profiling, have the potential to elucidate the complex inner workings of natural communities - be they the world's oceans or the human gut. A key step in exploring such data is the identification of dependencies between members of these communities, which is commonly achieved by means of correlation analysis.

However, it has been known since the days of Karl Pearson that the analysis of the type of data produced by such techniques (referred to as compositional data) can produce unreliable results since the observed data take the form of relative fractions of genes or species, rather than their absolute abundances.

Inferring correlation networks between genes or species is a common goal in microbial ecology. However, such networks are typically derived from genomic survey data, such as those obtained from 16S sequencing, which are subject to underappreciated mathematical difficulties that can undermine standard data analysis techniques. We show that these effects can lead to erroneous correlations despite the statistical significance of the associations. To overcome these difficulties, we developed SparCC; a novel procedure, tailored to the properties of genomic survey data that allow inference of correlations between genes or species.

201

ENIGMA Computation: Multiscale Modeling of Gene Regulatory Networks Across Evolutionary Timescales by Integrating "Top-Down" and "Bottom-Up" Approaches

Pavel Novichkov,[1]* David J. Reiss[2]* (dreiss@systemsbiology.org), Marcin Joachimiak,[1] Alexey Kazakov,[1] Serdar Turkaslan,[2] **Inna Dubchak**,[1] **Adam P. Arkin**,[1] and **Nitin S. Baliga**[2]

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif. and [2]Institute for Systems Biology, Seattle, Wash.

http://enigma.lbl.gov/ENIGMA_InformaticsKnowledgebase.html
http://baliga.systemsbiology.net/enigma

Project Goals: Environmental conditions and their associated cellular states dynamically shape regulatory responses. As the environment changes, the solution space for regulation of the same gene (or set of genes) moves between different states depending on the combinatorial interplay between transcription factors, their ligands, and binding sites. Such responses are encoded in gene regulatory networks by specific cis-regulatory signals encoded in the genome. We have developed a novel strategy for reverse engineering microbial gene regulatory networks that times context-specific system-scale regulatory phenomena directly to mechanism at single nucleotide-level resolution.

As a first step toward this goal we have integrated the data-driven and genomic-driven regulon discovery approaches developed within the ENIGMA team. The user can start from an automatic gene regulatory network inference conducted with the ***cMonkey*** integrated biclustering algorithm. ***cMonkey*** generates biclusters (predicting putative regulons with corresponding putative condition-specific cis-regulatory motifs) which can then be directly input to ***RegPredict*** for detailed comparative genomics analysis in order to verify evolutionary conservation of transcription factor binding sites (TFBS) and to refine the regulon content. ***RegPredict*** automatically suggests the most appropriate set of closely related genomes required for comparative analysis. Finally, the refined regulon can be submitted to the ***RegPrecise*** database, a community repository of manually curated regulons.

We have used this strategy to build transcriptional regulatory network models for ENIGMA keystone organisms *D. vulgaris* Hildenborough and *M. maripaludis*. The workflow was tested on several regulons (Rex, Fur, FliA, PerR, SahR) from *D. vulgaris* Hildenborough. For example, ***cMonkey*** predicted the Fur regulon and its corresponding TFBS, which was then refined and 19 additional gene members added using the comparative genomics approach of ***RegPredict***. Based on this model, we are also developing a framework for automated network inference plugged into a Network Portal, which includes a suite of analysis and visualization tools.

The data-driven systems approach has proven powerful for automatic reconstruction of mid- to large-size regulons. However, it is more challenging to identify small regulons due to their small size, a property that limits both gene expression data comparisons and sequence motif detection. Our novel method enables us to tackle the small regulon portion of the regulatory network reconstruction problem. The method uses a new biclustering engine along with heuristics designed from properties of known regulons to identify small regulon patterns in phylogenetic profiles.

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.

202

ENIGMA Computation: Evolutionary and Experimental Evidence-Based Functional Annotation of Genes

Morgan N. Price,[1] **Adam P. Arkin**,[1,2] Ben Bowen,[3] Barbara E. Engelhardt,[4,5] Ameet Talkwalkar,[4] Jeffrey M. Yunes,[6] Michael L. Souza,[7] Gaurav Pandey,[8,9] Susanna Repo,[8] Michael I. Jordan,[4,10] **Steven E. Brenner**,[1,2,6,7,8]* (brenner@compbio.berkeley.edu), Nitin Baliga,[11] and Paul D. Adams[1]

[1]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]Department of Bioengineering, University of California, Berkeley; [3]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; [4]Department of Electrical Engineering and Computer Science, University of California, Berkeley; [5]Currently at Biostatistics and Bioinformatics Department, Duke University, Durham, N.C.; [6]Joint Graduate Group in Bioengineering, University of California, Berkeley and San Francisco; [7]Biophysics Graduate Group, University of California, Berkeley; [8]Department of Plant and Microbial Biology, University of California, Berkeley; [9]Currently at Mount Sinai School of Medicine, New York; [10]Department of Statistics, University of California, Berkeley; and [11]Institute for Systems Biology, Seattle, Wash.

http://enigma.lbl.gov
http://compbio.berkeley.edu

Project Goals: The ENIGMA SFA aims to understand the architecture of microbial communities from a molecular level, which requires understanding in detail the molecular biology of key organisms. Although sequencing the genomes of these organisms is now straightforward, determining the molecular function of genes remains a challenge. However, many genome sequences are now available, and rich genome-wide functional data is becoming available outside of traditional model organisms. Therefore, we are developing improved tools for using evolutionary comparisons and functional-genomic data to predict the molecular function of proteins.

Phylogenetic analysis has been employed to infer the molecular function of a target gene by finding a function that is consistent with the evolutionary history of the gene. Over the past decade, this has been recognized as a highly accurate approach, but its manual application requires laborious effort by a domain expert. We have developed the SIFTER method, which automates phylogenetic-based function annotation by finding the most likely assignments of functions to proteins given a phylogenetic tree, model of evolution, and known functions. SIFTER uses a Bayesian graphical model framework to propagate molecular functions across the tree in a way that is statistically rigorous and robust. SIFTER explicitly takes account of evidence quality, to account for the variable quality of annotations from different sources.

Benchmarking studies of SIFTER show that it outperforms other widely-used homology-based approaches. Recently, we improved the core SIFTER algorithm, enabling it to run on large and diverse protein families, to work on a genome-scale, and to participate in the Critical Assessment of Function Annotation in 2011. We are extending it to share information between protein families based on gene-gene "association" relationships such as protein-protein interactions, co-expression, co-fitness, genome proximity, or genetic interactions. In doing so, we will be able to incorporate a larger variety of experimental data developed by and applicable to the ENIGMA project than other prediction approaches. We hope that with these enhancements, SIFTER will be the first successful method to statistically incorporate both homology and association data.

Another challenge is to interpret large-scale "fitness" data or knockout mutant phenotypes that are becoming available for diverse microbes due to approaches such as tagged transposon mutagenesis or TnSeq. In a pilot study in *Shewanella oneidensis* MR-1, we were able to confirm many annotations and to revise the annotations of 40 genes or operons, but this required extensive manual curation. To streamline the analysis, we are developing heuristics to find "re-annotatable" proteins to focus the manual curation. We are also automating the comparison of fitness data to metabolic models; in principle, it should be possible to automate much of the manual curation that now goes into producing a high-quality metabolic model.

This work partially conducted by ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) was partially 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. Other support for SIFTER comes from NIH K22 HG00056, NIH R01 GM071749, DOE SciDAC BER KP110201.

* Presenting author

203

ENIGMA Computational Core Group: Progress and Directions

David Reiss,[1] Serdar Turkarslan,[1] Jason Baumohl,[3] Paramvir Dehal,[3] Marcin Joachimiak,[3] Keith Keller,[3] Morgan Price,[3] Max Shatsky,[3] Alexey Kazakov,[3] Pavel Novichkov,[3] Yonatan Friedman,[2] Sonia Timberlake,[2] Michael Souza,[4] John-Marc Chandonia,[3] Steven Brenner,[4] Inna Dubchak,[3] Adam P. Arkin,[3] **Eric Alm**,[2] **Nitin Baliga**[1]* (Nitin.Baliga@systemsbiology.org), and Paul D. Adams[3]

[1]Institute for Systems Biology, Seattle, Wash.; [2]Massachusetts Institute of Technology, Boston; [3]Lawrence Berkeley National Laboratory, Berkeley Calif.; and [4]University of California, Berkeley

Project Goals: The ENIGMA Computational Core Group is tasked with data management analysis and dissemination for the ENIGMA Scientific Focus Area. The group develops and maintains widely used computational tools including the MicrobesOnline, RegTransBase, and RegPrecise databases, as well as data integration and visualization tools like the Gaggle, which also support a wide external user base. Analytical efforts have focused on regulatory network prediction (from genome sequence data), network inference (from functional data), gene functional annotation, and statistical approaches for environmental genomics. Efforts in the Computation Core Group integrate all of the projects in ENIGMA. In addition to developing algorithms to find patterns within noisy high throughput systems biology data (e.g. transcriptome structure, peptide atlas, protein-DNA interactions, and biodiversity from NextGen sequencing data), the Computation Core integrates diverse data types to elucidate modular architectures across scales (e.g. community assemblages and regulons) and infers genetic and environmental influences that encode dynamical interrelationships across these modules (e.g. microbial community networks and GRNs).

Using comparative analysis we have tracked the evolutionary history of gene functions to understand how novel functions evolve. One level up, we have used proteomics data, high-resolution genome tiling microarrays, and 5' RNA sequencing to revise genome annotations, discover new genes including ncRNAs, and map dynamically changing operon structures of five model organisms including *Desulfovibrio vulgaris* Hildenborough, *Pyrococcus furiosis*, *Sulfolobus solfataricus*, *Methanococcus maripaludis* and *Halobacterium salinarum* NRC-1. We have developed machine learning algorithms to accurately identify protein interactions at a near-zero false positive rate from noisy data generated using tagless complex purification, TAP purification, and analysis of membrane complexes. Further, we have developed algorithms to analyze and assign significance to protein interaction data from bait pull-down experiments and integrate these data with other systems biology data through associative biclustering in a parallel computing environment. We will "fill-in" missing information in these interaction data using a "Transitive Closure" algorithm and subsequently use "Between Commonality Decomposition" algorithm to discover complexes within these large graphs of protein interactions. To characterize the metabolic activities of proteins and their complexes we are developing algorithms to deconvolute pure mass spectra, estimate chemical formula for m/z values, and fit isotopic fine structure to metabolomics data. We have discovered that in comparison to isotopic pattern fitting methods restricting the chemical formula by these two dimensions actually facilitates unique solutions for chemical formula generators. To understand how microbial functions are regulated we have developed complementary algorithms for reconstructing gene regulatory networks (GRNs). Whereas the network inference algorithms cMonkey and Inferelator developed enable de novo reconstruction of predictive models for GRNs from diverse systems biology data, the RegPrecise and RegPredict framework developed uses evolutionary comparisons of genomes from closely related organisms to reconstruct conserved regulons. We have integrated the two complementary algorithms to rapidly generate comprehensive models for gene regulation of understudied organisms. Our preliminary analyses of these reconstructed GRNs have revealed novel regulatory mechanisms and cis-regulatory motifs, as well as others that are conserved across species. Finally, we are supporting scientific efforts in ENIGMA with data management solutions and by integrating all of the algorithms, software and data into a Knowledgebase. For instance, we have developed the RegPrecise database (http://regprecise.lbl.gov), which represents manually curated sets of regulons laying the basis for automatic annotation of regulatory interactions in closely related species. We are also in the midst of scaling up MicrobesOnline to handle the growing volume of sequence and functional genomics data. Over the last year our efforts have been focused on providing support for additional genomic and functional genomic data types. Similarly, we have developed several visualization tools to help with the exploration of complex systems biology datasets. A case in point is the Gaggle Genome Browser (GGB), which was enhanced with visualizations for plotting peptide detections and protein-DNA binding alongside transcriptome structure, plus the ability to interactively filter by signal intensity or p-value. Finally, we recognize that future advances to computational infrastructure cannot be anticipated and new software will be developed as dictated by scientific needs within ENIGMA and elsewhere. To account for this reality of how software environments evolve we have made advances to the Gaggle and Firegoose framework that enables interoperability and integration of diverse software and databases. Specifically, we have updated the R-goose package which provides connectivity between several Gaggle compliant bioinformatics tools and R; and prototyped a JSON based upgrade to the Gaggle protocol to make this environment extensible and more language neutral than the previous Java-based protocol.

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.

* Presenting author

204

ENIGMA Environmental Overview: Field to Lab to Field

Terry C. Hazen[1,2]* (tchazen@utk.edu), **Matthew W. Fields**,[3] Jizhong Zhou,[4] Joy Van Nostrand,[4] Dwayne A. Elias,[2] Jennifer J. Mosher,[2] Steven D. Brown,[2] Tommy J. Phelps,[2] Mircea Podar,[2] Anthony V. Palumbo,[2] Richard A. Hurt,[2] Anup Singh,[5] **Eric Alm**,[6] Mark B. Smith,[6] Dominique C. Joyner,[1,2] Romy Chakraborty,[7] Boris Faybishenko,[7] Kara DeLeon,[3] Jil Geller,[7] Bonita Lam,[7] Tamas Torok,[7] Julian Fortney,[1,2] Sharon E. Borglin,[7] David Stahl,[8] Judy Wall,[9] **Adam P. Arkin**,[7] and **Paul D. Adams**[7] (PDAdams@lbl.gov)

[1]University of Tennessee, Knoxville; [2]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]Montana State University, Bozeman; [4]University of Oklahoma, Norman; [5]Sandia National Laboratories, Livermore, Calif.; [6]Massachusetts Institute of Technology, Boston; [7]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [8]University of Washington, Seattle; and [9]University of Missouri, Columbia

http://enigma.lbl.gov

Project Goals: ENIGMA working hypotheses: 1) Key transects in the environment provide constraints on community composition and activity that are discernable at multiple scales. The uranium/nitrate/pH gradient may provide one such transect and we will examine the communities and activities in different *in situ* (and laboratory simulated) uranium/chromium contaminated environments. 2) These environmental constraints change in time due in part to the structure and function of these communities and causal relationships may be discovered and quantified. 3) Community structure is established and maintained by varying factors that include transient populations, niche diversification, optimized interactions, and resource competition. We will be attempting to dissect these by field perturbation and time series studies and laboratory simulations of the environment. 4) There are important stable communities that achieve the above, which we can dissect at the level of molecular interactions. We will drive toward isolation of key organisms, discovery of key activities and interactions, and dissect cellular networks that control when and how these activities are expressed.

The implications of the above hypotheses suggest: 1) There is a core set of metabolic factors that are the direct effectors of metal-reduction in soil and, while exact microbial species may vary, the molecular network will be relatively constant. This effort includes the identification of the biomolecular determinants of metal reduction in key populations that directly and/or indirectly interact with other populations that have programmed responses to important environmental parameters. 2) There are particular variants of these keystone microbes that adapt them for different metals and different concentrations and these are deployed in planktonic and attached communities over time during the reductive and reoxidative processes following stimulation by substrate amendment. 3) At a particular site there are core, relatively stable sub-communities of microbes whose interspecific interactions are responsible for the stable reduction (and ultimate reoxidation) of metals. There is also a core community structure of necessary functional classes of microbes that form a stable "food web" to exploit the available energy in the environment nearly optimally. For example at the Hanford 100-H study site, microbes from soil samples differed from the corresponding groundwater organisms (even at the phyla level) and were more diverse (p=0.001). While many of the populations were observed in both groundwater and surrogate sediments, the respective matrices appeared to enrich for particular OTUs. Results do not indicate a large shift in dominant organisms in soil from pre- to post- injection, and this may be due to the microbes remaining dominant from the first stimulation. However, a prevalence of core genera and rare genera were observed across 34 samples while urban and rural genera were less abundant.

Background and Significance. ENIGMA is planning on shifting the principle field focus from Hanford to Oak Ridge. For the last 12 years ORNL has been characterizing, monitoring, and conducting field experiments at DOE's BER ORNL Field Research site. The focus has been on elucidating the mechanisms and efficacy of bioreduction and bioimmobilization of U, one of the DOE's most common waste site contaminants. This has involved a number of field studies including pump tests, hydrological modeling, characterization of sediment and groundwater and amendments of ethanol, bromide, and nitrate, etc. More than 300 wells have been established and characterized and are available for analysis in a searchable database.

Research and Design

- Overarching Driver: Determine microbial community structure and function in both *in situ* environments and constructed consortia—environment to the laboratory.
- Elucidate structure to function during key biogeochemical transformations—*immobilization of metals*
- Determine key succession events and mechanisms—*stability in the context of geochemical and thermodynamic constraints*
- Parameterize critical microbes at the phylogenetic and functional level in conjunction with key biogeochemical variables that together, impact and control environmental activities of interest (*e.g.*, metal-reduction; N flow; C flow)
- Identify key populations, directly and/or indirectly related to activities of interest.
- *In situ* and laboratory consortia will be used to explicate levels of biological organization from populations to proteins, and
- Models will be developed with various bioinformatic tools (*e.g.*, AdaptML, random matrix theory, multivariate statistics) commonly used for genes and proteins but applied to populations overlaid with geochemical parameters and engineering controls.

* Presenting author

The current tasks for the Environmental Core are as follows:

Task 1. Optimize omics protocols for environmental samples.

Task 2. Use existing data from the ORNL FRC database to design efficient field sampling studies that maximize the geochemical diversity of study sites. This strategy is expected to enhance the resolution of associations between microbial communities and key geochemical features. Toward this end, we have developed a Monte Carlo search algorithm to optimize site selection for geochemical diversity. We are currently looking for 100 wells that we can do a metagenome analysis in cooperation with the existing, planned SBR IFRC program at ORNL. *Overarching principle for Field Studies and Field Linked Resources:* With the cooperation of the ORNL IFRC and as a team, establish the hypotheses, experiments, sampling, analyses, and schedule as a test plan, including all protocols, amounts, and responsibilities. Prioritize everything so that time, money, shipping, sample limitations, and unforeseen contingencies will not limit delivery of milestones. This includes 50%, 75%, and 95% evaluations of test plans by the team prior to execution, and fatal-flaw analyses at each step.

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.

205

ENIGMA Environmental Microbiology: Microbial Community Dynamics in Groundwater and Surrogate Sediments During HRC® Biostimulation of Cr(VI)-Reduction

K.B. De Leon,[1,2] B.D. Ramsay,[2] D.R. Newcomer,[3] B. Faybishenko,[4] T.C. Hazen,[5,6] M.W. Fields[1,2]* (matthew.fields@erc.montana.edu), and **Paul D. Adams**[4] (PDAdams@lbl.gov)

[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 Tennessee; and [6]Oak Ridge National Laboratory

http://enigma.lbl.gov

Project Goals: The elucidation of bacterial community dynamics for both groundwater and sediment-associated communities over time and space during biostimulation for chromate reduction in relation to geochemical variables.

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(VI) 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®. A second injection occurred November 2008 and geochemical data collected throughout the study showed an overall decrease in nitrate, sulfate, and chromium(VI). Spatial and temporal water and soil samples (n=34) 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. Sequences were refined by length, primer errors, and Ns, and sequences with a high percentage of low Phred quality score values were removed. Python scripts were developed to filter the pyrotag data with respect to quality scores, and the filtering technique was validated with environmental samples. Soil samples differed from the corresponding groundwater (even at the phyla level) and were more diverse (p=0.001). While many of the populations were observed in both groundwater and surrogate sediments, the respective matrices appeared to enrich for particular OTUs. Predominant populations for the sediments were *Psuedomonas*, *Acidovorax*, *Clostridium*, *Aquaspirillum*, *Methylibium*, *Anaeromyxobacter* while predominant populations for groundwater were *Psuedomonas*, *Pleomorphomonas*, *Ramlibacter*, *Arthrobacter*, and *Herbaspirillum*. Genera observed only in the sediment included *Marinomonas* while genera observed only in the groundwater included *Desulfonauticus*, *Desulfomicrobium*, and *Syntrophobacter*. 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 34 samples while urban and rural genera were less abundant. There was a shift from *Acidovorax* to *Aquaspirillum* from upstream (non-stimulated) to downstream soil both pre- and post-injection. Surrogate soil samples indicated similar changes in the soil community in the injection (Well 45) and downstream (Well 41) wells across time, while water samples seem to indicate more of a pre- and post-injection grouping instead of gradual changes across time. Furthermore, while post-injection soil samples indicate a continuing dominance of *Aquaspirillum*, corresponding water samples indicate *Pseudomonas* as a dominant genus. For each well, HRC® injection resulted in increased diversity, but the greatest changes during stimulation occurred in the populations of mid-dominance either between wells or across time. These organisms could be important to consider as possible indicator species in future work.

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.

---

* Presenting author

206

ENIGMA Environmental: Succession of Hanford Groundwater Microbial Communities During Lactate Amendment and Electron-Acceptor Limitation

Dwayne A. Elias[1]* (eliasda@ornl.gov), Jennifer J. Mosher,[1] Tommy J. Phelps,[1] Mircea Podar,[1] Richard A. Hurt,[1] James H. Campbell,[1] Meghan M. Drake,[1] James G. Moberly,[1] Christopher W. Schadt,[1] Steven D. Brown,[1] Terry C. Hazen,[2] Adam P. Arkin,[3] Anthony V. Palumbo,[1] Boris A. Faybishenko,[3] and **Paul D. Adams**[3]

[1]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]University of Tennessee, Knoxville; and [3]Lawrence Berkeley National Laboratory, Berkeley, Calif.

http://enigma.lbl.gov

Project Goals: Bioremediation strategies involving *in-situ* microbial stimulation for Cr(VI) -reduction/immobilization are ongoing, but determining their relative success is complex. By using controlled laboratory conditions, the influence of individual variables on the successful community structure, dynamics and the metal-reducing potential can be studied. The goal of the current work was to determine the impact of lactate stimulation during sulfate limitation on the succession of a native microbial community. Triplicate anaerobic, continuous-flow reactors were inoculated with Cr(VI) contaminated groundwater from the Hanford, Washington 100-H area and incubated for 95 days to obtain stable, enriched communities. The microbial community structure shifted with a significant loss of diversity. Final communities were dominated by *Pelosinus* spp. and to a lesser degree, *Acetobacterium* spp. with small levels of other organisms including methanogens. The resultant diversity decreased from 63 genera within 12 phyla to 11 bacterial genera (from 3 phyla) and 2 archaeal genera (from 1 phylum). Isolation efforts attained four new strains of *Pelosinus* spp. Three of the 4 *Pelosinus* strains were capable of Cr(VI)-reduction and one also reduced U(VI). Under the tested conditions of limited sulfate, it appears that the sulfate-reducers, including *Desulfovibrio* spp., were outcompeted even though they are capable of fermentative growth. These results suggest that lesser-known organisms, such as *Pelosinus* spp., may play a more important role in metal-reduction than currently suspected.

Currently, a similar bioreactor experiment is underway and builds from the above findings. Hanford groundwater was enriched in duplicate bioreactors with lactate and; 1) no Cr(VI) to emulate the earlier work and establish a baseline, 2) 0.1 mg/L Cr(VI) to reflect the low plume concentrations nearing the Columbia river and 3) 3.0 mg/L Cr(VI) to reflect the source metal concentration. Multiple levels of analysis are now underway after a 105 day experiment. These include temporal measurements of; 16S rRNA gene pyrosequencing, gas and organic acid quantification by GC and HPLC, Geochip, Phylochip, metagenomics, metatranscriptomics, metaproteomics, soluble Fe(III), HFO, Cr(VI) and U(VI) reduction assays, metal uptake characteristics for 36 metals, small metabolite quantification, and fluorescent antibody cell counts for species of *Desulfovibrio*, *Geobacter*, *Methanococcus*, and *Pelosinus*. These results will help to determine if and how the added complexity of Cr(IV) will influence the microbial community structure and metabolism as well as revealing the any acclimation of the community to Cr and quantifying the relative reduction potential.

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. Oak Ridge National Laboratory is managed by University of Tennessee UT-Battelle LLC for the Department of Energy under Contract No. DE-AC05-00OR22725.

207

ENIGMA Environmental Microbiology: Single-Cell Analysis Platforms for Genomic Analysis of Uncultivable Environmental Microbes

Suzan Yilmaz[1]* (syilmaz@sandia.gov), Peng Liu,[1] Robert J. Meagher,[1] Yooli Light,[1] Adam P. Arkin,[2] Terry C. Hazen,[3] **Anup K. Singh**,[1] and Paul D. Adams[2]

[1]Sandia National Laboratories, Livermore, Calif.; [2]Lawrence Berkeley National Laboratory, Berkeley, Calif.; and [3]University of Tennessee, Knoxville

http://enigma.lbl.gov

Project Goals: We are developing a pipeline for single cell genomics that utilizes FISH (fluorescence *in situ* hybridization) for targeting species of interest, FACS (fluorescence activated cell sorting) for high throughput isolation of single cells, and MDA (multiple displacement amplification) for production of sufficient DNA from single cells for genome sequencing. This pipeline is being used for a number of collaborative projects in ENIGMA.

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 cannot link the activity deterministically to species. We are developing high-throughput tools for studying bacteria one cell at a time, allowing us to unravel the complex 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 nanoliter 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. We are using this pipeline to analyze water samples from DOE bioremediation sites (e.g., Hanford) to identify

* Presenting author

keystone organisms and link their functions to species. Furthermore, we are also using our single-cell genomics pipeline to complement the metagenomic sequencing efforts in ENIGMA. Metagenomic sequencing typically fails to achieve complete assembly and metabolic reconstruction of individual genomes in a complex community. Single-cell sequencing, together with metagenomics, makes it possible to assemble genomes of novel uncultivated organisms.

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.

208

ENIGMA Environmental Microbiology: Metagenomics-Enabled Understanding of Microbial Communities at DOE Contaminated Sites

Jizhong Zhou[1,2]* (jzhou@ou.edu), Zhili He,[1] Ye Deng,[1] Aifen Zhou,[1] Ping Zhang,[1] Qichao Tu,[1] Joy Van Nostrand,[1] Hao Yu,[1] Zhou Shi,[1] James Voordeckers,[1] Yongjin Lee,[1] Rong Song,[1] Liyou Wu,[1] **Matthew W. Fields**,[3] **David A. Stahl**,[4] **Judy D. Wall**,[5] **Terry C. Hazen**,[6,7] **Adam P. Arkin**,[2] and **Paul D. Adams**[2] (PDAdams@lbl.gov)

[1]University of Oklahoma, Norman; [2]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [3]Montana State University, Bozeman; [4]University of Washington, Seattle; [5]University of Missouri, Columbia; [6]University of Tennessee, Knoxville; and [7]Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://enigma.lbl.gov

Project Goals: Although high throughput sequencing and associated (meta)genomic technologies provide an avenue to determine genetic and organismal diversity of an ecosystem, linking the genetic/population diversity to phenotypic diversity across different organizational levels (e.g., molecular, cellular, populations, communities and ecosystems) is extremely challenging. As a part of the overall ENIGMA goal, the ultimate aim of this project is to utilize (meta)genomic technologies to better understand the mechanistic connections between molecular-level interactions/processes and community-level processes/functions. The following three specific objectives have been pursued: (i) To determine adaptation and molecular mechanisms of *Desulfovibrio vulgaris* Hildenborough (DvH) in response to multiple environmental stresses; (ii) To understand microbial community functional diversity at U/Cr-contaminated sites and develop high-throughput functional gene arrays (FGAs) for microbial community analysis; and (iii) To determine the responses, interactions, mechanisms and dynamics of groundwater/sediment microbial communities to U/Cr contamination and bioremediation treatments.

Long-term experimental evolution of DvH. To better understand the mechanistic connections between molecular-level functions and community-level processes, experimental evolution has been carried out to determine molecular mechanisms of DvH in response to high salinity. Significantly increased salt resistance was observed in evolved DvH (eDvH) with increased biomass, higher growth rate and shorter lag phases. Whole genome sequencing of eDvH at 1200 generations revealed specific point mutations and deletions. Their contribution to increased salt resistance has been proven by mutagenesis and phenotype analyses. Glu and Ala significantly increased in eDvH. After 5000 generations, the final biomass and growth rate of eDvH in the medium with high salinity was similar to that in the medium without extra salt. To further investigate the dynamics of evolution, repeatability of the evolution, whole genome sequencing, fitness assay and site-directed mutagenesis are in progress.

GeoChip-based metagenomic technology development. We have developed the GeoChip 4.0 series (4.0-4.2) for characterizing microbial communities. The GeoChip 4.0 series are manufactured based on the NimbleGen microarray format. For example, GeoChip 4.0 contains 120,054 distinct probes, and covers 200,393 coding sequences for 539 gene families in different microbial functional processes. The StressChip subset contains 22,855 probes covering 79,628 gene sequences for 46 genes involved in microbial responses to environmental stresses (e.g., temperature, osmolarity, oxidative status, nutrient limitation). The specificity, sensitivity and quantification of the developed GeoChip 4.0 series were evaluated computationally and experimentally. High specificity was observed for both synthesized oligonucleotides and genomic DNA from pure strains; the sensitivity was estimated to be 0.5 µg of DNA; the log(signal intensity) vs. log(DNA concentration) was highly correlated ($R = 0.925$). All the results showed that the GeoChip 4.0 series are specific, sensitive, and quantitative tools for characterizing microbial communities.

GeoChip applications. GeoChips have been used to study groundwater microbial communities to examine sustained reduction of contaminants using slow-degrading/slow-hydrolysis e-donors. At the Oak Ridge site, a one-time injection of emulsified vegetable oil (EVO) was used to examine U(VI) bioreduction and immobilization. Samples collected from the control and treatment wells (W1-7) were analyzed using GeoChip 3.0. Acetate, from EVO biodegradation, stimulated $NO3^-$, Mn(IV), Fe(III), $SO4^{2-}$, and U(VI) bioreduction in W1-7 and increased functional gene diversity. After EVO depletion, functional gene diversity declined. Fe(III)- and sulfate-reducing bacteria could play key roles in U(VI) reduction, whereas acetogens, denitrifiers and methanotrophs could be important for e-donor production and maintaining favorable reducing conditions. At the Hanford site, a one-time injection of poly-lactate was used to test Cr(VI) bioreduction. The groundwater microbial communities were monitored for 390 days using GeoChip 4.0. Cr(VI) was effectively reduced and functional gene diversity increased. Fe(III)- and sulfate-reducing bacteria could play key roles in Cr(VI) reduction, whereas denitrifiers could be important for maintaining reducing conditions.

Metagenomic sequencing. We have sequenced or resequenced metagenomes and dominant isolates from Oak Ridge FRC wells FW106 (contaminated with uranium, nitric acid, organics, and mercury) and FW301 (pristine). The FW106 metagenome was previously sequenced, and the results suggest extensive lateral transfer of metal resistance and organic compound metabolism genes. To extend these analyses, the metagenomes of FW106 and FW301, and genomes of multiple isolates of the dominant *Rhodanobacter* strain found in FW106 were sequenced or resequenced using Illumina sequencing technology. We are currently conducting comparative analyses of FW106 to FW301 to identify ecological trends observed between pristine and highly stressed groundwater communities. Furthermore, we are comparing the FW106 metagenome to *Rhodanobacter* isolates to confirm predicted lateral transfer events.

Molecular ecological network analysis. A novel random matrix theory-based approach has been developed to construct molecular ecological networks (MENs) based on GeoChip hybridization or high-throughput sequence data. Various mathematic and statistical tools and methods have been integrated into a comprehensive MEN analysis pipeline (MENAP). We have applied this approach to construct and analyze MENs from the Oak Ridge EVO experiment described above. Functional MENs were constructed from three GeoChip datasets: (i) Early EVO injection (≤31 days), (ii) Days 80 to 140, and (iii) Control well (7 time points) and pre- injection (0 day). All three constructed functional MENs posed general network characteristics (scale-free, small world and modularity), and the topology of these functional MENs was distinctly different, suggesting that the interactions among different microbial functional groups/populations in each community were dynamically altered during uranium bioremediation. Additionally, the changes in network structure were significantly correlated with environmental geochemical dynamics and EVO concentrations.

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.

209

ENIGMA Environmental Microbiology: Effect of Nitrate Stress on Metal-Reducing Microbes and Results of Nitrate Push/Pull Field Tests at Hanford 100H

Romy Chakraborty[1]* (rchakraborty@lbl.gov), Sharon E. Borglin,[1] Boris Faybishenko,[1] Paramvir Dehal,[1] **Adam P. Arkin**,[1] **Terry C. Hazen**,[2,3] **Matthew W. Fields**,[4] Jil Geller,[1] Julian Fortney,[2,3] Dominique Joyner,[2,3] Mark Conrad,[1] and **Paul D. Adams**[1] (PDAdams@lbl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]University of Tennessee, Knoxville; [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [4]Montana State University, Bozeman

http://enigma.lbl.gov

Project Goals: The ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) project seeks to elucidate the mechanisms and key processes that enable microorganisms and their communities to function in metal-contaminated soil sites. One goal is to understand the effect of environmental stressors that enable ENIGMA-relevant microorganisms to thrive in such environments.

As part of the ongoing investigation of sustainable bioremediation of Cr(VI) in groundwater at the Hanford 100H area, we performed groundwater biostimulation tests by injecting Hydrogen Release Compound (HRC) and three lactate (17mM) injection experiments. To investigate the response of resident microbes to nitrate stress, a push-pull test was then conducted by injecting 55 gals of groundwater (collected from the background well) with KNO_3 (nitrate concentration 5,000 ppm) in October, 2010. After one day, pumping began from the same well, and lasted for 16 days. As a result of nitrate injection, total biomass decreased and sulfate concentration increased, but the sulfide and iron concentrations dropped. During pumping, the nitrate concentration decreased about 3 orders of magnitude. PLFA data showed biomass on the order of 10^7 cells/ml prior to push pull, and dropping off to 10^5 cells/ml during the test, but recovering toward the end back to 10^7 cells/ml by the end. Biomarker lipids indicate a shift toward monoenoics indicating an increase in gram negative bacteria and decrease in branched lipids (gram postive) and branched monoenoic (sulfate reducers).

We discuss the field test results in details and elucidate the effect of nitrate stress on environmentally relevant microbes *Geobacter metallireducens* and *Desulfovibrio vulgaris* as observed through controlled lab experiments. The lab studies and the field study together will help in understanding the overall fate of microbes under changing environmental conditions in the field and the key cellular mechanisms impacted by such stress conditions.

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.

* Presenting author

210

ENIGMA Microbiology Physiology: Dissecting the Physiology and Community Interactions of Environmental Bacteria

Adam P. Arkin[1]* (APArkin@lbl.gov), Kelly Bender,[2] David Stahl,[3] Aindrila Mukhopadhyay,[1] Jan Liphardt,[1] Swapnil Chhabra,[1] Adam Deutschbauer,[1] Michael W.W. Adams,[4] Steve Yannone,[1] Mark Biggin,[1] Paramvir Dehal,[1] Marcin Joachimiak,[1] Kimberly Keller,[8] Birte Meyer,[3] James Moberly,[6] Morgan Price,[1] Lara Rajeev,[1] Trent Northen,[1] Matthew W. Fields,[5] Dwayne Elias,[6] Nitin Baliga,[7] **Judy Wall**,[8] and Paul Adams[1]

[1]Lawrence Berkeley National Laboratory; [2]Southern Illinois University; [3]University of Washington; [4]University of Georgia; [5]Montana State University; [6]Oak Ridge National Laboratory; [7]Institute for Systems Biology; and [8]University of Missouri

http://enigma.lbl.gov

Project Goals: The ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) project aims to elucidate the mechanisms and key processes that enable microorganisms and their communities to function in metal-contaminated soil sites. The ENIGMA Laboratory Component focuses its efforts on uncovering the genetic and biochemical basis of physiology and metabolism in key microbial species in isolation and in laboratory-defined communities. These data are used to develop models of microbial community activity and principles of community organization in an effort to predict the role that microbial species and their interactions play in the dynamics of geochemical transformations in a changing environment.

The mission of the Laboratory Component involves four primary goals that are closely coordinated with the Environmental, Biotechnology, and Computational Components within ENIGMA. Here, we describe these goals and highlight recent progress in these areas with a focus on the biology and interactions of the sulfate-reducing bacterium (SRB) *Desulfovibrio vulgaris* Hildenborough (DvH) in microbial consortia (1). In collaboration with the Environmental Component, we will characterize key microbes and communities enriched directly from metal-contaminated sites and prioritize these systems for deep functional annotation based on physiological properties, interactions among microorganisms, and potential biogeochemical activities. With flexible experimental techniques developed in concert with the Biotechnology component, we will use metabolomics, proteomics, transcriptomics, high-resolution microscopy, and high-throughput mutagenesis/phenotyping approaches to define transcript architecture, signal transduction, gene function, sRNA function, protein interactions and localization, elucidate the structure and function of biofilm formation, the metalloproteome, metabolism and mechanisms of specific interaction in these key isolates and their relatives (2). With the Computational group, we will use these data to construct predictive models of metabolism and gene regulation, which in turn will be used as drivers for new hypotheses and further experimentation (3). Using key microbial isolates, we will constitute laboratory simulation of enriched and defined assemblages in planktonic and attached states and assess different cellular compositions under varying environmental constraints. As a two species community model system, we have used long-term evolution experiments to map the genetic basis of improved syntrophy between DvH and a methanogen. More recently, transcriptomics and whole-genome mutant fitness profiling have identified key genes and pathways involved in the syntrophy of *Desulfovibrio alaskensis* G20 with a methanogen (4). These form a basis for pilot studies on the molecular dissection of competitive association among multiple SRB and methanogens in common culture. Lastly, we aim to determine how information flows among microbes and to assess the specificity, mutualism, and antagonism among pairs and groups of microbes by application of genetic, biochemical and computational technologies.

This work conducted by ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) 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.

211

ENIGMA Microbial Physiology: Systematic Mapping of Two Component Response Regulators to Gene Targets in a Model Sulfate-Reducing Bacterium

L. Rajeev,[1] E. Luning,[1] K. Keller,[2] G. Zane,[2] P. Dehal,[1] M. Price,[1] A. Arkin,[1] J. Wall,[2] and **A. Mukhopadhyay**[1]* (amukhopadhyay@lbl.gov)

[1]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif. and [2]Biochemistry Division, University of Missouri, Columbia

http://enigma.lbl.gov/

Project Goals: The goal of this project is to elucidate the mechanisms by which bacteria connect core metabolic functions to environmental cues and stresses. Such signaling remains poorly understood even in the best-studied model organisms. As part of near-term goals, we developed methods to understand signal transduction pathways and the corresponding regulatory networks in the sulfate-reducing bacterium, *Desulfovibrio vulgaris*. Specifically we focused on two component signal transduction systems. In ongoing research our goal is to understand the physiological relevance of the regulatory maps we have discovered. Long-term goals of this project are to extend our methods to evaluate multiple organisms that coexist in an ecological niche and deduce the connections between the environment and a microbial community that exists in it.

Desulfovibrio vulgaris is an environmentally relevant bacterium that serves as a model system for dissimilatory sulfate reduction. It is an important member of anaerobic syntrophic communities and is of interest for its metal reduction ability. The strain Hildenborough encodes a large number of two component regulatory systems, none of which are characterized. We sought to map the transcriptionally acting response regulators of these signal transduction systems to their gene targets. In order to accomplish this goal, we developed an in vitro DNA-affinity-purified-Chip method. We successfully determined 200 gene targets for 24 response regulators, which constitute the majority of this class of regulators in *D. vulgaris*. Our results enabled functional predictions and the identification of binding site motifs for several regulators (1). As expected, several simplex and complex regulatory modules were discovered. Of these an important regulatory network uncovered in our study is centered on the lactate utilization pathway, which appears to be under the control of multiple response regulators. The regulators include a lactate-responsive, a nitrite-responsive, a phosphate-responsive regulator and a potential oxidative stress responsive regulator. Here we present the comprehensive set of regulatory maps obtained using the DAP-chip method. Further, we describe the results from our experiments to evaluate the response of *D. vulgaris* carbon utilization pathway to various stresses such as nitrite and phosphate.

Reference

1. Rajeev L, Luning EG, Dehal PS, Price MN, Arkin AP, Mukhopadhyay A* Systematic mapping of two component response regulators to gene targets in a model sulfate reducing bacterium. *Genome Biology* 2011

This work conducted by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies 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.

212

ENIGMA Microbial Physiology: Evidence-Based Annotation of Gene Function in Metal-Reducing and Sulfate-Reducing Bacteria Using Genome-Wide Mutant Fitness Data

Adam Deutschbauer[1]* (AMDeutschbauer@lbl.gov), Grant M. Zane,[2] Samuel Fels,[2] Morgan Price,[1] Jennifer V. Kuehl,[1] Kelly Wetmore,[1] Swapnil R. Chhabra,[1] Thomas R. Juba,[2] Gareth P. Butland,[1] **Adam P. Arkin**,[1] **Judy D. Wall**,[2] and Paul Adams[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif. and [2]University of Missouri, Columbia

http://enigma.lbl.gov

Project Goals: The ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) project seeks to elucidate the mechanisms and key processes that enable microorganisms and their communities to function in metal-contaminated soil sites. One goal is to develop and apply high-throughput genetic technologies to identify causative genes and pathways rapidly in ENIGMA-relevant microorganisms.

The development of genetic tools to study the functions of thousands of poorly characterized genes in newly isolated environmental microbes is a rate-limiting step in determining the activity of the keystone organisms in communities. One promising approach to meet this challenge is the large-scale generation of mutant libraries and gene fitness assays as a complement to standard whole genome sequencing and classical annotations based on homology and computation. Currently, large-scale mutant collections have been assembled for only a small sample of typically pathogenic bacteria. With the availability of a universal, sequence-verified TagModule collection (Oh et al. NAR 38:e146, 2010) that has been introduced into a hyperactive mini-Tn*5* (Larsen et al., Arch Microbiol. 178:193, 2002), the creation of such libraries has been facilitated for a number of environmental microbes. We have generated sequenced and archived transposon mutant collections of 24,688 mutants of *Shewanella oneidensis* MR-1, 15,477 of *Desulfovibrio alaskensis* G20, and 13,007 of *Desulfovibrio vulgaris* Hildenborough. Simple scanning of the genes not mutated provides a limited list that includes essential genes for the condition of mutant isolation. Pools of tagged or barcoded transposon mutants of most non-essential genes can be used to probe the gene fitness in different growth conditions and when subjected to various stresses. Recent work with *S. oneidensis* (Deutschbauer et al., PLOS Genetics, 7:e1002385, 2011) has shown that many hypothetical genes and potentially redundant genes have detectable phenotypes when this approach is applied across a large number of diverse growth conditions. In some instances, specific molecular functions can be predicted (evidence-based annotations) where nearly meaningless annotations were the extent of previous knowledge.

This work conducted by ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) 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.

213

ENIGMA Laboratory: Evolutionary and Ecological Origins of Community Assembly, Stability, and Efficiency

Birte Meyer[1]* (birtemey@u.washington.edu), Kristina L. Hillesland,[1] Jason Flowers,[1] Nicolas Pinel,[2] Nicholas Elliott,[1] **Judy Wall**,[3] Marcin Joachimiak,[4] Aifen Zhou,[5] **Joe Zhou**,[5] Jennifer Kuehl,[4] **Adam Deutschbauer**,[4] Paramvir Dehal,[4] Morgan Price,[4] Zhili He,[5] **Aindrila Mukhopadhyay**,[4] Lara Rajeev,[4] Edward Baidoo,[4] **Trent Northen**,[4] **Nitin Baliga**,[2] **Mark Biggin**,[4] Ming Dong,[4] **Jap Bing**,[4] Peter Walian,[4] **Anup Singh**,[6] Suzan Yilmaz,[6]

Dwayne Elias,[7] **Matthew Fields**,[8] Hector Garcia-Martin,[4] **Terry Hazen**,[4] **Adam Arkin**,[4] and **David Stahl**[1]

[1]University of Washington, Seattle/Bothell; [2]Institute for Systems Biology, Seattle, Wash.; [3]University of Missouri, Columbia; [4]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [5]University of Oklahoma, Norman; [6]Sandia National Laboratories; [7]Oak Ridge National Laboratory; and [8]Montana State University, Bozeman

http://enigma.lbl.gov

Project Goals: A goal of DOE and ENIGMA is to understand and ultimately predict microbialcommunity assembly and the adaptive response of communities to environmental change. To address these questions, we are examining assemblies of simple two-member communitiescomposed of a secondary fermenter like *Desulfovibrio* species and hydrogenotrophic methanogens. This type of association (syntrophy) is representative of a trophic interaction sustaining both pollutant transformation and organic matter mineralization in many anoxicenvironments typical of the subsurface. To study the metabolic versatility and specificity ofthese assemblies, we first investigated the association of twelve different pairings of *Dv.* species and two methanogenic species (*Methanococcus maripaludis* and *Methanospirillum hungatei*). The results demonstrated that different *Desulfovibrio*-methanogen pairings vary significantly in their growth characteristics, most notably in their ability to ferment lactate at elevated hydrogen levels, presumably reflecting differences in their syntrophy-associated enzyme systems (e.g., hydrogenases and electron transfer complexes). Those studies now serve to direct a systems-level approach to the study of common and divergent features of community interaction: focusing on the genetic and metabolic signatures of efficient species interaction, major determinants of community stability, and the capacity for these communities to improve through adaptive evolution.

Comparative studies have so far shown that both the electron transfer system and themediator for electron transfer differ among *Desulfovibrio* species. For example, comparison ofdifferent assemblies grown in chemostats under various dilution rates demonstrated theimportance of formate as a major mediator of electron exchange in *Dv. alaskensis* strain G20 syntrophic cocultures in contrast to the hydrogen exchange-based system of *Dv. vulgaris* strain Hildenborough. Notably, the transcript analyses revealed that gene expression during syntrophic growth of *Dv. alaskensis* str. G20 also varies with both growth rate and the methanogenic partner. These conclusions were subsequently confirmed using a tagged-transposon *Dv. alaskenis* G20 mutant mini-library (1200 strains) to examine the relative fitness of differentinsertion mutants grown syntrophically in chemostats. Complementary studies are examiningthe adaptive evolutionary response of the two species to syntrophic growth. Ongoing laboratory evolution experiments of 24 replicated lines have so far documented a remarkable capacity for rapid improvement in the stability and efficiency of this mutualism after only 1000 generationsof cooperative growth. The genetic basis of improved cooperation is now being examined by genome resequencing, initially of twelve of the pairings at 1000 generations using both Illuminaand SOLiD next generation sequencing platforms and microarrays. Since some lines have evolved to obligate syntrophy, the history of their adaptive evolution will be reconstructed using single cell genome sequencing of earlier generations. The first mutations in *Dv. vulgaris* to become fixed in multiple lines were in an outer membrane porin (DVU0799), suggesting thatamino acid replacements near the outer face of this porin alters the flux of metabolites and/or nutrients. Other notable and frequent mutations were in genes implicated in EPS synthesis and regulation of lactate metabolism. Collaborative biophysical studies with ENIGMA members arenow exploring altered function of the different porin mutants and mutants implicated in EPSsynthesis. Of particular note was the replacement in all evolved lines of six amino acids in thesensory PAS domain of a histidine kinase (DVU3022) implicated in the regulation of lactatemetabolism. The high frequency replacement with an identical short amino acid sequencehaving predicted sensory function suggests a novel phase variation-like mechanism of adaptive response. Together these ongoing studies point to both common and divergent mechanisms ofinterspecies interaction, and offer a framework to better resolve genetic, structural, and metabolicfeatures contributing to stability and efficiency of community assembly.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biologicaland Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC0205CH11231

214

ENIGMA Microbial Physiology: Cooperation Impacts Structure in a Syntrophic Biofilm of *Methanococcus maripaludis* and *Desulfovibrio vulgaris*

Kristen A. Brileya,[1,2] Matthew W. Fields[1,2]* (matthew.fields@erc.montana.edu), Adam P. Arkin,[3] and **Paul D. Adams**[3] (PDAdams@lbl.gov)

[1]Center for Biofilm Engineering, [2]Department of Microbiology, Montana State University; and [3]Lawrence Berkeley National Laboratory

http://enigma.lbl.gov

Project Goals: Elucidate the relationship between structure and function in a methanogenic biofilm that consists of a sulfate-reducing bacterium and methanogenic archaeon. The functions of interest are carbon-compound oxidation and methanogenesis between syntrophic populations. The structure refers to population organization within a self-assembled, biofilm community.

Transfer of reduced carbon and electrons between microbial community members is of interest in anoxic systems, and methanogenesis represents a crucial trophic level that can include sulfate-reducing bacteria and methanogenic archaea. The current work uses a dual-culture approach to examine the structure of a syntrophic biofilm formed by the sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough and the methanogenic archaeon *Methanococcus maripaludis.* Biofilm was grown in a continuously stirred reactor where cells could attach to a silica surface or remain suspended. Under the tested conditions, *D. vulgaris* formed monoculture biofilm, but *M. maripaludis* did not. However, *M. maripaludis* did form pellicles in static, batch cultures while *D. vulgaris* did not form a pellicle. Under syntrophic conditions, a methanogenic biofilm formed and reached steady-state in approximately 7 days based upon protein levels and methane mass flux. Biofilm establishment was dependent upon initial colonization by *D. vulgaris* that was followed by recruitment of *M. maripaludis* into the biofilm matrix. The initial *Desulfovibrio*:*Methanococcus* biofilm ratio was approximately 375:1 but steady-state biofilm reached a ratio of 4:1. Steady-state biofilm was fixed for Fluorescence *in situ* Hybridization (FISH) and confocal laser scanning microscopy (CLSM). FISH revealed a framework of *D. vulgaris* with both single cells and large micro-colonies of *M. maripaludis* interspersed throughout the biofilm. 3D-FISH and CLSM of hydrated intact biofilm confirmed steady-state biofilm irregularity, with ridge, valley and spire macro-architecture. Key structural signatures were observed that confirmed the cooperative nature of the community using a newly developed model. Colorimetric assays indicated cell-associated carbohydrate was composed of .035 μg hexose/μg protein, .017 μg pentose/μg protein and .011 μg uronic acid/μg protein, similar to *D. vulgaris* mono-culture biofilm and approximately 5 times less than *M. maripaludis* pellicles. Filaments presumed to be protein have been observed in dual-culture biofilm matrix with electron and atomic force microscopy, and matrix was sensitive to proteinase K treatment during preliminary work with Catalyzed Reporter Deposition FISH. Syntrophic biofilm 3-D structure appears to be initialized by *D. vulgaris* that provides an advantageous environment for *M. maripaludis* to establish micro-colonies throughout the *D. vulgaris* scaffold. The coculture biofilm growth mode resulted in a 10-fold higher methane production per *M. maripaludis* biomass than the planktonic only growth mode, and these results indicated that the structure of cooperative interactions between a bacterium and archaeon positively impacted function.

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.

215

ENIGMA Microbial Physiology: Assimilatory and Dissimilatory Metallomics of *Desulfovibrio vulgaris* and *Pelosinus* Strain A11

W. Andrew Lancaster* (lancaste@uga.edu),[1] Israel Scott,[1] Brian Vacarro,[1] Angeli Lal Menon,[1] Farris L. Poole,[1] Jil Geller,[2] Jennifer J. Mosher,[3] Terry C. Hazen,[2,3] Dwayne A. Elias,[3] Michael W.W. Adams,[1] Adam P. Arkin[2] and Paul Adams[2]

[1]University of Georgia, Athens; [2]Lawrence Berkeley National Laboratory / ENIGMA, Berkeley, Calif.; and [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://enigma.lbl.gov/

Project Goals: The goals of the ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) project are to understand at a molecular systems biology level the microbial communities at DOE sites contaminated with heavy metals or radionuclides with sufficient detail to predictively model interactions within microbial and community processes that drive complex geochemistry in key environments. We expect to define biological principles governing selection of microbial community function and composition in given environments.

While some heavy metals are environmental contaminants, metals in general have a very positive role in biological systems as they afford proteins virtually unlimited catalytic potential, enable electron transfer reactions and greatly impact protein stability. Consequently, metal-containing proteins play key roles in virtually all biological processes. However, the full complement of metalloproteins within a given cell cannot be predicted solely from bioinformatic analyses of a genome sequence since metal coordination sites are diverse and poorly recognized. Hence it is not possible to predict the number and types of metals that a microorganism utilizes and how these might vary with the growth conditions. Determining the metals that are taken up during microbial growth and the metal content of fractionated native biomass can provide insight into these issues and, using coupled MS/MS analyses with extensive fractionations, can reveal completely new aspects of metal metabolism (1). Herein we report comparative metallomic analyses of the model microbe, *Desulfovibrio vulgaris* strain Hildenborough (DvH), and a newly described microbe, *Pelosinus* strain A11, which was recently isolated from an enrichment of a groundwater sample from the Hanford site by Elias and coworkers at ORNL. DvH was grown on lactate under sulfate-reducing conditions in a 600-liter metal fermenter and in a glass 5-liter fermenter and the nature of metals assimilated were compared. *Pelosinus* A11 was also grown at the 600-liter scale on fructose in the absence of sulfate and the assimilated metals were determined using ICP-MS (53 elements). Both organisms assimilated 20 or so metals into their cytoplasmic fractions, but the types and amounts were species specific. For example, DvH assimilated cobalt, cadmium and tungsten into high molecular weight complexes (>3 kDa) but *Pelosinus* A11 did not. In contrast to DvH, *Pelosinus* A11 assimilated vanadium, copper and uranium into >3kDa complexes. A number of additional factors affecting metal metabolism were analyzed, including growth with and without added chromium or uranium to study assimilation and dissimilatory reduction of these metals. The results will be presented in terms of the ranges and types of metals assimilated by DvH and *Pelosinus*

* Presenting author

A11 and the metals available in the organisms' natural environments, including Hanford groundwater.

Reference

1. Cvetkovic, A., Menon, A. L., Thorgersen, M., Scott, J. W., Poole, F. L., Jenney, F. E., Lancaster, W. A., Praismann, J. L., Shanmukh, S., Vaccaro, B., Trauger, S. A., Kalisiak, E., Apon, J. V., Siuzdak, G., Yannone, S. M., Tainer, J. A. and Adams, M. W. W. (2010) Microbial metalloproteomes are largely uncharacterized. Nature 466, 779-782

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.

216

ENIGMA Microbiology Physiology: Accurate, High-Throughput Identification of Stable Protein Complexes in *Desulfovibrio vulgaris*

John-Marc Chandonia[1,3]* (JMChandonia@lbl.gov), Maxim Shatsky,[1,3] Ming Dong,[1] Haichuan Liu,[2] Lee Yang,[1] Jil T. Geller,[1] Megan Choi,[1] Barbara Gold,[1] Nancy L. Liu,[1] Marjon Khairy,[1] Sonia Reveco,[1] Tom R. Juba,[5] Bonita R. Lam,[1] Evelin D. Szakal,[2] Simon Allen,[2] Sunil Kumar,[4] Farris L. Poole,[4] **Steven E. Brenner**,[1,3] **Steven C. Hall**,[2] **Susan J. Fisher**,[1,2] **Michael Adams**,[4] **Terry C. Hazen**,[6] **Judy D. Wall**,[5] **Swapnil Chhabra**,[1] **Jian Jin**,[1] **H. Ewa Witkowska**,[2] **Adam P. Arkin**,[1,3] **Gareth P. Butland**,[1] **Mark D. Biggin**,[1] and **Paul D. Adams**[1]

[1]Lawrence Berkeley National Laboratory; [2]University of California, San Francisco; [3]University of California, Berkeley; [4]University of Georgia; [5]University of Missouri, Columbia; and [6]Oak Ridge National Laboratory

http://enigma.lbl.gov

Project Goals: *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 component of our strategy is to develop and use high throughput pipelines to purify and characterize soluble protein complexes. We expect that these interaction data will improve our ability to produce accurate metabolic and regulatory models of key members of microbial communities.

The group led by Mark Biggin has developed a novel method for identification of stable, soluble protein complexes in microbes. In a small-scale pilot study, we showed that many protein complexes survive intact through a series of orthogonal chromatographic methods, with complex components having correlated elution profiles. These profiles were measured with the aid of mass spectrometry (MS) and iTRAQ reagents (Dong et al., 2008). We developed statistical and machine learning methods to analyze a full-scale data set, which were required in order to obtain biologically meaningful results due to the high potential for false positives (FP) caused by co-elution of proteins that are not part of a complex. Our methods were tuned using a manually curated gold standard (GS) set. As a first high-throughput study, we demonstrated this technique in identifying a high-precision subset of stable protein complexes in *Desulfovibrio vulgaris*. We have shown that our predicted network of interactions is significantly enriched in pairs with similar functional annotations. The quantitative information from elution profiles allowed us to develop a statistical model to estimate the false discovery rate in our predictions; because this varies according to how "crowded" the eluted fractions are, we are able to identify a subset of hundreds of highly reliable (i.e., with very low false discovery rate) interactions, as well as a much larger set of interactions that can be predicted with known false discovery rates. Advantages of the tagless approach include not requiring a mutant library (needed for alternative tag-based approaches such as TAP), and a false discovery rate comparable to TAP.

The group led by Gareth Butland has identified a number of protein complexes using TAP. We have developed an automated pipeline for synthesis of tagged gene constructs in collaboration with Swapnil Chhabra (Chhabra et al., 2011). To date, over 700 pulldowns (comprising more than 600 unique *D. vulgaris* strains) have been subject to analysis. In these experiments, more than 10,000 interactions were detected with over 1,000 distinct prey proteins. Using curated GS datasets (as in the tagless analysis), we filtered out ubiquitous proteins and other likely FP, resulting in a set of high-confidence interactions. A number of these interactions have been reciprocally confirmed, using strains in which the original prey protein was tagged and used as bait. Preliminary analysis of the data have identified several novel complexes, including multiple paralogous versions of the DnaJK-GrpE chaperone complex, each of which is bound to a small protein that may act as an allosteric regulator.

References

1. Chhabra SR, Butland GP, Elias D, Chandonia JM, Fok OY, Juba T, Gorur A, Allen S, Leung CM, Keller K, Reveco SA, Zane GM, Semkiw ES, Prathapam R, Gold B, Singer M, Ouellet M, Szakal ED, Jorgens D, Price MN, Witkowska HE, Beller HR, Arkin AP, Hazen TC, Biggin MD, Auer M, Wall JD, Keasling JD. 2011. Generalized Schemes for High Throughput Manipulation of the *Desulfovibrio vulgaris Hildenborough* Genome. *Applied and Environmental Microbiology* 77:7595-7604.
2. Dong M, Yang LL, Williams K, Fisher SJ, Hall SC, Biggin MD, Jin J, Witkowska HE. 2008. A "tagless" strategy for identification of stable protein complexes genome-wide by multidimensional orthogonal chromatographic separation and iTRAQ reagent tracking. *J Proteome Res* 7:1836-1849.

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.

* Presenting author

217

ENIGMA Microbiology Physiology: A Revised Bioenergetic Model of *Desulfovibrio vulgaris* Strain Hildenborough

Romy Chakraborty[1]* (RChakraborty@lbl.gov), Marcin Joachimiak,[1] Aifen Zhou,[2] Julian L. Fortney,[1] Zhili He,[2] Paramvir Dehal,[1] Morgan R. Price,[1] Jizhong Zhou,[2] Terry C. Hazen,[3] Adam P. Arkin,[1] Judy D. Wall,[4] Adam Deutschbauer,[1] **Swapnil R. Chhabra**,[1] and Paul Adams[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif., [2]University of Oklahoma, Norman; [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [4]University of Missouri, Columbia

http://enigma.lbl.gov

Project Goals: The ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) project seeks to elucidate the mechanisms and key processes that enable microorganisms and their communities to function in metal-contaminated soil sites. One goal is to understand the bioenergetic processes that enable ENIGMA-relevant microorganisms to thrive in such environments.

Sulfate reducing bacteria (SRB) play important roles in global sulfur and carbon cycling through their ability to completely mineralize organic matter while respiring sulfate to hydrogen sulfide. They are ubiquitous in anaerobic environments and have the ability to reduce toxic metals such as Cr(VI) and U(VI). While SRB have been studied for over three decades, bioenergetics of this clade of microbes is poorly understood. *Desulfovibrio vulgaris* strain Hildenborough (DvH) has served as a model SRB for the last several years with the accumulation of transcriptomic, proteomic and metabolic data under a wide variety of stressors. To further investigate the three proposed modes of energy generation in this anaerobe we conducted a systematic study involving multiple electron donor and acceptor combinations for growth. DvH was grown at 37°C in a defined medium with (a) 60mM Lactate + 30mM Thiosulfate, (b) 60mM Lactate + 40mM Sulfite (c) 60 mM lactate, 30 mM sulfate, (d) 120mM Pyruvate + 30mM Sulfate, (e) H_2 + 10mM Acetate + 30mM Sulfate, (f) formate 100mM + 10mM acetate + 25mM sulfate. g) no lactate, 5 mM acetate, 30 mM sulfate. Cells were harvested for transcriptomics at mid-log phase of growth for all conditions, when the optical density at 600nm was in the range 0.42-0.5. Initial results indicate that in spite of different electron donors, cells grown with sulfate, thiosulfate or sulfite as the electron acceptor show commonalities in gene expression and cluster together. Cells grown with H_2 and formate as electron donors as well as by pyruvate fermentation form a separate cluster consistent with differential pathway utilization. It was also observed that the ability of DvH to reduce metals like Cr(VI) when grown with lactate was greater than with formate as electron donor.

This work conducted by ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) 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.

218

Berkeley Synchrotron Infrared Structural Biology (BSISB) Program: A Bioanalytical Facility for Characterizing Chemical Changes in Living Cells

Hoi-Ying N. Holman[1]* (hyholman@lbl.gov), Michael C. Martin,[2] and Peter Nico[1]

[1]Ecology Department and [2]Advanced Light Source, Lawrence Berkeley National Laboratory

Project Goals: A DOE Structural Biology User Facility for accelerating biological research.

The Berkeley Synchrotron Infrared Structural Biology program (BSISB) is funded by the Department of Energy to provide support to researchers whose work can benefit from real-time measurements and imaging of chemical changes in living cells while biological processes are underway. The BSISB program has developed synchrotron radiation-based Fourier transform infrared (SR-FTIR) spectromicroscopy, which is ideally suited to measure at a chemicallevel how microbes adapt to changing environments. SR-FTIR spectromicroscopy is similar to the well-established non-invasive molecular analytical technique Fourier transform infrared (FTIR), except SR-FTIR is capable of discovering molecular information with micrometer spatial accuracy and signal-to-noise ratios 100 to 1000 times greater. Aqueous environments hinder SR-FTIR's sensitivity to bacterial activity, but the recent development of integrated in situ open-channel microfluidic culturing platforms circumvents the water-absorption barrier. These platforms enable real-time chemical imaging of bacterial activity in biofilms and facilitate comprehensive understanding of structural and functional dynamics in a wide range of microbial systems. Recently, the BSISB program has added visible (VIS) hyperspectral/fluorescence microscopy to its multi-modal imaging capability. Users can simultaneously track changes in cellular morphology, structure, and biological processes like gene expression and signaling during SIR experiments. The presentation will include examples that demonstrate the use of BSISB in wide ranging research projects that are important to DOE missions.

219

Structural and Functional Properties of Transporter Solute Binding Proteins

Sarah Zerbs[1]* (szerbs@anl.gov), Lisa M. Miller,[3] Ruth Pietri,[3] Jamey C. Mack,[1,2] Andrew Binkowski,[1,2] Changsoo Chang,[1,2] Marianne Cuff,[1,2] Yao Fan,[1,2] Karolina Michalska,[1,2] Jureck Osipiuk,[1,2] Kemin Tan,[1,2] Andrzej Joachimiak,[1,2] and **Frank Collart**[1]

[1]Biosciences Division, Argonne National Laboratory, Lemont, Ill.; [2]Midwest Center for Structural Genomics, Argonne National Laboratory, Lemont, Ill.; and [3]National Synchrotron Light Source, Brookhaven National Laboratory, Upton, N.Y.

http://www.bio.anl.gov/molecular_and_systems_biology/proteins.html

Project Goals: The Argonne "Environment Sensing and Response" Scientific Focus Area (ESR-SFA) program addresses the hypothesis that cellular behavior can be modeled through an understanding of the biological interface with the environment and the cellular responses that originate from the cell/environment interaction. The overall objective of the ESR-SFA program is to identify the molecular basis of cellular transport and sensory pathways that mediate the response to environmental nutrients. We recently identified four possible solute binding proteins (SBPs) in *R. palustris* that demonstrated binding to aromatic lignin monomers. We used molecular crystallography to examine the molecular basis for the aromatic compound binding specificity of these proteins and assess the feasibility for the development of sequenced based prediction methods to determine ligand binding specificity.

Rhodopseudomonas palustris metabolizes aromatic compounds derived from lignin-degradation products and has the potential for bioremediation of xenobiotic compounds. In this study, we exploit the ligand binding specificity of solute binding proteins (SBPs) to evaluate the transport specificity of ABC transporters for lignin degradation products for multiple transporters. We recently identified four possible SBPs in *R. palustris* CGA009 that demonstrated binding to aromatic ligands similar to lignin monomers. To expand on these initial results, we selected homologues from *R. palustris* strains HaA2 and BisB5, *B. japonicum* USDA110 and *S. meliloti* 1021 and subjected targets to a directed screen against potential aromatic ligands. A fluorescent thermal shift (FTS) assay was used to screen for ligands of purified SBPs using a library of 46 compounds. Individual proteins exhibited binding to multiple structurally similar molecules, but also demonstrated clear ligand preferences arising from carbon chain length and ring modifications. Arrangement of FTS data by reference SBP group and common chemical features of the ligands enables organization of the binding capabilities into distinct groups of protein-ligand interactions. Overlay of the chemical specificity on an unrooted phylogenetic tree suggests a consistent relationship of the binding profile inferred by the FTS assay with the phylogenetic relatedness.

To establish a molecular basis for the aromatic compound binding specificity of these proteins, 13 proteins were produced in a bacterial expression system, purified and routed to crystallization screens in the presence and absence of ligands. Each protein was screened with either no ligand (apo) or one of two preferred ligands (ligand1, ligand 2) indicated by induction of the highest melting temperature stabilization in FTS assays. These 39 combinations yielded multiple crystals, 9 of which gave refraction patterns suitable for solving structures. Solved structure resolutions are generally less than 2 Å and all structures indicate SBPs in a closed formation with a ligand in the binding pocket. Most targets were solved as monomers, however two targets exhibited dimers in the space group. In cases where an FTS-identified ligand was found in the active site, structure resolution was high enough to calculate hydrogen bond coordination in the residues that comprise the binding pocket. Based on these structures, ligand coordination differs between the two groups described by phylogenetic and *in vitro* functional analysis of these targets.

The crystal structures of these novel aromatic binding proteins from *R. palustris* confirmed the inferred functional assignment as binding proteins for aromatic compounds. The variances in binding pocket geometry and thermodynamic parameters for these SBPs are consistent with the observed sequence variance and sequence phylogeny for this protein family. This information will enable an understanding of the molecular mechanisms that underlie microbial interactions and thereby facilitate the development of system-level models predictive of cellular response to environmental conditions or changes.

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).

220

KBase: An Integrated Knowledgebase for Predictive Biology and Environmental Research

Adam Arkin[1]* (aparkin@lbl.gov), Robert Cottingham,[2] Sergei Maslov,[3] Rick Stevens,[4] Dylan Chivian,[1] Paramvir Dehal,[1] Christopher Henry,[4] Folker Meyer,[4] Jennifer Salazar,[4] Doreen Ware,[5] David Weston,[2] Brian Davison,[2] and Elizabeth M. Glass[4]

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]Brookhaven National Laboratory, Upton, N.Y.; [4]Argonne National Laboratory, Argonne, Ill.; and [5]Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

http://kbase.us/
http://outreach.kbase.us

Project Goals: The Systems Biology Knowledgebase (KBase) has two central goals. The scientific goal is to produce predictive models, reference datasets and analytical tools and demonstrate their utility in DOE biological research relating to bioenergy, carbon cycle, and the study of subsurface microbial communities. The operational goal is to create the integrated software and hardware infrastructure needed to support the creation, maintenance and use of predictive models and methods in the study of microbes, microbial communities and plants.

KBase is a collaborative effort designed to accelerate our understanding of microbes, microbial communities, and plants. It will be a community-driven, extensible and scalable open-source software framework and application system. KBase will offer free and open access to data models and simulations, enabling scientists and researchers to build new knowledge, test hypotheses, design experiments, and share their findings to accelerate the use of predictive biology. Our immediate 18-month goal is to have a beta-version completed by February 2013.

The KBase microbial science domain will enable the reconciliation of metabolic models with experimental data with the ultimate aim of manipulating microbial function for applications in energy production and remediation. In order to accomplish this, we will enable users to expand on a strong foundation of quality genome annotations, reconstruct metabolism and regulation, integrate and standardize ‘omics data, and construct models of genomes. In doing so we will vastly improve the planning of effective experiments, maximize understanding of microbial system function, and promote sharing of data and findings.

The plants science domain will initally target linking genetic variation, phenotypes, molecular profiles, and molecular networks, enabling model-driven phenotype predictions. We will also map plant variability onto metabolic models to create model-driven predictions of phenotypic traits. Initial work will focus on creating a workflow for rapidly converting sequencing reads into genotypes. We will also build tools for data exploration, and the linking of gene targets from phenotype studies such as genome-wide association studies, with co-expression, protein-protein interaction, and regulatory network models. Such data exploration will allow users to narrow candidate gene lists by refining targets, or be able to visualize a sub-network of regulatory and physical interactions among genes responsible for a phenotype. Users can also highlight networks or pathways impacted by genetic variation.

Through comparative analysis of metagenomes acquired over different spatial, temporal or experimental scales, it is now possible to define how communities respond to and change their environment. Our microbial communities team will build the computational infrastructure to research community behavior and build predictive models of community roles in the carbon cycle, other biogeochemical cycles, bioremediation, energy production, and the discovery of useful enzymes. We are building the next-generation metagenomic platform that provides scalable, flexible analyses, data vectors for models, tools for model creation, data quality control, application programing interfaces, and GSC-compliant standards for data collection. Initial efforts will target the development of bioprospecting and experimental design tools.

KBase will be composed of a series of core biological analysis and modeling functions, including an application programming interface that can be used to connect different software programs within the community. These capabilities will be constructed from the popular analysis systems at each of the KBase sites. Their integration into KBase will combine individual functions to create the next generation of biological models and analysis tools. The KBase application programming interface will also enable third-party researchers from our diverse community of users to design new functions. KBase will be supported by a computing infrastructure based on the OpenStack cloud system software, distributed across the core sites.

The success of the KBase project depends not only on producing a large-scale, open computational capability for systems biology research data management and analysis but also on positioning these tools to be used by the community. Our outreach program is designed to target different user groups: data providers, tool builders, and users of both data and tools. A significant effort will be made to connect the user groups and broader systems biology science communities to the KBase resources and efforts. We will provide

* Presenting author

educational support, including providing access to outreach and technical staff and online venues in which to express questions, suggestions and needs to other users and our entire team. In addition to this, we will provide transparency to KBase, providing information about the project, team, and development with the scientific community.

New functionality will allow users to visualize data, create powerful models or design experiments based on KBase-generated suggestions. Although the integration of different data types will itself be a major offering to users, the project is about much more than data unification. KBase is distinguished from a database or existing biological tools by its focus on interpreting missing information necessary for predictive modeling, on aiding experimental design to test model-based hypotheses, and by delivering quality-controlled data.

This work is supported by the U.S. Department of Energy, Office of Biological and Environmental Research under Contract DE-AC02-06CH11357.

221

The DOE Systems Biology Knowledgebase: Microbial Communities Science Domain

Folker Meyer[1]* (folker@anl.gov), Dylan Chivian,[2] Andreas Wilke,[1] Narayan Desai,[1] Jared Wilkening,[1] Kevin Keegan,[1] William Trimble,[1] Keith Keller,[2] Paramvir Dehal,[2] Robert Cottingham,[3] Sergei Maslov,[4] Rick Stevens,[1] and **Adam Arkin**[2]

[1]Argonne National Laboratory, Argonne, Ill.; [2]Lawrence Berkeley National Laboratory, Berkely, Calif.; [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [4]Brookhaven National Laboratory, Upton, N.Y.

http://kbase.us/

Project Goals: The Systems Biology Knowledgebase (KBase) has two central goals. The scientific goal is to produce predictive models, reference datasets and analytical tools and demonstrate their utility in DOE biological research relating to bioenergy, carbon cycle, and the study of subsurface microbial communities. The operational goal is to create the integrated software and hardware infrastructure needed to support the creation, maintenance and use of predictive models and methods in the study of microbes, microbial communities and plants. The microbial communities component will be focused on building the computational infrastructure to understand the community function and ecology through study of genomic and functional data and integration of community models with single-organism models. This will allow for researching community behavior and building predictive models of communities in their role in the environmental processes and the discovery of useful enzymes. The microbial communities infrastructure will support the overall KBase goal to provide a framework for experimental decision support and data interpretation.

The KBase microbial communities team will integrate both existing and new tools and data into a single unified framework that is accessible programmatically and through web services. The framework will allow the construction of sophisticated analysis workflows by facilitating the linkages between data and analysis methods. The standardization, integration and harmonization of diverse data types housed within the KBase and data located on servers maintained by the larger scientific community will allow for a single point of access ensuring consistency and quality-assurance/quality-control checks of data quality.

We have begun by creating KBase data and analysis services that will link our core resources: MG-RAST [1], metaMicrobesOnline [2], SEED [3], IMG/M [5] and ModelSEED [4]. These services will allow clients to access data and analysis methods across these tools without the burden of reconciling identifiers, learning different data access and programmatic access methods, ensuring data quality, and maintaining relevant metadata. New functionality, not currently available in our core tools, is being created within KBase using the programmatic interfaces.

Protoypical applications

Bioprospecting: Microbial diversity is a key element in the search for new, valuable compounds such as enzymes with novel properties. Elucidating novel proteins from microbial communities is a function that integrates metaMicrobesOnline functions with MG-RAST data through the KBase programmatic interface. It will allow for deep comparative analysis of protein families, expanding significantly the current functionality in MG-RAST. This includes detailed trees and alignments combining metagenomic sequences and sequences from complete genomes. The initial version of this tool will allow in-depth characterization of novel members of existing protein families; future versions will allow characterization of completely novel protein families. This will exercise the communities part of the API and also the microbes set of API calls and provide a useful, missing component to the combined tool suite.

Metagenomic Experimental Design Wizard: The "wizard" will assist in the design of metagenomic experiments. Using information on sequence data (including sequence quality) and the detailed community analysis, it will provide guidance on the analyses/experiments that are supported by the data. The wizard provides guidance on, for example, identifying microbial communities using sequence assembly, and providing a confidence value for community reconstructions obtained from metagenomic data.

References

1. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, et al. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics. 2008 Sep 19;9:386.
2. Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, et al. MicrobesOnline: an integrated portal for comparative and functional genomics. Nucleic Acids Res. 2010 Jan;38(Database issue):D396-400.

* Presenting author

3. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. 2005 Oct 7;33(17):5691-702.
4. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol. 2010 Sep;28(9):977-82.
5. Victor M. Markowitz, I-Min A. Chen, Ken Chu, Ernest Szeto, Krishna Palaniappan, Yuri Grechkin, et al. IMG/M: the integrated metagenome data management and comparative analysis system, Nucleic Acid Res. 2011.

This work is supported by the U.S. Department of Energy, Office of Biological and Environmental Research under Contract DE-AC02-06CH11357.

222

The DOE Systems Biology Knowledgebase: Microbial Science Domain

Paramvir S. Dehal[1]* (psdehal@lbl.gov), Chris S. Henry,[2] Ben Bowen,[1] Steven Brenner,[1] Ross Overbeek,[2] John-Marc Chandonia,[1] Dylan Chivian,[1] Pavel S. Novichkov,[1] Keith Keller,[1] **Adam P. Arkin**,[2] Robert Cottingham,[3] Sergei Maslov,[4] and Rick Stevens[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]Argonne National Laboratory, Argonne, Ill.; [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [4]Brookhaven National Laboratory, Upton, N.Y.

http://kbase.us

Project Goals: The Systems Biology Knowledgebase (KBase) has two central goals. The scientific goal is to produce predictive models, reference datasets and analytical tools and demonstrate their utility in DOE biological research relating to bioenergy, carbon cycle, and the study of subsurface microbial communities. The operational goal is to create the integrated software and hardware infrastructure needed to support the creation, maintenance and use of predictive models and methods in the study of microbes, microbial communities and plants. The microbes component of the KBase will be centered on an analysis pipeline that will include annotation of genome sequences, reconstruction of metabolic pathways and regulons, generation of metabolic and regulatory models, and reconciliation of models with existing 'omics datasets and new datasets uploaded by a user. KBase will provide access to this database and analysis pipeline via a powerful programmatic interface and an intuitive socially-enabled web interface.

The microbes component of the KBase project aims to unify existing 'omics datasets and modeling toolsets within a single integrated framework that will enable users to move seamlessly from the genome annotation process through to a reconciled metabolic and regulatory model that is linked to all existing experimental data for a particular organism. More importantly, we will embody tools for applying these models and datasets to drive the advancement of biological understanding and microbial engineering.

In order to drive the development of the microbes area and enable new science, we will focus on accomplishing prototype science workflows rather than general tasks. Work will be bootstrapped by leveraging data sets and tools developed and maintained by the MicrobesOnline, SEED, RegPrecise and ModelSEED resources. The initial microbes efforts will integrate prototype workflows for: (1) genome annotation and metabolic reconstruction, (2) regulon reconstruction, (3) metabolic and regulatory model reconstruction, and (4) reconciliation with experimental phenotype and expression data.

1. Evidence Based Genome Annotation and Metabolic Reconstruction: While the rate of genome sequencing continues to advance at an exponential pace, our ability to confidently assign structural and functional gene annotations has not kept pace. High quality gene annotations with confidence measures are a critical component of all genome scale modeling. Efforts to create genome scale regulatory and metabolic models are held back by the poor quality of existing gene models. To help resolve this, we are proposing a workflow that takes as input a genome sequence, RNASeq and/or high density tiling array data, and functional 'omics datasets. Through an iterative process combining the experimental data sets and comparative genomics, structural annotations will be improved and integrated into the RAST annotation server. High quality transcription start and operon predictions will be used to improve promoter and regulatory predictions. Gene functional annotations will be improved by combining model predictions to identify missing gene functions.

2. Regulon Reconstruction: Given accurate gene models, the KBase framework will provide integrated pipelines for building and refinement of higher level, regulatory and metabolic models. Reconstruction of genome-wide transcriptional regulatory network (TRN) is a necessary step toward the ultimate goal, building a ***predictive*** model of microbial organism.

3. Seamless Integration of Modeling into the Biological Research Process: Computational modeling of biological systems provides a mechanism for rapidly exploring a wide variety of alternative theories of complex phenomena observed in the wetlab. Yet modeling is restrained by a lack of interoperability of models and data, a high degree of mathematical and computational expertise required to utilize models, and an inability to rapidly build new models accurately capture the complete body of our current biological understanding. The KBase microbes team aims to integrate reconstruction and analysis algorithms for metabolic models, regulatory models, and ultimately many other modeling abstractions with a unified compilation of 'omics data to produce a framework that facilitates the process of building biological understanding. Algorithms will be included for simulating microbial behavior in a specified environment, predicting and modifying pathways according to data or design, and designing experiments to test biological theories.

4. Testing and Improving Consistency of Biological Understanding and Experimental Data: The microbes Kbase area will combine ‘omics data and modeling algorithms within a single platform, making it possible to easily cross-validate data and models by comparing predictions with experimental observations. Additionally, tools and interfaces will be provided to guide the process of adapting the biological understanding that underlies our models to reconcile conflicts with experimental data. The results of this adaptation process will automatically feed back into the annotation algorithms built into KBase, globally improving annotations of all genomes. Initially this effort will focus on the interpretation and reconciliation of growth phenotype and gene expression data, but will ultimately be expanded to all types of ‘omics data.

This work is supported by the U.S. Department of Energy, Office of Biological and Environmental Research under Contract DE-AC02-06CH11357.

223

The DOE Systems Biology Knowledgebase: Plant Science Domain

Doreen Ware[1,2]* (ware@cshl.edu), Sergei Maslov,[4] Shinjae Yoo,[4] Dantong Yu,[4] Michael Schatz,[1] James Gurtowski,[1] Matt Titmus,[1] Jer-ming Chia,[1] Sunita Kumari,[1] Andrew Olson,[1] Shiran Pasternak,[1] Jim Thomason,[1] Ken Youens-Clark,[1] Mark Gerstein,[5] Gang Fang,[5] Darryl Reeves,[5] Pam Ronald,[6] Chris Henry,[7] Sam Seaver,[7] David Weston,[3] Priya Ranjan,[3] Robert Cottingham,[3] Rick Stevens,[7] and **Adam Arkin**[8]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.; [2]United States Department of Agriculture–Agriculture Research Service; [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [4]Brookhaven National Laboratory, Upton, N.Y.; [5]Yale University, New Haven, Conn.; [6]University of California, Davis; [7]Argonne National Laboratory, Argonne, Ill.; and [8]Lawrence Berkeley National Laboratory, Berkely, Calif.

http://kbase.us

Project Goals: The Systems Biology Knowledgebase (KBase) has two central goals. The scientific goal is to produce predictive models, reference datasets, and analytical tools and to demonstrate their utility in DOE biological research relating to bioenergy, carbon cycle, and the study of subsurface microbial communities. The operational goal is to create the integrated software and hardware infrastructure needed to support the creation, maintenance, and use of predictive models and methods in the study of microbes, microbial communities and plants. The plant component of the KBase will allow users to model genotype-to-phenotype relationships using metabolic and functional networks as well as phenotype measurements and ‘omic data. It will also support the reconstruction of new metabolic and functional networks based on expression profiles, protein-DNA, and protein-protein interactions. To accomplish this, we will provide interactive, data-driven analysis and exploration across multiple experiments and diverse data-types. We will provide users access to comprehensive collections of ‘omics datasets together with relevant analytical tools and resources.

The major goal for the KBase plants area is to model genotype-to-phenotype relationships through analysis and integration of genomic, transcriptomic, methylomic sequencing data, metabolite and phenotype measurements, and the reconstruction of metabolic and functional networks based on expression profiles, protein-DNA, and protein-protein interactions. To accomplish this goal, KBase will provide interactive, data-driven analysis, and exploration across multiple experiments and diverse data-types. We will provide users access to comprehensive datasets from high-throughput experiments together with relevant analytical tools and resources. Users will be provided with a platform to analyze their own experimental data, integrate publicly available data from other ‘omics’ platforms, and have these results incorporated into a data exploration framework. We will aid in the translation of basic science through exploration across large ‘omics experiments and the initiation of hypothesis-driven genetics studies without the overhead of data flow and management.

The KBase plants effort will consist of two major components, 1) genotyping workflows and 2) data exploration and prediction tools.

Genotyping workflows: Exponential growth in digital demands has motivated extensive research into improved algorithms and parallel systems, especially for genotyping samples, monitoring expression levels, and a host of other important biological applications.

Genotyping workflows will leverage our recent development of Jnomics, as our new Hadoop-based open-source package for rapid development and deployment of cloud-scale sequence analysis tools. Jnomics provides many pre-built tools out-of-the-box that accelerate common tasks, such as mapping, sorting, merging, filtering, and selection, to be performed as distributed tasks spread across a cluster. New tools can be easily created using an open-source Java API, especially for large-scale genotyping and expression analysis. Because it builds on Hadoop, Jnomics tools inherit Hadoop’s efficiency and scalability for very large datasets such as the billions of short reads necessary for genotyping many large plant genomes. Furthermore, Jnomics is “file format agnostic”, allowing it to seamlessly read and write most common sequence file formats (SAM, BAM, BED, fastq, fasta), making it easy for Jnomics to interface with other components.

Data exploration and prediction: While the benefits of integrating large independent data sets for value-added research are vast, the influx of raw biological data presents unprecedented challenges in data management and representation. We believe that effective visualization will play a key role in mining and exploration of ‘omics data sets. While genome browsers are useful for anecdotal relationships between feature sets and genome maps—especially for bench scientists studying a region of interest—they are not ideal for correlational interrogation of the data. Effective visualization is crucial for generating hypotheses. Displays must be flexible insofar as they allow the end-user to represent highly dimensional data in various forms without compromising pairwise relationships among data points. We intend to develop aesthetic and interactive interfaces that enable researchers to intuitively draw relationships in their complex queries. The visualizations will allow end-users to transform massive data sets in real-time in order to explore latent system-wide relationships among the data. Data exploration tools will focus on integration and visualization of data from multiple ‘omics studies, genetic variation, network and pathways models as well as phenotypic association. Users will, for example, map variability onto metabolic models, highlighting pathways predicted to be impacted by genetic variation. By overlaying genetic variation within the context of gene network models, we can predict pathways that are impacted by the pool of genetic diversity in the population.

This work is supported by the U.S. Department of Energy, Office of Biological and Environmental Research under Contract DE-AC02-06CH11357.

224

The KBase Architecture and Infrastructure Design

Tom Brettin*[1] (brettints@ornl.gov), Bob Olson,[2] Ross Overbeek,[2] Terry Disz,[2] Bruce Parello,[2] Shiran Pasternak,[5] Folker Meyer,[2] Michael Galloway,[1] Steve Moulton,[1] Dan Olson,[2] Shane Canon,[3] Dantung Yu,[4] Shiran Paternak, Pavel Novichkov,[3] Daniel Quest,[1] Narayan Desai,[2] Jared Wilkening,[2] Miriam Land,[1] Scott Deviod,[2] **Adam Arkin**,[3] Robert Cottingham,[1] Sergei Maslov,[4] and Rick Stevens[2]

[1]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Argonne National Laboratory, Argonne, Ill.; [3]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [4]Brookhaven National Laboratory, Upton, N.Y.; and [5]Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

http://kbase.us/

Project Goals: The Systems Biology Knowledgebase (KBase) has two central goals. The scientific goal is to produce predictive models, reference datasets and analytical tools and demonstrate their utility in DOE biological research relating to bioenergy, carbon cycle, and the study of subsurface microbial communities. The operational goal is to create the integrated software and hardware infrastructure needed to support the creation, maintenance and use of predictive models and methods in the study of microbes, microbial communities and plants. The driving objectives of the KBase architecture and infrastructure design focus on creating an unprecedented user experience. The integrated software and hardware infrastructure supporting the user experience comprises a continuously expanding collection of software and services. These are hosted on a physical infrastructure consisting of high speed wide area networking, cloud computing resources, and state of the art cluster computing resources.

Achieving our goal of an integrative architecture to support predictive modeling requires enabling a user experience that covers a range of users. These include senior biologists determined to understand and create biological models, computational biologists developing new algorithms and statistical models that form a basis for the biological models, and bioinformaticists chaining together complex workflows to generate, summarize and integrate data that feed into the biological models. These user activities are enabled at various levels of abstraction, including a) knowledge creation, reproduction and sharing; b) rich web applications; c) programmatic Application Programming Interface (API) libraries and scripts; and d) wire level communication access.

The development of the KBase Unified API is based on a service-oriented approach to deliver both functionality and data to the community. Behind this lies a set of services backed by servers. Initially, these services will be developed by the KBase infrastructure team and will support a long term goal of community developed and contributed services. Our initial set of services will be backed by the following servers:

1. **Genomic Servers** that provide access to a rapidly growing set of genomes, features of those genomes, and annotations of both genomes and features.
2. **Expression Data Servers** creating access to a growing body of expression data, along with the underlying encoding of metadata needed to support interpretation.
3. **Protein Family Servers** supporting access to a variety of the existing collections of protein families.
4. **Polymorphism Servers** capturing various genetic polymorphisms such as single nucleotide polymorphisms, tandem repeats, and copy number variations.
5. **Phenotype Servers** enabling the relationships between genotype and phenotype to be understood.
6. **Compound and Reaction Data Servers** supporting a unified and maintained representation of reaction networks.
7. **Metabolic Modeling Servers** that support the construction and maintenance of metabolic models.
8. **Regulatory Models Servers** that support the construction and maintenance of regulatory models.

Our KBase physical infrastructure builds on the successes of DOE investment in our national scientific cyber infrastructure and can therefore leverage enormous intellectual resources present in the DOE community.

Building on ESNet allows us to construct a wide area network between the partner labs that enables a virtual hardware infrastructure. In the first quarter of the project we have established 10Gbit data transfer connectivity between KBase data transfer nodes.

Enabling cloud computing on Magellan will create new opportunities that range from rapid deployment of developer environments to highly scalable production servers. The acceptance of virtualization technology is growing, and the use of machine images produced by others is already visible in our core services. In the near-term, we are establishing the infrastructure to run existing images, both community based and internally created, on which the infrastructure is dependent. Examples include images supporting microbial community models and plant genome wide association studies. For the mid-term, we plan to contribute machine images to the community by creating snapshots of parts of our environment so that others can use them on the hardware of choice. For the long term objective, the ability to host running machine images on KBase hardware is a means for promoting collaboration and community support.

Cluster Computing has long been a critical part of biological data analysis. In collaboration with computing centers created by the Office of Advanced Computing Research such as NERSC, our underlying cluster services can leverage these resources and scale to meet needs.

KBase aims to power the next wave of biological research in DOE and beyond. Enabling these capabilities requires a software and hardware infrastructure that is integrated, extensible, and scalable. The architecture is designed to meet these needs and support user functionality to visualize data, create models or design experiments based on KBase-generated suggestions.

This work is supported by the U.S. Department of Energy, Office of Biological and Environmental Research under Contract DE-AC02-06CH11357.

225

The InterBRC Knowledgebase Data Registry

Dylan Chivian[1,2]* (DCChivian@lbl.gov), Guruprasad Kora,[3,4] Mustafa Syed,[3,4] Thomas Brettin,[4] Keith Keller,[2] Jason Baumohl,[2] Yury Bukhman,[5,6] Richard LeDuc,[5,6] Adam Arkin,[1,2] David Benton,[5,6] **Steve Slater**,[5,6] and **Edward Uberbacher**[3,4]

[1]Joint BioEnergy Institute, Emeryville, Calif.; [2]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [3]Bioenergy Science Center, Oak Ridge, Tenn.; [4]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [5]Great Lakes Bioenergy Research Center, Madison, Wis.; and [6]University of Wisconsin, Madison

Project Goals: The InterBRC Knowledgebase project will provide a mechanism for the integration of data obtained by the DOE Bioenergy Research Centers (BRCs) with the DOE Systems Biology Knowledgebase (KBase).

The DOE Systems Biology Knowledgebase (KBase) is building a system that allows predictive modeling of Microbes, Microbial Communities, and Plant Systems Biology for the scientific community. The DOE Bioenergy Research Centers (BRCs) are large producers of systems biology data in all three areas, generating genome sequence, expression, proteomic, metabolomic, metabolic flux, growth, and phenotype data for microbes; metagenome, metatranscriptome, and metaproteome data for microbial communities; and genome, protein interaction, protein localization, allelic variation, and mutant phenotype data for plants. These data are rich sources for modeling via the KBase, but are currently either not easily accessible, or housed in a wide range of data repositories and thus challenging to bring together for comparative analysis. The InterBRC Knowledgebase will serve as the bridge between BRC data stored in dedicated systems and the common infrastructure of the KBase. Data will be indexed and searchable via the InterBRC Knowledgebase Data Registry, which will additionally provide the location and access protocol for retrieving data sets of interest from the dedicated services. These will be incorporated in turn into the KBase infrastructure and be available for comparative analysis and systems biology modeling by BRC researchers and the greater community.

This work is part of the InterBRC Knowledgebase project, the Joint BioEnergy Institute (JBEI), the Great Lakes Bioenergy Research Center (GLBRC), the Bioenergy Science Center (BESC), and the Systems Biology Knowledgebase project supported by the U.S. Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between the U.S. DOE and Lawrence Berkeley National Laboratory (LBNL), through contract DE-FC02-07ER64494 between the U.S. DOE and the Great Lakes Bioenergy Research Center (GLBRC) Cooperative Agreement, through contract DE-AC05-00OR22725 between the U.S. DOE and Oak Ridge National Laboratory (ORNL) administered by UT-Battelle, LLC, for the U.S. DOE, and through subcontract 4000105297 between ORNL and the GLBRC Cooperative Agreement.

226

GGKbase, a Portable Knowledgebase for Analysis and Integration of "Omic" Data From Microbial Communities

Brian C. Thomas,[1]* Ken-ichi Ueda,[1] Rahul Basu,[2] Chongle Pan,[3] Trent Northen,[3] Benjamin Bowen,[2]* and **Jill Banfield**[1]* (jbanfield@berkeley.edu)

[1]University of California, Berkeley; [2]Lawrence Berkeley National Laboratory; and [3]Oak Ridge National Laboratory

http://genegrabber.berkeley.edu/

Project Goals: To develop a knowledgebase for analysis and integration of 'omics' data from microbial communities.

Cultivation-independent approaches provide access to the wide diversity of microorganisms in natural environments. Sequence data (metagenomic information) is foundational to most studies of natural microbial communities as it enables functional analysis through proteomics (proteogenomics) and provides context for transcriptomic and metabolomics information.

Given the vast dataset sizes, extraction of biological and biogeochemical insight from 'omic' datasets is challenging. In the current project, we have constructed a workflow and knowledgebase that enables recovery, efficient and effective display, manipulation, and interrogation of such information. Most important, the structure is portable. Developed initially for analysis of data from acid mine drainage microbial communities, the structure has already been populated by information from multiple other projects, including the DOE Rifle IFRC metagenomics and proteomics efforts.

The pipeline includes all components required for analysis of next-generation sequencing information, from assembly through binning and functional annotation. An explicit goal of GGKbase is the recovery of near-complete genomes, a feature that distinguishes our approach from others (e.g., MGRAST). Curated and binned genome fragments are grouped into organismal "bins" from which inferences about metabolic potential can be made. Genes for which proteins have been identified from one or a series of samples using the open reading frames predicted from the metagenomic data are flagged at the "genome browsing" level, and detailed information about abundance and distribution can be accessed, gene-by-gene.

We have developed GeneGrabber (a component of GGKbase), a multi-user, list-based, social/sharing approach for analysis of the metabolism of individual organisms and comparative metabolic analysis at the community level. Individual genes or groups of genes belonging to a pathway can be assigned to one of more lists, as determined by the investigator, and these lists can be shared with other users, including the ability to invite new users to participate in curating a list. Because the lists are driven by a keyword search (or EC number, GO term etc.), genes can be identified and classified simultaneously across the entire dataset. This establishes metabolic profiles using tens, hundreds, and potentially thousands of genes at a time.

Understanding an ecosystem's metabolic potential is a complex task. Leveraging the extensive, content-based lists created for each metagenomics resource, we developed a tool within GeneGrabber for visualizing the extent of metabolic machinery present in the data. This visualization, termed "genome summary," is invaluable for exploring metabolic pathways and can identify which organisms in the community are responsible for a process. The genome summary is a useful tool for investigating the molecular underpinnings of ecosystem metabolic processes.

The GGKbase system has been engineered to access other 'omics' data resources, without having to resort to database federation. GGKbase uses representational state transfer (REST) to tap into other information sources and we have developed a caching system using Redis to accelerate user-centric data access. We have also developed an API to access the GGKbase resource both in Ruby as well as just using simple URL access. Currently, GGKbase includes both metagenomic and proteomic data. Additionally, we have expanded the GGKbase structure to now include metabolomics data and have developed a system for processing and displaying this data called MetaboliteAtlas. MetaboliteAtlas, like all components in the GGKbase utilizes a RESTful architecture and includes a separate front-end web display for in-depth metabolomics investigations.

GGKbase is under continual development, currently focused around addition of metabolomic and transcriptomic data. As more researchers begin using GGKbase, new ideas are captured into the structure. Current challenges in increasing the scale of the community metagenomics approach include automation of time-intensive binning steps and aspects of time series data analysis.

227

The Ribosomal Database Project: Tools and Sequences for rRNA Analysis

Benli Chai[1]* (chaibenl@msu.edu), Qiong Wang,[1] Jordan Fish,[1] Donna McGarrell,[1] C. Titus Brown,[2] Yanni Sun,[2] James M. Tiedje,[1] and **James R. Cole**[1]

[1]Center for Microbial Ecology, and [2]Computer Science and Engineering, Michigan State University, East Lansing

http://rdp.cme.msu.edu/
http://fungene.cme.msu.edu/

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.

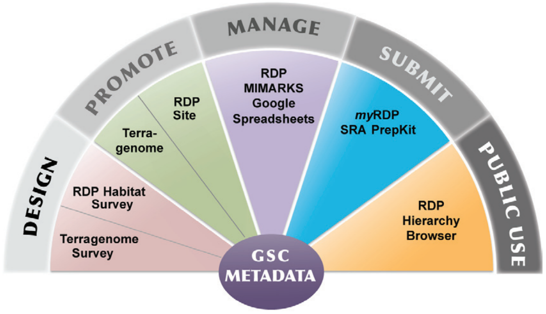
Our view of the evolutionary relationships among life forms on Earth has been revolutionized by the comparative analysis of ribosomal RNA sequences. Life is now viewed as belonging to one of three primary lines of evolutionary descent: Archaea, Bacteria and Eucarya. This shift in paradigm has not only 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. 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 (August 2011), RDP offers 1,921,179 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, 11,892 researchers have over 7.4 million private pre-publication sequences in the *my*RDP account system. On average, over the past year, the RDP website was visited by over 10,100 researchers (unique ip addresses) in over 24,500 analysis sessions each month, an increase of 20% from the previous year, 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, during the past 12 months the RDP Pyro Pipeline has been used by over 446 researchers (unique e-mail addresses) to process their own high-throughput current-generation rRNA sequence data.

The popular RDP naïve Bayesian Classifier has been cited in over 380 publications in over 100 journals since its publication in 2007. The open-source command line RDP Classifier package is freely available from SourceForge. This package has been downloaded 2,745 times during the past 12 months and a total of 6,025 times since first offered in 2006. A new command line tool developed to meet the growing need for taxonomy-based analyses of large numbers of sequences in multiple samples, MultiClassifier, is also freely available from the RDP site.

We are now offering a fungal version of our RDP Classifier, in collaboration with Andrea Porras-Alfaro, Gary Xie, Cheryl Kuske and their co-workers (supported through a DOE Science Focus Area grant to Los Alamos National Laboratory). This Fungal Classifier is trained on 8,506 curated fungal 28S rRNA gene reference sequences, along with a hand-vetted fungal taxonomy including 1,702 genera plus higher-level taxa. This is an important addition to the RDP tools, as fungi play a major role in carbon cycling and plant health. We will continue to work with our LANL colleagues to improve and extend our set of fungal tools.

We are 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. Our work was incorporated into the new MIMARKS (Minimal Information about a MARKer gene Sequence) standard (Yilmaz *et al.,* 2011). RDP's Hierarchy Browser has been updated to allow searching on MIMARKS attributes. To help our user community comply with these new community standards, RDP has added informative web pages to make our users aware of the GSC standards and developed tools to assist our users to collect and prepare compliant metadata (contextual data). RDP has created specialized MIMARKS templates using Google Docs office suite. These Google Spreadsheet templates for all 14 MIMARKS environmental packages provide embedded help and validation for MIMARKS attributes. By leveraging Google Docs, our users have a familiar tool that provides remote collaborative support, data storage, and a powerful user interface, all without need of local IT infrastructure. When researchers are ready to submit data, they can use the RDP's *my*RDP SRA PrepKit, which helps researchers prepare richly annotated sequence data in the complex XML documents required for submission to the NCBI and ENA SRA repositories. RDP GoogleSheets also contain macros that produce data formats compatible with standard ENA and NCBI sequence submission tools.



Beyond rRNA, RDP is leveraging its tools and services to provide support for analysis of coding for key environmental functions (functional genes). Like rRNA genes, protein-coding genes can also be selectively targeted for deep sequencing coverage. Genes that are important for environmental processes can thus be used to identify functional, ecological and evolutionary patterns. RDP's FunGene repository provides support for amplicon design and analysis for genes involved in a range of processes. Integrated tools are provided for primer testing, phylogenetic analysis and model refinement. FGPipeline (http://fungene.cme.msu.edu/FunGenePipeline/) has been recently developed to analyze this type of sequencing data. The major tools underpinning this pipeline include RDP FrameBot, which extends and implements an existing dynamic programming algorithm to detect and correct frameshift artifacts and filter out non-target reads, the HMMER3 aligner with preconfigured HMM models or user-supplied protein seeds, and RDP mcClust, which implements a memory-constrained hierarchical clustering algorithm for clustering large numbers of protein reads and includes a distributed computing option. This pipeline has been tested on important functional genes from the carbon and nitrogen cycles, and genes important for human-microbe interactions, including biphenyl dioxygenase (*bph*), important for bioremediation, nitrogenase reductase (*nif*H), a key component in nitrogen fixation, and butyryl-CoA transferase (*but)* and butyrate kinase (*buk*), important for production of butyrate, the main energy source of human colonic epithelial cells.

The RDP's mission includes user support. Help is available online, through e-mail (rdpstaff@msu.edu), and by phone (517-432-4998). In the past year, RDP staff has helped users through over 900 emails and phone conversations.

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. Yilmaz, P., Kottmann, R., Field, D., Knight R, Cole, J.R., Amaral-Zettler L. et al. 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nature Biotechnology* 29: 415-420.
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.

The RDP is supported by the Office of Science (BER), U.S. Department of Energy, Grant No. DE-FG02-99ER62848.

228

The NamesforLife Semantic Index of Phenotypic and Genotypic Data

Charles T. Parker,[1] Catherine Lyons,[1] and **George M. Garrity**[1,2]* (garrity@namesforlife.com)

[1]NamesforLife, LLC, East Lansing, Mich. and
[2]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing

http://www.namesforlife.com

Project Goals: Predictive models depend on high quality input data. But not all data are of similar quality nor are all of the data amenable to computational analysis without extensive cleaning, interpretation and normalization. Key among those needed to make the projects such as the DOE Knowledgebase (Kbase) operational are phenotypic data, which are more complex than sequence data, occur in a wide variety of forms, use complex and non-uniform descriptors and are scattered about the literature and specialized databases. Incorporating these data into the Kbase will require expertise in harvesting, modeling and interpreting the data. The *NamesforLife Semantic Index of Phenotypic and Genotypic Data* seeks to address this problem by taking the first steps toward building an ontology of bacterial and archaeal based on the taxonomic literature through the dvelopment of a draft vocabulary of phenotypic features of the taxonomic type strains.

Unlike sequence data, which are essentially universal, uniform and predictable, phenotypic data are inherently complex, noisy and "taxonomically parochial". The same trait may vary significantly under different conditions of growth, at different times during the life of a cell and under different environmental conditions. The language of phenotype is also complex and may be limited in taxonomic scope, requiring expert interpretation, as there is no equivalent to BLAST for searching for phenotypic data, and there is no central repository for such data. In some cases an entire language exists to describe the phenotypic features that apply to a single phylum, class or order (e.g., Cyanobacteria, Actinobacteria) or a particular class of features (e.g., lipids, structural carbohydrates). In addition, phenotypic data must be viewed in a historical perspective (time when the data were collected) to understand what was measured and how it was measured. As was the case with microbial physiology, which had fallen out of fashion as a field of research, so too has "classical" or polyphasic microbial taxonomy resulting in a dwindling community of experts who can readily interpret the existing information in the literature and in various specialized databases. How might that expertise be captured and applied to developing a standardized language and ontology of microbial phenotype?

In 2003, Garrity and Lyons proposed a novel approach to resolving ambiguity of biological nomenclature. Their approach provided a means of resolving the complex relationships that exist among names and the concepts and objects to which names apply. When coupled with Digital Object Identifiers (DOIs) their method provided a means by which names in digital content (e.g. journal articles, technical reports, web pages) and databases could be made actionable and directly linked to expertly curated information about the name, including its history of changes. NamesforLife, LLC has developed a suite of web services and applications based on this method that can be used to semantically enrich or enhance digital in a variety of formats. The Company has already demonstrated that vectors of names (Semiotic Fingerprints) can be used to index and cross-classify large corpora of scientific and patent literature based on the relationship between named organisms and the underlying subject matter of subsets of documents. The methods and tools are not, however, restricted to biological nomenclature and can be applied to terminologies of all types.

The long-term objective of this STTR project is to develop a semantic index of bacterial and archaeal phenotypes that can be used to augment annotation efforts and to provide a basis for predictive modeling of microbial phenotype. The index will be based on published descriptions of taxonomic type and non-type strains that have been the subject of ongoing genome sequencing efforts as this will provide a mechanism whereby hypotheses can be tested and verified, reproducibly. This project is tightly coupled with ongoing DOE projects (Genomic Encyclopedia of Bacteria and Archaea, the Microbial Earth Project, the Community Sequencing Project) and with two key publications, Standards in Genomic Sciences (SIGS) and the International Journal of Systematic and Evolutionary Microbiology

* Presenting author

(IJSEM). The first step towards accomplishing this goal, and the primary objective of this Phase I project is the development of a draft vocabulary of phenotypic features.

Our approach towards developing a draft vocabulary of bacterial and archeal phenotype is based on a textual analysis of the richest source of descriptive information; the taxonomic literature. It follows a well-established path used for ontology construction based on derivation of domain-dependent hyponymy (is-a relationships) from a corpus and leverages tools, data resources and expertise that the Company has already developed. For this Phase I project, our target corpus consists of a subset of taxonomic literature of type strains from the IJSEM (2003-2009) and SIGS (2009-2011). These articles have been further subdivided into those pertaining to a single bacterial species or multiple species and higher taxa. The articles have been indexed with Apache Lucene to produce two separate indices; one of the complete documents and one of the descriptions of each new organism alone. We are currently developing a KWIC (KeyWord In Context) interface to permit location and display of a given word in the corpus in its surrounding context to understand usage variations within and across different taxa. Selection of terms for analysis is being done using Apache Luke, which provides facilities for determining usage frequency, coupled with curatorial review for relevance, categorization and synonymy. Then end goal of the Phase I project is to allow end users to view an article with all of the phenotypically relevant terms highlighted based on the KWIC index, to use the KWIC index to auto-populate phenotypic characteristics of a given strain based on the published literature and to allow endusers to adjust/negate/select phenotypic characteristics and their values for a strain, using a normalized set of terms.

Funding for the this project was provided through the DOE SBIR/STTR program (DE-SC0006191). Public funding for development of the NamesforLife infrastructure was received from the DOE SBIR/STTR program (DE-FG02-07ER86321), the Michigan Small Business Technology Developemt Corporation, the Michican Statgetic Fund, and the Michigan Universities Commercialization Initiative.

229

Development of Novel Random Network Theory-Based Approaches to Identify Network Interactions Among Nitrifying Bacteria

Ye Deng* (dengye@glomics.com), Zhili He, Feifei Liu, Yujia Qin, Joy Van Nostrand, and **Jizhong Zhou**

Institute for Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, Norman

Project Goals: Interactions among various microbial populations in a community play critical roles in determining the functioning of an ecosystem. However, little is known about such network interactions. Also, nitrification is an important step in the nitrogen cycle, and two groups of microorganisms, ammonia-oxidizing bacteria (AOB) and ammonia-oxidizing archaea (AOA) are thought to play essential roles in this microbe-mediated process. Thus the ultimate goal of this project is to develop a novel molecular ecological network approach and analysis pipeline to understand the genetic diversity and interaction of soil AOA and AOB in a grassland ecosystem. Three specific objectives have been pursued: (i) To develop a molecular ecological network approach and analysis pipeline; (ii) To determine the genetic diversity of soil AOA and AOB populations by pyrosequening of bacterial and archaea *amoA* genes; and (iii) To examine interactions of soil AOA and AOB populations by molecular ecological network analysis of *amoA* pyrosequening data.

Development of functional molecular ecological network approach and pipeline. Theoretically, the interactions among various microbial populations in a community play critical roles in determining the functioning of an ecosystem; however, these relationships remain unclear due to the lack of appropriate experimental data and computational analytic tools. To address such challenges, a mathematical approach for identifying molecular ecological networks (MENs) from high-throughput metagenomics sequencing data has been developed. Two major steps are involved: (i) construction of network based on the Random Matrix Theory; and (ii) network characterization using various mathematical approaches. Compared to other network reconstruction methods, the RMT-based approach is remarkable in that the networks are automatically defined and robust to noise, thus providing excellent solutions to several common issues associated with high-throughput metagenomics data. The MEN topological analyses have integrated the most recent findings in social and biological network analyses, such as general network structures (small-world, scale-free, modularity and hierarchy structures) and module-based eigengene analysis, for dissecting community organizations at the whole and module levels. Furthermore, we associated module characteristics with environmental traits to understand the importance of network interactions in determining community functions. To facilitate application by the scientific community, all of these methods and statistical tools have been integrated into a comprehensive Molecular Ecological Network Analysis Pipeline (MENAP), which is open-accessible through the internet.

Pyrosequencing analysis of bacterial and archaeal *amoA* genes. We have designed universal primers for amplifying archaeal or bacterial *amoA* genes (ammonia monoxygenase subunit A). Both bacterial *amoA* and archaeal *amoA* genes were amplified from soil samples in a grassland experiment site, BioCON (CO_2 concentrations, nitrogen fertilization and plant species), followed by 454 pyrosequencing. A total of 1.2M reads were obtained. After preprocessing, about 651K reads remained with 140K for bacteria and 511K for archaea. A total of 1911 OTUs for bacteria and 7922 OTUs for archaea were obtained at a similarity cutoff 0.95. Ammonium oxidizing bacteria (AOB) were dominated by species in *Nitosospiram, Nitrosovibrio* and *Nitrosomonas* genera (beta-proteobacteria) and Ammonium oxidizing archaea (AOA) were mostly from uncultured species in the *Crenarchaeota* phylum. Further analysis showed that plant diversity and nitrogen fertilization had clear impacts on the community structures of the nitrifying communities, while elevated CO_2 did not have significant effects.

To understand the interactions of AOA and AOB populations in a community, the novel RMT-based network analysis were performed based on the sequencing data. Networks with expected characteristics were obtained. Further topological analysis of AOA and AOB interactions is in progress.

230

The PhyloFacts Phylogenomic Encyclopedia of Microbial Gene Families

Yaoqing Shen,[1] Bushra Samad,[2] Ajithkumar Warrier,[1] Ruchira Datta,[1] and **Kimmen Sjölander**[1,2,3]* (kimmen@berkeley.edu)

[1]QB3 Institute, [2]Department of Bioengineering, and [3]Department of Plant and Microbial Biology, University of California, Berkeley

http://makana.berkeley.edu/phylofacts/

Project Goals: To improve the precision of microbial (meta)genome functional annotation by providing phylogenomic analyses of microbial gene families, providing access to these analyses to the scientific community in the PhyloFacts Database. This pipeline includes: clustering all microbial genes from whole genomes into gene families including homologs from other species; constructing multiple sequence alignments and estimating protein family trees; ortholog identification; integrating experimental and annotation data; computationally scalable methods for HMM classification of (meta)genome sequences to PhyloFacts families and orthology groups.

PhyloFacts is a phylogenomic encyclopedia of gene family trees across the Tree of Life (1). Gene families are defined based on (1) agreement at the multi-domain architecture, and (2) on containing a single Pfam domain in common. The FlowerPower algorithm (2) is used to retrieve homologs from the UniProt database, parameterized separately for the two homology clustering criteria. For each gene family, we construct a multiple sequence alignment and phylogenetic tree; phylogenetic trees are used to identify orthologs in different species using the PHOG algorithm (3). This combination of single domain and domain-architecture clustering enhances the recall and precision of functional classification and orthology identification (4). We construct a hidden Markov model (HMM) for the family and use it to identify homologous protein structures by scoring the Protein Data Bank. Finally, we retrieve experimental and annotation data from various external resources, including UniProt, Gene Ontology, Pfam and KEGG and use these to provide informative descriptions of each family and orthology group, from which the functions of family and orthology-group members can be inferred.

We have focused during the last year on expanding our coverage of microbial gene families. More than 7M proteins are included in PhyloFacts families, representing a broad swath of species. For instance, within Archaea, >90% of *Halobacterium salinarum and* >87% of *Sulfolobus solfataricus* are represented by at least one PhyloFacts family. Within Bacteria, 100% of *Escherichia coli K12*, >94% of *Bacillus subtilis*, >91% of *Thermotoga maritime*, >90% of *Geobacter sulfurreducens*, >87% of *Sulfolobus solfataricus* and >79% of *Deinococcus radiodurans* are represented. Within Eukarya, >90% of *Saccharomyces cerevisiae* and >86% of *Arabidopsis thaliana* genes are included. Detailed coverage of representative species with whole genomes is presented at http://makana.berkeley.edu/phylofacts/coverage/.

Users can access the data in PhyloFacts in several ways, including sequence accession and inputting protein sequences in FASTA format for HMM classification. Data can be downloaded from individual PhyloFacts family pages and can also be downloaded in bulk from http://makana.berkeley.edu/phylofacts/downloads/.

We have also developed a prototype phylogenomic HMM classification system we call FAT-CAT (for Fast Approximate Tree Classification), to allow the functional classification of novel proteins and the simultaneous taxonomic and functional classification of metagenome reads using HMMs placed at internal nodes of gene trees. We will present the results of this analysis at the meeting.

References

1. Krishnamurthy, N., Brown, D.P., Kirshner, D. and Sjölander, K. (2006) PhyloFacts: an online structural phylogenomic encyclopedia for protein functional and structural classification. *Genome biology*, 7, R83.
2. Krishnamurthy, N., Brown, D. and Sjölander, K. (2007) FlowerPower: clustering proteins into domain architecture classes for phylogenomic inference of protein function. *BMC evolutionary biology*, 7 Suppl 1, S12.
3. Datta, R.S., Meacham, C., Samad, B., Neyer, C. and Sjölander, K. (2009) Berkeley PHOG: PhyloFacts orthology group prediction web server. *Nucleic acids research*, 37(Web Server issue): W84-W89.
4. Sjölander, K., Datta, R.S., Shen, Y. and Shoffner, G.M. (2011) Ortholog identification in the presence of domain architecture rearrangement. *Briefings in bioinformatics*. 12 (5): 413-422.

The PhyloFacts Microbial Encyclopedia is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

231

Microbial ENergy Processes Gene Ontology (MENGO): New Gene Ontology Terms Describing Microbial Processes Relevant for Bioenergy

Trudy Torto-Alalibo[1]* (trudy@vbi.vt.edu), Endang Purwantini,[1] Joao C. Setubal,[1,2] Brett M. Tyler,[1,3] and **Biswarup Mukhopadhyay**[1]

[1]Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg; [2]Department of Biochemistry, Institute of Chemistry, University of São Paulo, Brazil; and [3]Center for Genome Research and Biocomputing, Oregon State University, Corvallis

http://mengo.vbi.vt.edu

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.

The MENGO project 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 GO consortium was formed in 1998 to create universal descriptors, which can be used to describe functionally similar gene products and their attributes.

MENGO, an interest group of the GO consortium seeks to expand term development for microbial processes useful for bioenergy production. Currently, there are over 200 MENGO terms added to the GO. Areas covered include carbohydrate catabolic processes, oligosaccharide binding and transport, hydrogen production and methanogenesis. Additionally, over 200 GO annotations of bioenergy relevant gene products from microbes such as *Clostridium thermocellum*, *Methanosarcina barkeri*, *Bacteroides thetaiotaomicron* and *Chlamydomonas reinhardtii* have been made. A selection of terms and annotations will be highlighted in this presentation.

The MENGO interest group will also host a workshop right after the DOE contractor-grantee meeting on February 29th at the same venue. This workshop will highlight progress made in GO term development and microbial gene annotation as well as some of the challenges encountered. Additionally, we will have an open forum to hear from participants on other bioenergy areas to be targeted for further term development and microbial genomes to be annotated.

Funding for the MENGO project is provided by the Department of Energy as part of the Systems Biology Knowledgebase program- Grant# DE-SC000501.

232

Gene Ontology Terms Describe Biological Production of Methane

Endang Purwantini[1]* (epurwant@vbi.vt.edu), Trudy Torto-Alalibo,[1] Joao C. Setubal,[1,2] Brett M. Tyler,[1,3] and **Biswarup Mukhopadhyay**[1]

[1]Virginia Bioinformatics Institute, Virginia Polytechnic Institute, Blacksburg; [2]Department of Biochemistry, Institute of Chemistry, University of São Paulo, Brazil; and [3]Center for Genome Research and Biocomputing, Oregon State University, Corvallis

http://mengo.vbi.vt.edu

Project Goals: The MENGO consortium in collaboration with the community of microbiologists engaged in bioenergy research and the Gene Ontology (GO) consortium aims to develop a comprehensive set of Gene Ontology terms that will describe bioenergy related biological processes and to annotate relevant microbial genomes with appropriate GO terms.

The MENGO project is a community-oriented multi-institutional collaborative effort that aims to develop new Gene Ontology (GO) terms to describe microbial processes of interest to bio-energy production. Such terms will aid in the comprehensive annotation of relevant genes in diverse microbial genomes. Among the 200 terms developed so far are a comprehensive set that describes processes involved in the biological production of methane/methanogenesis.

Biologically, methane is generated by methanogenic archaea from H_2 + CO_2, secondary alcohol + CO_2, formate, carbon monoxide, acetate, methanol, methylamines, and methanethiols. Pathways for methanogenesis from these substrates use unusual coenzymes such as coenzyme F_{420}, methanofuran, tetrahydromethanopterin, coenzyme M, cofactor F_{430}, and coenzyme B. Methanogenesis allows efficient mineralization of biological polymers in anaerobic niches of nature and thereby plays an important role in carbon cycle. This integrated process is leveraged for the production of methane from renewable resources and for waste treatment. The MENGO team has created some terms that are useful for describing the biological processes allowing methanogenesis from carbohydrates, including the biosynthesis of relevant coenzymes. Additionally, methanogenesis related gene products of certain methanogenic archaea such as *Methanocaldococcus jannaschii, Methanosarcina barkeri, Methanosarcina thermophilla, Methanosaeta concilii, Methanopyrus kandleri,* and *Methanothermobacter marburgenensis* have been manually annotated with GO terms.

Funding for the MENGO project is provided by the Department of Energy as part of the Systems Biology Knowledgebase program—grant# DE-SC000501.

233

Toward System Biology KnowledgeBase on Transcriptional Regulation in Bacteria

Pavel S. Novichkov[1]* (PSNovichkov@lbl.gov), Dmitry A. Ravcheev,[2,3] Alexey E. Kazakov,[1] Semen A. Leyn,[2,3] Adam P. Arkin,[1] Inna Dubchak,[1] and **Dmitry A. Rodionov**[2,3]* (rodionov@burnham.org)

[1]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]Sanford-Burnham Medical Research Institute, La Jolla, Calif.; and [3]Institute for Information Transmission Problems RAS, Moscow, Russia

http://regprecise.lbl.gov

Project Goals:

1. **Develop methods for genome-scale regulon reconstruction utilizing the comparative genomics approach and analysis of RNA and DNA regulatory sites.**
2. **Infer the regulatory interactions and reconstruct transcription regulatory networks (TRNs) in multiple groups of microbial genomes important for DOE mission.**
3. **Integrate the diverse sets of predicted and experimental data on microbial transcriptional regulation into a RegKnowledgeBase.**

Transcriptional regulation of gene expression in response to extracellular and intracellular signals is a key mechanism for successful adaptation of microorganisms to changing environmental conditions. Genes and operons directly co-regulated by the same transcription factor (TF) or by an RNA motif are considered to belong to a *regulon*. All regulons taken together form the transcriptional regulatory network (TRN) of the cell. Availability of complete genomes stimulated wide application of a computational genomics-based approach implemented in the RegPredict Web-server platform for fast and accurate inference of microbial regulons. During the past decade a large number of manually-curated high quality inferences of transcriptional regulons were accumulated for diverse taxonomic groups of bacteria.

We have developed the RegPrecise database (http://regprecise.lbl.gov) for capturing, visualization, and analysis of computationally predicted regulons in microbial genomes. The primary object of the database is a single regulon in a particular genome. Each regulon description contains a regulator, its effector, a set of target genes and operons, and their associated *cis*-regulatory sites. Each TF-operated regulon also has a DNA-binding site model (profile) represented as a nucleotide logo.

Bacterial TRNs are highly flexible in evolution of microbial genomes. The effective large-scale reconstruction of TRNs by comparative genomics requires selection of optimal sets of closely-related genomes. Therefore, the central strategy for regulon analysis in microbial genomes in RegPrecise is based on subdivision of all microbial species into small taxonomic groups that are analyzed independently. The highest level of regulon organization in the database is represented by taxonomic collections of regulons. The current version of RegPrecise contains 13 taxonomic collections of regulons covering major phyla of Bacteria (Proteobacteria, Firmicutes, Cyanobacteria, Actinobacteria etc.).

The total number of TF and RNA-motif regulons in the current release of RegPrecise database exceeds seven thousands. RegPrecise provides three classifications of regulons implemented as controlled vocabularies: (i) biological processes /metabolic pathways; (ii) effectors /environmental signals; (iii) TF protein families. Biological processes attributed to regulons in the database covers a wide spectrum of the cellular metabolism (Fig. 1).The current list of effectors of analyzed TFs includes more than 200 metabolites from the following major classes: amino acids, carbohydrates, nucleotides, lipids and fatty acids, co-enzymes, peptides and antibiotics, secondary metabolites, and inorganic chemicals. Regulons represented in RegPrecise includes ~5400 TFs from >50 TF protein families.

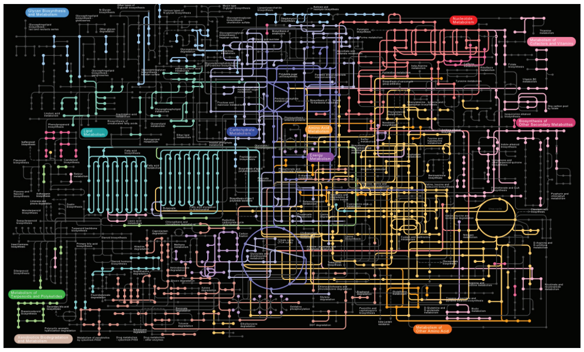


Figure 1. Overview of metabolic pathways covered by reconstructed TF regulons in RegPrecise DB.

The content of RegPrecise database is publically available through the RESTful web-services in JSON format.

In the next release of RegPrecise we will continue extending the regulon content to cover other diverse taxonomic groups. We are also planning to conduct the large-scale assignment of confidence levels to the predicted regulons based on available experimental evidences from literature and external web-resources (EcoCyc, CoryneRegNet, DBTBS, RegTransBase).

RegPrecise is a key component of the upcoming DOE Systems Biology KnowledgeBase. It will provide essential datasets of reference regulons in diverse microbes to enable automatic reconstruction of draft TRNs in newly sequenced genomes.

This research is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), under contract DE-SC0004999 with Sanford-Burnham Medical Research Institute and Lawrence Berkeley National Lab.

234

Reference Collection of Transcriptional Regulons in Bacillales

Semen A. Leyn[1,2]* (sleyn@burnham.org), Marat D. Kazanov,[2] **Pavel S. Novichkov**,[3] and **Dmitry A. Rodionov**[1,2]

[1]Sanford-Burnham Medical Research Institute, La Jolla, Calif.; [2]Institute for Information Transmission Problems RAS, Moscow, Russia; and [3]Lawrence Berkeley National Laboratory, Berkeley, Calif.

http://regprecise.lbl.gov/

Project Goals:

1. **Develop integrated platform for genome-scale regulon reconstruction utilizing the comparative genomics approach and analysis of RNA and DNA regulatory sites.**
2. **Infer the regulatory interactions and reconstruct transcription regulatory networks in several groups of microbial species important for DOE mission.**
3. **Develop RegKnowledgeBase on microbial transcriptional regulation.**

Gram-positive facultative anaerobic bacteria from the Bacillales order were isolated from diverse habitats including soil, sea water, plants and animals. Bacillales use various strategies to respond and survive in a variety of stresses and environmental conditions including resistance to multiple antibiotics. *Bacillus subtilis* str. 168 is one of the best-characterized Gram-positive bacteria and a model organism for studying sporulation, cell differentiation, stress response and social behavior of bacteria. According to the DBD database, *B. subtilis* genome encodes 238 DNA-binding transcription factors (TFs) classified in 45 protein families. Of them, 120 TFs were studied experimentally and the respective regulatory interactions were captured in the DBTBS database. However, many of the previously studied TF regulons were studied insufficiently, providing an incomplete knowledge on the range of target genes and associated TF-binding sites (TFBSs).

Continuously growing number of available complete genomes allows successful application of the comparative genomics approaches for regulon analysis. We used a "knowledge-driven" approach, which combines the accumulated experimental information from literature and databases with novel bioinformatics tools for genomic reconstruction of regulatory interactions. We perform the comparative genomics reconstruction of regulons operated by either TFs or RNA-regulatory element using the RegPredict Web-server (http://regpredict.lbl.gov). RegPredict allows prediction of TFBS and RNA motifs in a group of selected genomes, with further identification and annotation of candidate members of the respective regulons. Functional analysis of target genes was based on annotations from the SEED database that were validated by the genomic context analysis in MicrobesOnLine.

In this study we carried out large-scale comparative genomics analysis of regulatory interactions in *B. subtilis* and 10 related species from the Bacillales order. For TF regulons, we first analyzed 59 regulons with previously known TFBS motifs according to literature and the DBTBS database on transcriptional regulation in *B. subtilis*. These known regulons were expanded by prediction of novel targets in *B. subtilis* and propagated to other studied genomes of Bacillales, resulting in refinement of TFBS motifs and identification of novel regulon members. Then we predicted novel TFBS motifs and reconstructed 32 TF regulons for which target genes have been previously defined in *B. subtilis* but whose TFBSs were unknown. Finally, we discovered novel TFBS motifs and reconstructed regulons *de novo* for 34 previously uncharacterized TFs. Novel regulons involve genes from the following biological processes: utilization of various carbohydrates; metabolism of glutamate, histidine and thiamine; stress responses; drug/metabolite transport. Totally, more than 3500 TFBSs have been predicted in the *Bacillales* group (from 200 to 600 sites per genome).

For RNA-operated regulons, we used bacterial RNA regulatory motifs collected from the Rfam database, scanned the studied genomes with these motifs to identify new occurrences of each RNA family, and annotated the respective target operons. Among 37 reconstructed RNA regulons there are 11 families of metabolite-sensing riboswitches, 17 types of aminoacyl-tRNA-responsive T-boxes, one regulon controlled by the RNA-binding protein PyrR, five predicted ribosomal protein leader RNA structures and five regulons for *in silico* predicted RNA motifs of unknown function. The reconstructed RNA motif-operated regulons in Bacillales control key metabolic pathways including biosynthesis of vitamins and cofactors (cobalamin, riboflavin, thiamine, nucleoside queuosine), biosynthesis of glucosamine, metabolism of most amino acids, biosynthesis and salvage of purines and pyrimidines, and magnesium homeostasis.

The reference collection of transcriptional regulons in the Bacillales group of bacteria is available in the RegPrecise database (http://regprecise.lbl.gov/). Currently, this collection constitutes the biggest transcriptional regulatory network among various taxonomic groups with reconstructed TRNs in RegPrecise.

This research is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), under contract DE-SC0004999 with Sanford-Burnham Medical Research Institute and Lawrence Berkeley National Lab.

* Presenting author

235

Community-Based Approach for Genome-Wide Regulon Annotation in Bacteria

Dmitry A. Ravcheev[1,2]* (dravcheev@sanfordburnham.org), Aaron A. Best,[3] Mikhail S. Gelfand,[2,4] **Pavel S. Novichkov**,[5] and **Dmitry A. Rodionov**[1,2]

[1]Sanford-Burnham Medical Research Institute, La Jolla, Calif.; [2]Institute for Information Transmission Problems RAS, Moscow, Russia; [3]Hope College, Holland, Mich.; [4]Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia; and [5]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.

http://regprecise.lbl.gov

Project Goals:

1. **Develop integrated platform for genome-scale regulon reconstruction utilizing the comparative genomics approach and analysis of RNA and DNA regulatory sites.**
2. **Infer the regulatory interactions and reconstruct transcription regulatory networks (TRNs) in several groups of microbial species important for DOE mission.**
3. **Develop RegKnowledgeBase on microbial transcriptional regulation.**

In the light of the constantly growing number of complete genomic sequences accurate genome-scale annotation of regulatory features is a one of the critical task of modern genomics and system biology. Inference of regulatory interactions in newly sequenced genomes allows reconstruction of the transcriptional regulatory networks (TRNs) and effective modeling of cellular metabolic and signal pathways. TRN is a fine-tuned system including controllers, such as transcription factors (TFs), their binding sites (TFBSs), and RNA regulatory elements (e.g., riboswitches), and sets of target genes whose expression is regulated by these controllers. Genome-scale reconstruction of TRNs in bacteria requires accurate prediction of a large number of regulatory interactions between controllers and their targets. A regulon, defined as a set of genes under the direct control of a certain controller, constitutes a 'building block' of each TRN. The comparative genomics approach was successfully applied for reconstruction of multiple regulons in diverse bacterial groups. A regulog concept is used to represent a regulon inferred and projected in a group of closely-related bacterial genomes. High-quality reconstruction of each regulog constituting the combined TRN in a group of taxonomically-related genomes is a labor intensive work that can be carried out by the scientific community.

We developed a workflow for coordinated reconstruction of multiple bacterial regulons by the community annotators (Fig. 1). The workflow utilizes the comparative genomics-based approach implemented in the RegPredict Web-server (http://regpredict.lbl.gov). This reconstruction pursues two main tasks: (i) propagation of the known regulon to new genomes; (2) expansion of the known regulon by prediction of new regulon members. Initial datasets of regulatory interactions including information about TFs, the previously mapped TFBSs and sets of regulated genes (mostly experimentally determined) are extracted from literature and public databases, such as DBD, EcoCyc, CoryneRegNet, DBTBS, RegTransBase, and RegPrecise. These data are mapped to complete genomes forming initial sets of tasks for community annotators. Each regulon is reconstructed by a community annotator in the reference set of taxonomically-related genomes using the computational platform RegPredict. After the completion of regulon annotation task, each output regulog undergoes quality control by curators. Finally approved regulogs include information on regulators and their targets in a group of taxonomically-related genomes.

The community-based workflow was efficiently used for initial TRN inference in four taxonomic groups of bacteria, including Enterobacteriaceae, Lactobacillaceae, Streptococcaceae, and Corynebacteriaceae. The reference set of inferred transcription factor regulogs is available in the RegPrecise database (http://regprecise.lbl.gov) and includes 173 regulogs described in 50 genomes.

In the Enterobacteriaceae lineage (12 species including *Escherichia coli*), 64 regulons were reconstructed by a community of 27 undergraduate students from Moscow State University (MSU) in Russia. These include the global regulons ArgR, Crp, Fur, FruR, LexA, and PurR that control from tens to a hundred of targets per genome. The obtained collection of the Enterobacteriales regulogs contains more than 600 target genes, including ~100 new regulon members. Another community of 36 MSU students performed annotation of 43 regulons in the Corynebacteriaceae lineage (8 species including *Corynebacterium glutamicum*). The final set of reconstructed regulons comprises more than 150 target operons, including multiple novel members of regulons. Streptococcaceae and Lactobacillaceae are two closely related lineages of the Firmicutes phylum. Regulon reconstruction in 30 selected genomes from these two taxonomic groups was carried by a community of 18 undergraduate students from Hope College (Holland, MI). As a result of this community effort, 33 regulons were inferred in each of these two lineages. Among them, 22 regulons are shared between the two lineages, whereas the remaining 18 regulons are lineage-specific. Final collections of regulons in Streptococcaceae and Lactobacillaceae contain 200-300 regulated operons per genome with multiple novel predicted regulon members.

The community-based approach for *in silico* reconstruction of multiple regulons applied in this work to four taxonomically diverse groups of bacterial genomes is a promising approach for large-scale annotation of regulatory features. Detailed TRNs obtained by this approach for DOE-mission genomes will constitute an important dataset for the forthcoming DOE Systems Biology KnowledgeBase.

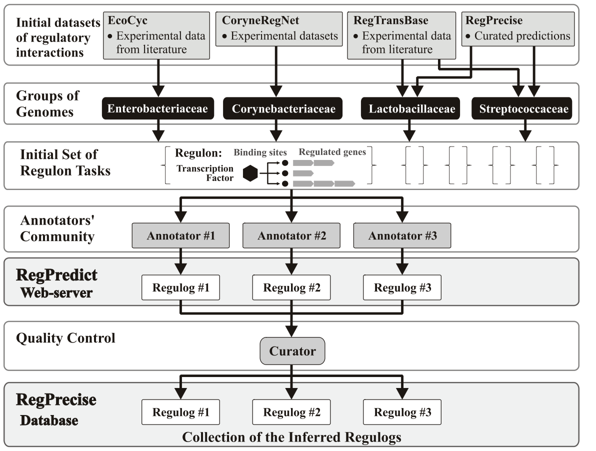

Initial datasets of regulatory interactions
EcoCyc
• Experimental data from literature
CoryneRegNet
• Experimental datasets
RegTransBase
• Experimental data from literature
RegPrecise
• Curated predictions
Groups of Genomes
Enterobacteriaceae
Corynebacteriaceae
Lactobacillaceae
Streptococcaceae
Initial Set of Regulon Tasks
Regulon:
Binding sites
Regulated genes
Transcription Factor
Annotators' Community
Annotator #1
Annotator #2
Annotator #3
RegPredict Web-server
Regulog #1
Regulog #2
Regulog #3
Quality Control
Curator
RegPrecise Database
Regulog #1
Regulog #2
Regulog #3
Collection of the Inferred Regulogs

Figure 1. Workflow for genome-wide regulon annotation by scientific community.

This research is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), under contract DE-SC0004999 with Sanford-Burnham Medical Research Institute and Lawrence Berkeley National Lab.

236

Computational Methodologies for Identification of Phenotype-Specific Biological 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), **Jill Banfield**[3] (jbanfield@berkeley.edu), Matthew C. Schmidt[1,2] (mcschmid@ncsu.edu), William Hendrix[1,2] (wthendri@ncsu.edu), Andrea M. Rocha[4] (amrocha@mail.usf.edu), and Kanchana Padmanabhan[1,2]* (kpadman@ncsu.edu)

[1]North Carolina State University, Raleigh; [2]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]University of California, Berkeley; and [4]University of South Florida, Tampa

http://freescience.org/cs/gtl/

Project Goals: Develop computational methodologies to identify phenotype-related biological systems and their interplays. The crux of the project is a procedure for: (1) identification and characterization of phenotype-related genes; (2) identification of phenotype-biased cellular subsystems; (3) reconstruction of phenotype-specific metabolic pathways; and (4) elucidation of symbiotic/competing crosstalks between these pathways.

Phenotypes that certain microorganisms express assist in activities like breaking down the lignocellulosic barrier of biomass, and the biodegradation of various environmental contaminants. Tackling the various bioremediation and bioenergy problems with the help of genetic engineering requires the understanding of the cellular subsystems that help with the phenotype-expression in the organism. To supplement experimentation methods, computational methodologies need to be used. These methods could reveal phenotype-related "signals'" and their combinatorial interplay by comparing potentially hundreds of microorganisms with millions of genes organized into thousands of cellular subsystems.

We developed **graph-theoretical and statistical methods and released open-source software for *in silico* prediction of cellular subsystems related to the expression of a target phenotype**.

The Network Instance-Based Biased Subgraph Search (**NIBBS)** [5] is capable of comparing hundreds of genome-scale metabolic networks to identify ***metabolic subsystems that are statistically biased toward phenotype-expressing organism***. NIBBS identifies the set of all phenotype-*biased* metabolic network motifs. From the results obtained, for example, for bio-hydrogen production phenotype, NIBBS was able to identify metabolic subsystems like acetate and butyrate fermentation, fatty acid biosynthesis, amino acid metabolism, and nitrogen metabolism. The validation for those results was found in literature. In addition, NIBBs was also able to provide clues about pathway cross-talks, including those involved in production of Acetyl-CoA.

The **α,β-motif finder** [2] and **bi-clustering** [3] approach allow for identification of ***phenotype-related functional modules*** that, in addition to metabolic subsystems, could include their regulators, sensors, transporters. The functional modules identified are present across a set of phenotype-expressing organisms. This approach can identify conserved modules across any subset of the input organisms and hence can identify sub-phenotype-specific modules as well. From the results obtained, for example, for the light fermentation hydrogen production phenotype, the modules associated with N-fixation, iron regulation, and ammonia uptake were found. When applied to the bio-hydrogen production phenotype, this method was able to identify modules 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. On further analysis, various pathway crosstalks including those between iron and nitrogen related metabolic pathways and iron uptake and ammonia assimilation were predicted.

The Dense ENriched Subgraph Enumeration (**DENSE**) [1] algorithm allows for ***incorporating partial prior knowledge about the proteins involved in a phenotype-related process into the identification of sets of functionally associated proteins in a phenotype-expressing organism***. This method, when applied to the protein functional association network of the *Clostridium acetobutylicum*, a dark fermentative, hydrogen producing bacterium was able to predict known and novel associations including those with regulatory, signaling, and uncharacterized proteins.

We supplemented the computational methodologies discussed so far with a method to computationally analyze the results of the methods for biological significance and improve functional annotation [4]. This method is different from other existing significance analysis techniques in that it takes into account an inherent design principle of biological networks, *hierarchical modularity*. Functionally homogenous modules combine in a hierarchical manner into larger, less cohesive subsystems. The method quantifies biological significance both with a score known as **Hierarchical Modularity Score (HMS)** and a confidence of the score via a *p*-value. Additionally, it provides the hierarchical functional annotation of the modules.

References

1. Willam Hendrix, Andrea M Rocha, Kanchana Padmanabhan, Alok Choudhary, Kathleen Scott, James R Mihelcic and Nagiza F Samatova, *DENSE: efficient and prior knowledge-driven discovery of phenotype-associated protein functional modules*, BMC Systems Biology 2011, 5:172.
2. Mathew C Schmidt, Andrea M Rocha, Kanchana Padmanabhan, Zhengzhang Chen, Kathleen Scott, James R Mihelcic and Nagiza F Samatova, *Efficient alpha,beta-motif Finder for Identification of Phenotype-related Functional Modules*, BMC Bioinformatics 2011, 12:440.
3. Kevin Wilson, Andrea Rocha, Kanchana Padmanabhan, Kuanyu Wang, Zhengzhang Chen, Ye Jin, James R Mihelcic and Nagiza F Samatova, *Detecting Pathway Cross-talks by Analyzing Conserved Functional Modules across Multiple Phenotype-Expressing Organisms*, IEEE International Conference Bioinformatics and Biomedicine, 2011.
4. Kanchana Padmanabhan, Kuanyu Wang, Nagiza Samatova, *Functional Annotation of Hierarchical Modularity,* (under review).
5. Mathew C Schmidt, Andrea M Rocha, Kanchana Padmanabhan, Yekaterina Shpanskaya, Jill Banfield, Kathleen Scott, James R Mihelcic and Nagiza F Samatova, *NIBBS-Search for Fast and Accurate Prediction of Phenotype-Biased Metabolic Systems,* (under review).
6. Zhengzhang Chen, Kanchana Padmanabhan, Andrea M. Rocha, Yekaterina Shpanskaya, James R. Mihelcic and Nagiza F. Samatova, *SPICE: Discovery of Phenotype-Determining Component Interplays*, (under review).

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.

237

OptCom: A Multi-Level Optimization Framework for the Metabolic Modeling and Analysis of Microbial Communities

Ali R. Zomorrodi* (zomorrodi@engr.psu.edu) and **Costas D. Maranas**

Department of Chemical Engineering, The Pennsylvania State University, University Park

Project Goals: The goal of this project is to develop an efficient and comprehensive computational framework for the flux balance analysis of microbial communities using genome-scale metabolic models.

Microorganisms rarely live isolated in their natural environments but rather function in consolidated and socializing communities. Despite the growing availability of high-throughput sequencing and metagenomic data, we still know very little about the metabolic contributions of individual microbial players within an ecological niche and the extent and directionality of interactions among them. This calls for development of efficient modeling frameworks to shed light on less understood aspects of metabolism in microbial communities. Here, we introduce OptCom, a comprehensive flux balance analysis framework for microbial communities, which relies on a multi-level/objective optimization formulation to properly describe trade-offs between individual vs. community level fitness criteria (see Figure 1). In contrast to earlier approaches that rely on a single objective function, here, we consider species-level fitness criteria for the inner problem while relying on community-level objective maximization for the outer problem. OptCom is general enough to capture any type of interactions (positive, negative or combinations thereof) and is capable of accommodating any number of microbial species (or guilds) involved.

To quantify the deviation of community members from their optimal behavior, we introduce a metric called 'optimality level' (c^k) for each species involved. The optimality level for each one of the microorganisms is quantified using a variation of OptCom which we refer to as *descriptive* through incorporating all available experimental data for the entire community (e.g., community biomass composition) as constraints in the outer problem and all data related to individual species as constraints in the respective inner problems while allowing the biomass flux of individual species to fall below (or rise above) the maxima of the inner problems. An optimality level of less than one for a microorganism *k* implies that it grows sub-optimally at a rate equal to $100c^k$ % of its maximum to optimize a community-level fitness criterion while matching experimental observations. Alternatively, an optimality level of one implies that microorganism *k* grows exactly optimally, whereas a value greater than one indicates that it achieves a higher biomass production level than the community-specific maximum (i.e., super-optimality) by depleting resources from one or more other community members.

We applied OptCom to quantify the syntrophic association between *D. vulgaris* and *M. maripaludis* and assess the optimality levels of growth in phototrophic microbial mats of Octopus and Mushroom Springs of Yellowstone National Park. We also used OptCom to elucidate the extent and direction of inter-species metabolite and electron transfer in a model microbial community and examine the possibility of adding a new member to this community. Our study demonstrates the importance of trade-offs between species- and community-level fitness driving forces and lays the foundation for metabolic-driven analysis of various types of interactions in multi-species microbial systems using genome-scale metabolic models.

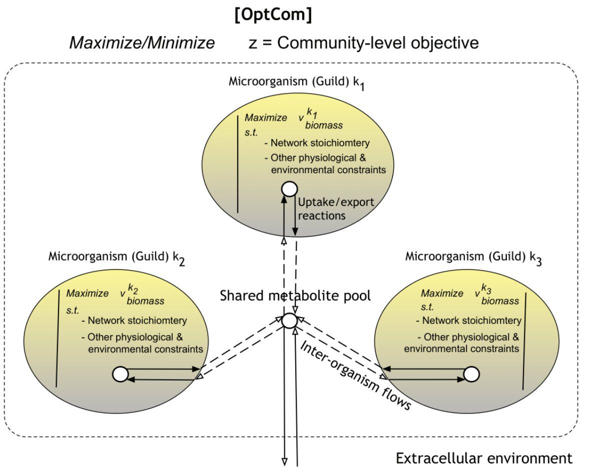

[OptCom]
Maximize/Minimize
z = Community-level objective
Microorganism (Guild) k_1
Maximize $v^{k_1}_{biomass}$
s.t.
- Network stoichiomtery
- Other physiological & environmental constraints
Uptake/export reactions
Microorganism (Guild) k_2
Maximize $v^{k_2}_{biomass}$
s.t.
- Network stoichiomtery
- Other physiological & environmental constraints
Shared metabolite pool
Microorganism (Guild) k_3
Maximize $v^{k_3}_{biomass}$
s.t.
- Network stoichiomtery
- Other physiological & environmental constraints
Inter-organism flows
Extracellular environment

Figure 1. **Pictorial illustration of OptCom.** OptCom relies on a multi-level optimization structure where a separate biomass maximization problem is defined for each species as inner problems. These inner problems are then integrated in the outer stage through the inter-organism flow constraint to optimize a community-level objective function.

238

MetRxn: A Knowledgebase of Metabolites and Reactions Spanning Metabolic Models and Databases

Akhil Kumar[1]* (azk172@psu.edu), Patrick F. Suthers,[2] and **Costas D. Maranas**[2]

[1]Bioinformatics and Genomics, [2]Department of Chemical Engineering, The Pennsylvania State University, University Park

Project Goals: The goal of this project is to develop a comprehensive knowledgebase that standardizes metabolite and reaction entries consolidated from a wide range of databases and existing genome-scale metabolic models.

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. A key bottleneck in the pace of reconstruction of new high quality metabolic models is our inability to directly make use of metabolite/reaction information from biological databases (e.g., BRENDA, KEGG, MetaCyc, EcoCyc, BioCyc, BKM-react, UM-BBD, Reactome.org, Rhea, PubChem, ChEBI etc.) or other models due to incompatibilities of representation, duplications and errors.

A major impediment is the presence of metabolites with multiple names across databases and models, and in some cases within the same resource, which significantly slows down the pooling of information from multiple sources. Therefore, the almost unavoidable inclusion of multiple replicates of the same metabolite can lead to missed opportunities to reveal (synthetic) lethal gene deletions, repair network gaps and quantify metabolic flows. Moreover, most data sources inadvertently include some reactions that may be stoichiometrically inconsistent and/or elementally / charge unbalanced, which can adversely affect the prediction quality of the resulting models if used directly. Finally, a large number of metabolites in reactions are partly specified with respect to structural information and may contain generic side groups (e.g., alkyl groups -R), varying degree of a repeat unit participation in oligomers, or even just compound class identification such as “an amino acid” or “electron acceptor”.

MetRxn is a knowledgebase that includes standardized metabolite and reaction descriptions by integrating information from BRENDA, KEGG, MetaCyc, Reactome.org and 44 metabolic models into a single unified data set. All metabolite entries have matched synonyms, resolved protonation states and are linked to unique structures. All reaction entries are elementally and charge balanced. This is accomplished through the use of a workflow of lexicographic, phonetic, and structural comparison algorithms (see Figure 1). MetRxn allows for the download of standardized versions of existing genome-scale metabolic models and the use of metabolic information for the rapid reconstruction of new ones.

We describe the development and highlight applications of the web-based resource MetRxn that integrates, using internally consistent descriptions, metabolite and reaction information from 8 databases and 44 metabolic models. The MetRxn knowledgebase contains over 76,000 metabolites and 72,000 reactions (including unresolved entries) that are charge and elementally balanced. By conforming to standardized metabolite and reaction descriptions, MetRxn enables users to efficiently perform queries and comparisons across models and/or databases. For example, common metabolites and/or reactions between models and databases can rapidly be generated along with connected paths that link source to target metabolites. MetRxn supports export of models in SBML format. New models are being added as they are published or made available to us. MetRxn uses relational database models (MySQL) and is available as a web-based resource at http://metrxn.che.psu.edu.

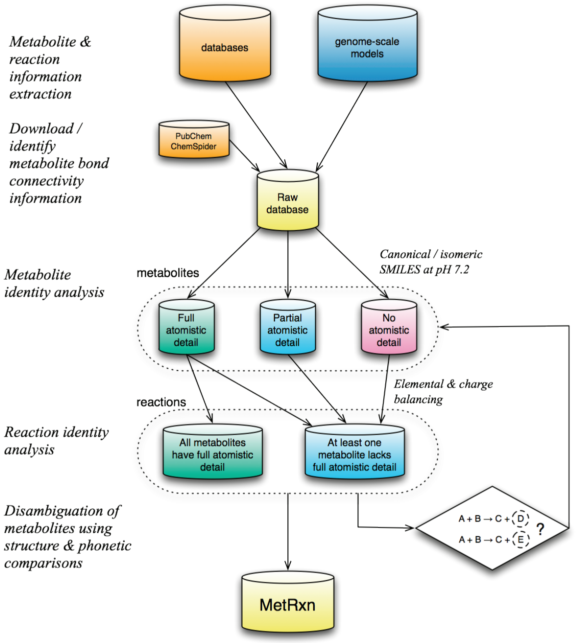

Metabolite & reaction information extraction
Download / identify metabolite bond connectivity information
Metabolite identity analysis
Reaction identity analysis
Disambiguation of metabolites using structure & phonetic comparisons
databases
genome-scale models
PubChem ChemSpider
Raw database
metabolites
Canonical / isomeric SMILES at pH 7.2
Full atomistic detail
Partial atomistic detail
No atomistic detail
Elemental & charge balancing
reactions
All metabolites have full atomistic detail
At least one metabolite lacks full atomistic detail
A + B → C + D
A + B → C + E
?
MetRxn

Figure 1: Flowchart outlining the construction of MetRxn.

239

Computational Design of Bioenergy-Related Metabolic Pathways

Mario Latendresse* (latendre@ai.sri.com), Mike Travers, and **Peter Karp**

SRI International, Menlo Park, California

http://metacyc.org
http://brg.ai.sri.com/ptools

Project Goals: The general goal of our project is to support computational design of metabolic pathways for metabolic engineering applications. Our approach combines the following elements. (1) Users will specify a target metabolite, a feedstock compound, and other constraints on their design problem. (2) A pathway search algorithm will construct alternative pathways by combining reactions from the MetaCyc database. MetaCyc is a highly curated multi-organism metabolic database that currently contains more than 9500 metabolic reactions. The search algorithm will rank pathways according to multiple criteria. (3) Users will view the output of the pathway search algorithm using a graphical interface that facilitates user comprehension and evaluation of the pathways. The resulting software will be an add-on module to SRI's Pathway Tools software.

Pathway Search

We have implemented an initial version of a graph search algorithm to design novel pathways, given a source and target metabolite. This software is capable of integrating a variety of metrics and filters that affect the search. Metrics implemented to date include atom conservation, molecular similarity, avoiding certain molecules, penalties for using a reaction that is outside of the taxonomic range of the target organism, and penalties for generating or consuming certain kinds of side products/reactants. The search engine is capable of using reactions from any PGDB including MetaCyc. The results of the search can be displayed using the web-based reaction atom-mapping viewer, a new capability of Pathway Tools that can display the trajectories of individual atoms through a sequence of reactions.

Future planned work includes adding new search metrics (such as thermodynamic constraints), and an enhanced user interface that allows scientists to interactively view and control the search process.

Automated Computation of Reaction Atom Mappings

An atom mapping describes explicitly the one to one transfer of each atom from the reactants of a reaction to its products. A reaction might have several chemically valid mappings due to the symmetries of reactants and products or for other reasons. These mappings allow the computation of the flow of essential atoms from source to target metabolites in a pathway. Note that a typical reaction equation used by biologists or chemists does not describe the atom mapping, as such details are difficult to provide and might be overwhelming to the reader. When provided, atom mappings are typically described as supplementary data for each reaction.

Some previous work on computing atom mappings does not compute all possible atom mappings and is therefore incomplete. For example, the KEGG RPAIR database describes only one atom mapping per reaction.

As far as we know, all computational approaches to atom mapping published so far, are combined with a post-processing step involving manual curation. That is, a scientist reviews the computed atom mappings for possible errors and appropriate corrections are applied when necessary. But the necessary corrections are not applied to the computational approach itself to avoid future computed errors.

In the approach we have taken, all the chemically valid mappings of each reaction of MetaCyc are computed. Moreover, we aim to have a computational approach that does not require manual post-processing.

Technically, our approach is based on mixed integer linear programming: for each reaction, a linear program is generated from all possible valid mappings of the reaction where the objective is to minimize the sum of the costs of the bonds broken and made. We currently use bond costs that have been determined by a chemist. A linear solver solves the linear program, giving all possible optimal solutions, that is, all possible correct mappings for one reaction. This technique has been applied to almost all the reactions of MetaCyc. It is more computationally efficient than all other known techniques published so far.

The computation of correct mappings depends on the bond costs used. In the near future we intend to systematically validate these bond costs using a linear program based on a sample of correct mappings (positive examples) and incorrect mappings (negative examples). We also plan to include atom mappings in future versions of MetaCyc and of other Pathway/Genome Databases.

---

* Presenting author

This work was funded by the Department of Energy under grant DE-SC0004878.

240

Bioenergy Curation in the MetaCyc Database of Metabolic Pathways

Ron Caspi* (caspi@ai.sri.com), Deepika Weerasinghe, and **Peter Karp**

Bioinformatics Research Group, SRI International, Menlo Park, Calif.

http://metacyc.org
http://biocyc.org

Project Goals: Two goals of our project are (1) To expand the coverage of bioenergy-related metabolic information in the MetaCyc database and (2) To generate within the BioCyc database collection organism-specific PGDBs for all energy-relevant organisms sequenced by JGI.

MetaCyc (metacyc.org) is a literature-curated database containing more than 1,800 metabolic pathways, collected from a wide variety of organisms, with an emphasis on microorganisms and plants. The goal of MetaCyc is to contain a representative sample of every experimentally elucidated pathway, thereby cataloging the universe of known pathways and enzymes. MetaCyc contains rich, detailed, and high quality data, including minireview summaries with extensive literature citations, enzyme information, evidence codes, and links to other databases.

In addition to its role as an encyclopedic resource for metabolic pathways, MetaCyc also serves as a reference database for the Pathway Tools software, which predicts the metabolic pathway complement of an organism from its annotated genome, creating a Pathway/Genome Database (PGDB) for that organism.

Since only pathways that already exist in MetaCyc can be inferred by Pathway Tools in organism-specific databases, the pathway content of MetaCyc strongly influences the metabolic networks predicted by Pathway Tools. Two goals of our project are (1) To expand the coverage of bioenergy-related metabolic information in MetaCyc, and (2) To generate within our BioCyc database collection organism-specific PGDBs for all energy-relevant organisms sequenced by JGI.

During the past year we have done the following:

- Greatly expanded the coverage in MetaCyc of compounds found in cellulosic biomass, with an emphasis on natural cellulosic and hemicellulosic polymers.
- Created 8 pathways for the degradation of important biomass compounds such as celluloses, rhamnogalacturonans, xylans, arabinans, xyloglucans, and carrageenans.
- Curated many enzymes that are involved in cellulosic biomass degradation.
- Curated eight naturally occurring and/or bioengineered hydrogen production pathways and related enzymes.
- Curated ten naturally-occurring and/or engineered biosynthetic pathways for potential biofuel compounds, including 1-butanol, 3-methyl butanol, isopropanol, all-trans-farnesol, long chain fatty acid esters, the algal lipid diacylglyceryl-N,N,N-trimethylhomoserine, the algal fatty acids docosahexanoate, arachidonate, and eicosapentaenoate, and the algal triterpinoid botryococcene.

In the remainder of the second year of this project we will add a new type of pathway diagram that will facilitate the presentation of enzymatic degradation of complex polymers, and will use this new tool to create many pathway diagrams for lignocellulose-degradation pathways.

Reference

1. Caspi et al (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. Database Issue[Epub ahead of print] PMID 22102576.

MetaCyc and BioCyc are funded by grant GM80746 from the NIH National Institute of General Medical Sciences. Bioenergy curation is funded by grant DE-SC0004878 from the Department of Energy.

241

Enhancing the SEED Framework for Curation and Analysis of Genomic Data and Genome-Scale Metabolic Models

Christopher Henry*[1] (chenry@mcs.anl.gov), **Ross Overbeek**,[2] Robert Olson,[1] Terrence Disz,[1] Bruce Parello,[2] Rob Edwards,[3] Matt DeJongh,[4] Aaron Best,[4] Fangfang Xia,[1] Scott Devoid,[1] and **Rick Stevens**[1]

[1]University of Chicago, Ill.; [2]Fellowship for Interpretation of Genomes, Ill.; [3]San Diego State University, Calif.; and [4]Hope College, Mich.

Project Goals: The objective of the SEED Knowledgebase project is encompassed by four goals: (1) enhance the computational infrastructure behind the SEED framework to improve extensibility, accessibility, and scalability; (2) integrate extensions into the SEED database to accommodate new computational and experimental data

* Presenting author

types including regulatory networks, biochemistry and models, eukaryotic genome data, and growth phenotype data; (3) develop web services providing access to all SEED data and algorithms including genome annotation, model reconstruction, flux balance analysis, and data query and access; and (4) apply the SEED framework to annotate and model organisms with applications to bioenergy, carbon cycle, and bioremediation. This project will provide users with programmatic access to SEED data and algorithms, it will produce new models of bioenergy organisms, and it will enable integration of expression data and regulation with annotations and models.

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. In the SEED Knowledgebase project, we have endeavored to enhance the SEED environment in three areas: (1) development of a web API offering programmatic access to SEED data and algorithms (including trees and expression data); (2) extension of the SEED interface to enable curation of public genome annotations by registered users; and (3) reconstruction and analysis of metabolic models for all available prokaryotic genome sequences.

1. Web API for Programmatic Access to Data and Algorithms
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; enabling users to query the SEED database for genome features, functional annotations, gene orthologs, and sequence similarities; and enabling 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. Specifically, we apply the web services to: (1) identify new genome annotations based on model gapfilling; (2) identify commonly clustered and co-expressed sets of functional roles across all known genomes; and (3) to study redundancy of essential metabolic functions across all known genomes.

2. Extension of the SEED to Enable Curation of Public Genomes
We recently launched a new version of the SEED website called the Public SEED (pubseed.theseed.org). This new site now contains over 3000 annotated prokaryotic genomes, and it is continuously updated to include every complete prokaryotic genome sequence that is available in GenBank. In addition to offering a greatly enhanced database of genomes, the Public SEED also offers improved access to genomes, through a powerful new search feature. This feature enables users to rapidly search Public SEED content for organism names, gene names, locus IDs, and many other queries and returns results sorted by object type. The Public SEED also offers a unique ability enabling registered users to alter any genome annotation in the Public SEED, providing a unique resource for the community annotation of genomic data.

3. Reconstruction and Analysis of Metabolic Models for all Prokaryotic Genomes
Over the past decade, 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. We have applied the Model SEED framework to produce 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 and improve the automated generation of biomass composition reactions. These algorithms were applied to assess the quality of annotations in the SEED framework, and to identify high-priority gaps to filled in these annotations. Finally, we applied 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.

242

Tools and Approaches for Integrating Multiple Genetic and Cellular Networks

Gang Fang,[1] Darryl Reeves[1]* (darryl.reeves@yale.edu), Koon-Kiu Yan,[1] Nitin Bhardwaj,[1] Declan Clarke,[1] Chao Cheng,[1] Pedro Alves,[1] Sergei Maslov,[2] and **Mark Gerstein**[1]

[1]Computational Biology and Bioinformatics, Medical School, Yale University, New Haven, Conn. and [2]Department of Condensed Matter Physics and Materials Science, Brookhaven National Laboratory, Upton, N.Y.

http://networks.gersteinlab.org/

Project Goals: Our overall goals of this project are the development of tools for the analysis of networks and pathways in plants and microorganisms related to enable the Systems Biology Knowledgebase proposed by the DOE.

One important task of the future KBase is to provide a platform to help users analyze gene biological function and inspire experiments for the purpose of biofuel development. A gene's biological function is essentially its relationships or interactions with other biological objects within the cell and around the environment. It is necessary to understand gene function on a genomic scale, and from the integration of genetic and cellular networks. We approach this through the prediction and analysis of biological networks, focusing on protein-protein and transcription-factor-target interactions. We describe how these networks can be determined through integration of many genomic features and how they can be analyzed in terms of various topological statistics. In particular, we will report a number of recent analyses: (1) Improving the prediction of molecular networks through systematic training-set expansion; (2) Showing how the analysis of pathways across environments potentially allow them to act as biosensors; (3) Analyzing the structure of the regulatory network which indicates that it has a hierarchical layout with the "middle-managers" acting as information bottlenecks; (4) Showing these middle managers tend be arranged in various "partnership" structures giving the hierarchy a "democratic character"; (5) Comparing the topology and variation of the regulatory network to the call graph of a computer operating system; (6) Developing a framework to integrate together various kinds of biological networks (e.g. relating to TFs and miRNAs) into an integrated meta-network; (7) Integrating this meta-network with actual molecular structures; and (8) Creating practical web-based tools for the analysis of these networks (DynaSIN and tYNA).

We acknowledge funding from the DOE DE-SC0004856 grant.

243

Biological Significance of Gene Modules in an *Arabidopsis thaliana* Co-Expression Network

Darryl Reeves[1]* (darryl.reeves@yale.edu), Gang Fang,[2] and **Mark Gerstein**[1,2,3]

[1]Program in Computational Biology and Bioinformatics, [2]Department of Molecular Biophysics and Biochemistry, and [3]Department of Computer Science, Yale University, New Haven, Conn.

Project Goals: Construction of a biologically meaningful co-expression network for *Arabidopsis thalinia* for use in the Systems Biology Knowledgebase (KBase).

Gene expression is subject to environmental and cell growth conditions. Gene co-expression modules which are associations between gene expression and experimental perturbations and/or cell phenotypes provide important clues for understanding gene biological function. Such modules can assist researchers in the design of additional experiments for identifying favorable genes and metabolic pathways for biofuel development. Thousands of gene expression datasets, mainly derived from microarray experiments, presenting a large range of conditions are publicly available now, and will be integrated into KBase. It is necessary to develop an online toolset to allow users to query, display and make *in silico* analysis of expression data. As one of the major packages for analyzing gene expression, WGCNA[1] (weighted correlation network analysis) will be used to build gene co-expression modules and to cluster cell phenotypes in KBase. In this poster, we present a work-flow showing the steps and options to perform this analysis.

Based on the test run of a number of *Arabidopsis thaliana* datasets, we showed how the choice of tree cutting algorithm affects co-expression module generation and, ultimately, functional annotation of the modules. The clustered co-expression network was grouped into modules using three tree-cutting methods provided by the WGCNA package: static, dynamic, and hybrid. We tested each method and optimized their module generation. The criteria for optimization were how enriched modules are for Gene Ontology (GO) terms and KEGG pathways. The goal is to help users determine biological significance of the modules generated by each set of parameters tested.

Reference

1. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559.

244

Dynamic Metabolic Model Building Based on the Ensemble Modeling Approach

Jimmy G. Lafontaine Rivera[1]* (lafonj@gmail.com), Ali R. Zomorrodi[2]* (zomorrodi@engr.psu.edu), Thomas Wasylenko[3]* (tmwasylenko@gmail.com), **Costas D. Maranas**,[2] **Greg Stephanopoulos**,[3] and **James C. Liao**[1]

[1]Department of Chemical and Biomolecular Engineering, University of California, Los Angeles; [2]Department of Chemical Engineering, The Pennsylvania State University, University Park; [3]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge

Project Goals: The goal of this project is to develop a novel modeling approach to describe the dynamic behavior of metabolic systems (in particular, flux changes upon enzyme tuning) by integrating multiple data platforms including flux, metabolite, transcriptome, and enzyme tuning data. Although the utility of such models is undeniable, their development has been impaired by inadequate modeling approaches, the sheer size of the problem, and difficulties in accessing the intracellular environment. As a result, little progress has been made in realizing such dynamic models despite the continuously increasing number of intracellular measurements that are becoming available by high throughput methods. The resulting models from the proposed research will account for pathway enzyme kinetics and aim to predict the effects of genetic manipulations designed to bring about changes in metabolic flux and overproduction of metabolites, such as tuning various enzyme levels or the Michaelis-Menten constants (K_m) of key enzymes. In this context, such models will be instrumental for constructing microbial strains to produce various biofuels such as ethanol, 1-butanol, and isobutanol from renewable resources. We will use

production of these fuels in *Escherichia coli* as a model system, because of *E. coli*'s central role as a test bed in systems biology, the wealth of kinetic and regulatory information available and its successful usage for the production of biofuels. While the *E. coli* focus will facilitate model development, the approach developed will be general and applicable to other microorganisms and eventually plants. The project is based on the Ensemble Modeling (EM) approach, robust flux and metabolite measurements, and an efficient optimization scheme developed in the PIs' laboratories.

Presently, there are no satisfactory dynamic models of cellular function. This unique deficiency persists despite recent advances in the areas of high throughput measurement of cell-wide intracellular biomolecules and molecular level simulations of various systems. Current approaches for creating dynamic models of cellular function attempt to do so by fitting transient metabolite concentration data to various kinetic rate expressions. These data are difficult to obtain and often have large experimental errors, making it impossible to scale up to the levels required by our current understanding of cellular metabolism. The Ensemble Modeling (EM) framework was proposed to address this issue by relying only on steady-state measurements (although transient measurements can also be used) to create accurate kinetic models of cellular metabolism. Here we introduce the EM framework and show how it can be integrated with robust flux measurement techniques and efficient optimization schemes in order to arrive at such models.

245

Reliable Numerical Methods for FBA and FVA

Yuekai Sun (yuekai@gmail.com)*,[1] Ronan M.T. Fleming,[2] and **Michael A. Saunders**[3]

[1]ICME, Stanford University; [2]Center for Systems Biology, University of Iceland, Iceland; and [3]Department of Management Science and Engineering, Stanford University

http://www.stanford.edu/group/SOL/

Project Goals:

1. **Develop algorithms for solving optimization problems involving large stoichiometric matrices. (a) Extend existing sparse linear programming algorithms to enable the solution of such systems, in which the matrix coefficients represent reactions at multiple timescales and thus vary over many orders of magnitude. (b) 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. (c) Implement a parallel convex optimizer in to enable sampling of the thermodynamically feasible set. (d) Disseminate software to the systems biology community.**
2. **Investigate cyclic dependency between metabolic and macromolecular biosynthetic networks. (a) Predict the material and energy cost of macromolecular synthesis in an integrated metabolic, transcriptional and translational model of *Escherichia coli*. (b) Reconstruct and analyze the macromolecular synthesis network of *Thermotoga maritima*.**
3. **Quantify the significance of thermodynamic constraints on prokaryotic metabolism. (a) Simultaneous prediction of metabolic fluxes and concentrations in *Escherichia coli*. (b) Validate and interpret flux and concentration prediction in *Escherichia coli* and *Thermotoga maritima*. (c) Predict thermodynamically favorable pathways for hydrogen production by *Thermotoga maritima* on a range of substrates. (d) Numerically sample mass conserved, thermodynamically feasible steady state fluxes and concentrations in *Escherichia coli*.**

Concerning project goal 1(a):
Integrated networks of organism metabolism and expression are inherently multi-scale because typical fluxes can vary over eight orders of magnitude. Such networks require special methods to analyze them accurately, and naive use of off-the-shelf optimization software for flux balance analysis (FBA) can produce severely inaccurate solutions. We describe methods for obtaining greater reliability.

The multi-scale nature of integrated networks is also problematic for flux variability analysis (FVA). The traditional FVA formulation sacrifices sharpness in the calculated bounds to ensure feasibility of the sequences of linear programs. In practice the bounds can be off by orders of magnitude. We describe an FVA formulation that guarantees both feasible linear programs and sharp calculated bounds.

246

A Systems Biology Knowledgebase and Analysis Platform for *De Novo* Phenotype Predictions, Integrated Omics Analysis, and Iterative Model Improvement

Edward J. O'Brien[1]* (ejobrien@ucsd.edu), Joshua A. Lerman,[1] Roger L. Chang,[1] Daniel R. Hyduke,[1] Karsten Zengler,[1] Bernhard Ø. Palsson,[1] and **Michael A. Saunders**[2]

[1]Department of Bioengineering, University of California–San Diego, La Jolla and [2]Department of Management Science and Engineering, Stanford University, Stanford, Calif.

http://systemsbiology.ucsd.edu

Project Goals: This project aims to: (1) reconstruct and refine genome-scale models of macromolecular synthesis and metabolism for *Thermotoga maritima* and *Escherichia*

***coli* (2) develop modeling conventions and simulation procedures for phenotype prediction with the integrated models, (3) guide the development of algorithms capable of finding optimal steady-state solutions despite the large, sparse, and ill-scaled constraint matrices, (4) use these prototype models to design a software platform and standard operating procedures to reconstruct general multi-subsystem stoichiometric models, (5) develop and enable methods to parameterize and constrain the network with biophysical models and the direct mapping of diverse omics data, and (6) classify the failure modes of the model to prioritize subsystems and regulatory circuits for model expansion.**

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 metabolic models for both *de novo* phenotype prediction and analysis of omics datasets. To this end, constraint-based algorithms and model-driven omics analysis have provided insight into gene regulation, adaptive evolution, microbial communities, metabolic engineering, and drug response and design. While metabolic models have proven to be a powerful tool, the genetic content of these models is formalized by a Boolean mapping between genetic loci and metabolic reactions, termed the gene-protein-reaction (GPR) relationship. This modeling paradigm only allows for heuristic analysis of transcriptomic and proteomic data and Boolean predictions of genetic requirements.

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 gene Expression). An analogous macromolecular synthesis reconstruction has also been completed for *Thermotoga maritima*, which relied heavily on experimental refinement of the transcription unit architecture. The metabolic and the macromolecular synthesis networks have subsequently been merged into integrated models (termed the 'ME' matrix, for Metabolism and Expression), which allow for explicit analysis and simulation of transcriptomes and proteomes in the context of the underlying reaction network. Not only does inclusion of gene expression increase the scope of the model, but the interdependency of gene expression and metabolism also constrains and refines the metabolic solution space, leading to more accurate predictions.

The ME model formulation additionally 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 to invest its finite resources to synthesize them. This framework is shown to capture known trends in the cellular composition of RNA and protein at various growth rates. Allowing for variable cell composition makes the model more generally applicable to diverse environments whereas the biomass objective function tunes the model to a particular condition.

Our experience reconstructing and analyzing these prototype models made it clear that future model refinement, expansion to other subsystems, omics analysis, and reconstructions for other organisms requires a new software platform and standard operating procedures (SOPs). Analogous software and SOPs exist for metabolic models, but certain features of multi-subsystem models necessitate a redesign of the reconstruction framework including: 1) an order of magnitude larger number of reactants and reactions, 2) many different types of molecules (e.g. rna, proteins, metabolites), 3) 'template' reactions for common cellular processes, 4) efficiency parameters coupling different subsystems, 5) the need for direct integration with growing omics datasets, and 6) future expansions to include other cellular processes and subsystems. Due to these challenges and opportunities, we have developed a database-driven solution to reconstruct, query, and generate these multi-subsystem models. We have designed this software to directly couple with model simulation and analysis software, and have included features and flexibility to address the points listed above.

Future efforts will focus on refining model parameters and constraining condition-specific flux variables with omics datasets and biophysical models. Model improvement will then be mediated by the identification of failure modes to prioritize scope expansions (including signaling and regulatory interactions). We demonstrate the promise for long-term applications of this type of model for metabolic and protein engineering, interpretation of adaptive evolution, and analysis of cellular regulation and optimality.

Grant Information: Numerical Optimization Algorithms and Software for Systems Biology (DOE Award DE-SC0002009)

247

Elucidation of Distinct Transcriptional Regulatory Logic in Bacteria: Applications of a Constraints-Based Systems Biology Knowledgebase

Ali Ebrahim* (aebrahim@ucsd.edu), Stephen A. Federowicz, 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 tran-

scriptional regulatory network with metabolic and macromolecular synthesis models, and (5) provide a platform for genome scale metabolic engineering and synthetic design.

Constructing a systems biology knowledgebase requires the synthesis of a number of critical components into a single platform to allow for diverse computations and analyses. These components include 1) an underlying data model and database 2) tools for integrating, comparing, and analyzing diverse sets of data, and 3) a computational model which mathematically relates the underlying biochemical information. The knowledgebase is built using an iterative workflow: biological experiments are performed to generate data, data is analyzed and its results are integrated into a model, and the model is used to direct future experiments. Here we detail the process of all three steps and display the biological insight gained through the successful utilization of steps one and two.

We first performed an integration and analysis of ChIP-chip, gene expression, and transcription start site (TSS) data obtained at the genome scale for *Escherichia coli*. These specific experiments allow for the elucidation of distinct logical programs and genome scale regulatory mechanisms. These two pieces can then be combined to build a comprehensive model of transcriptional regulation. Logical programs are executed by bacteria in response to common environmental signals or physiological shifts and often include a small molecule signal and associated transcription factor. Initial studies of amino acid metabolism and the transcription factors ArgR and Lrp revealed that arginine and leucine can act as signaling molecules to regulate the transport, biosynthesis, or utilization of 16 amino acids[2]. Similar network motifs governing the flow of an effecter molecule were also shown for purine metabolism and the transcription factor PurR[1]. This has led us to investigate the aerobic-anaerobic shift regulated by ArcA and Fnr, along with the phenomenon of catabolite repression regulated by Crp and Cra. Obtaining a holistic understanding of these systems along with amino acid metabolism allows us to gain an understanding of regulation in response to carbon, nitrogen, and electron acceptor shifts.

In addition to systems level regulation it is possible to gain an understanding of specific regulatory mechanisms at the genome scale in the form of diverse promoter architectures and transcription factor mediated bidirectional transcription. Here we show how promoter architectures occurring in both a unidirectional and bidirectional fashion confer patterns of activation and repression on associated transcription units. Combining this information with systems level principles forges a tie between network and sequence level mechanisms to provide a powerful modeling framework. Utilization of this framework for engineering and synthetic approaches promises to enable a new era of molecular engineering. Overall, the integrated knowledgebase enables a wide range of analysis and is poised to guide future experiments in a model driven fashion towards a comprehensive understanding of transcriptional regulation.

References

1. Cho B-K, Federowicz S, Embree M, Park Y-S, Kim D, Palsson BØ. 2011. The PurR regulon in *Escherichia coli* K-12 MG1655. Nucleic acids research. 39(15):6456-64
2. Cho B-K, Federowicz S, Park Y-S, Zengler K, Palsson BØ. 2011. Deciphering the transcriptional regulatory logic of amino acid metabolism. Nature Chemical Biology. 8(1):65-71

248

Multi-Scale Spatially Distributed Simulations of Microbial Ecosystems

William J. Riehl,[1] William Harcombe,[2] Niels Klitgord,[3] Amrita Kar[1]* (akar@bu.edu), Pankaj Mehta,[4] Nathaniel C. Cady,[5] Christopher J. Marx,[2] and **Daniel Segrè**[1,6]* (dsegre@bu.edu)

[1]Graduate Program in Bioinformatics, Boston University, Boston, Mass.; [2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Mass.; [3]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; [4]Department of Physics, Boston University, Boston, Mass.; [5]College of Nanoscale Science and Engineering, University at Albany, Albany, N.Y.; and [6]Department of Biology and Department of Biomedical Engineering, Boston University, Boston, Mass.

http://comets.bu.edu

Project Goals: The goal of this project is to develop a tool for facilitating simulation, validation and discovery of multiscale dynamical processes in microbial ecosystems. Our Computation Of Microbial Ecosystems in Time and Space (COMETS) is a an open-source platform for performing spatially distributed time-dependent flux balance based simulations of microbial metabolism. Our plan involves building the software platform itself, calibrating and testing it through comparison with experimental data, and integrating simulations and experiments to address important open questions on the evolution and dynamics of cross-feeding interactions between microbial species.

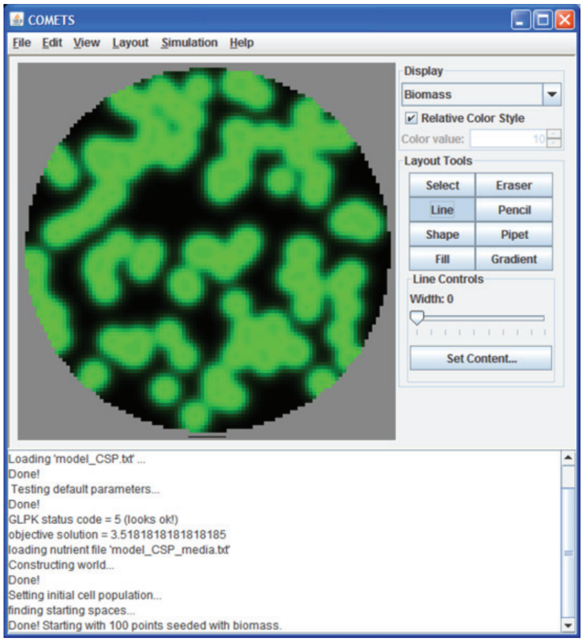
COMETS (Computation Of Microbial Ecosystems in Time and Space) is a broadly applicable and user-friendly platform for modeling metabolic interactions between microbial species. This platform builds on dynamic flux balance analysis (dFBA [1]) to perform time-dependent metabolic simulations of microbial ecosystems, bridging the gap between stoichiometric and environmental modeling. Simulations occur on a spatially structured lattice of interacting metabolic subsystems, representing a level of detail that is intermediate between fine-grained single-cell modeling and a global population modeling approach.

The current version of COMETS incorporates three fundamental steps: *(i)* Implementation of cellular growth (increase of biomass), using a hybrid kinetic-dFBA solver for every point in the 2D lattice. Upper bounds on uptake

* Presenting author

fluxes for the dFBA calculation are estimated based on a concentration-dependent saturating function, in analogy with Michaelis-Menten kinetics. Each grid point may contain biomass for an arbitrary number of different species; *(ii)* Advance of the front of biomass, which we treat as an incompressible fluid (in analogy to [2]). This step involves the solution of the Laplace equation, followed by a calculation of the new biomass front using a level set method; *(iii)* Implementation of a finite differences approximation of the diffusion equation for modeling the diffusion of extracellular small molecules, i.e. environmentally available nutrients and secreted byproducts.

To correctly perform this last part of the simulation, we exploit the natural separation of time scales between growth and diffusion of small molecules. The typical time scale associated with growth, t_{growth}, is set by cell doubling times and is of order 10^3 seconds. The dependence of growth rates on the external nutrients is incorporated by solving a spatially dependent FBA on a time scale $t_{\text{FBA}} << t_{\text{growth}}$, and is typically of order 10^2 seconds. Diffusion is then performed on time scales t_D ~10 seconds, an order of magnitude smaller than the FBA update times. This separation of time scales allows us to efficiently model the complex dynamics in a computationally tractable manner while ensuring that our scheme is physically consistent. An important consequence of this scheme is that it associates a length scale l with each point in our lattice. The scale is set by the smallest diffusion constant, $D_{\min}$, and is given by $l^2 \sim D_{\min}/t_D$. Typically, for small metabolites $D_{\min} \sim 10^{-5}$ cm^{2}s, implying that we have a spatial resolution of about 10^{-2} cm per grid point.



Display: Biomass

☑ Relative Color Style

Color value: 10

Width: 0

```
Loading 'model_CSP.txt' ...
Done!
 Testing default parameters...
Done!
GLPK status code = 5 (looks ok!)
objective solution = 3.5181818181818185
loading nutrient file 'model_CSP_media.txt'
Constructing world...
Done!
Setting initial cell population...
finding starting spaces...
Done! Starting with 100 points seeded with biomass.
```

Our prototype of COMETS uses the open-source GNU Linear Programming Kit (GLPK) for performing the dFBA calculations, and a Java platform for coordinating the simulations and for rapid visualization. Several microbial species have been imported into COMETS, including *Escherichia coli*, *Salmonella typhimurium*, *Shewanella oneidensis*, *Lactococcus lactis*, and *Saccharomyces cerevisiae*. New species and new environmental settings can be easily incorporated through a macro language, or using a custom graphical user interface (see Figure). COMETS is being tested by comparing computational simulations to available and newly measured spatial distributions of biomass in individual *E. coli* colonies on agar [3]. In addition, we are applying this platform to study the population dynamics of two syntrophic bacteria [4], with special attention to the effects of initial density in a diverse spatially organized system that contains a population of cheaters.

References

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.
2. Xavier J, Martinez-Garcia E, and Foster KR. *Social evolution of spatial patterns in bacterial biofilms: when conflict drives disorder.* Am Nat 2009, Jul;174(1):1-12.
3. Pipe LZ and Grimson MJ. *Spatial-temporal modeling of bacterial colony growth on solid media.* Mol. BioSyst. 2008, Mar;4(3):192-198.
4. Harcombe W. *Novel cooperation experimentally evolved between species.* Evolution, Jul;64(7):2166-72.

249

A Multi-Scale Approach to the Simulation of Lignocellulosic Biomass

Goundla Srinivas,[1] Glass Dennis,[1] Sergiy Markutsya,[2] Yana Kholod,[2] Ajitha Devarajan,[2] John Baluyut,[2] Monica Lamm,[2] Theresa L. Windus,[2] Xiaolin Cheng[1,3]* (chengx@ornl.gov), **Mark S. Gordon**,[2] and **Jeremy C. Smith**[1,3]

[1]Oak Ridge National Laboratory, Oak Ridge, Tenn.;
[2]Ames Laboratory, Iowa State University, Ames; and
[3]University of Tennessee, Knoxville

http://cmb.ornl.gov/

Project Goals: In concert with the imminent increase in the Department of Energy's leadership supercomputing power to the petaflop range, the objective of this project is to develop multiscale methods for extending the time- and length-scales accessible to biomolecular simulation on massively parallel supercomputers. This project also aims to apply the developed multiscale approaches to obtain an understanding of the structure, dynamics and degradation pathways of extended cellulosic and lignocellulosic materials. Information from multiscale simulation, when closely integrated with experiment, will provide fundamental understanding needed to overcome biomass recalcitrance to hydrolysis.

The multiscale simulation methods, ranging from highly accurate quantum mechanical (QM) methods to coarse-grained molecular dynamics (MD), have been used to obtain an understanding of the structure, dynamics and degradation pathways of extended cellulosic and lignocellulosic materials using capability high-performance simulation. Treating solvent implicitly is a critical multiscale concept, and to

* Presenting author

this end we have developed a parallel order-*N* Poisson-Boltzmann solver (1,2) and a treecode-based Generalized Born electrostatic solvation method (3). Furthermore, a statistical mechanical multiscale approach was derived that was found to describe the temperature dependence of cellulose fiber stability (7), and complementary Fragment Molecular Orbital (FMO) and all-atom MD simulations have been performed of cellulose crystal structures.

A range of coarse-grained models have been developed on various length scales (8). Based on the effective fragment potential, coarse-grained models have been developed for benzene and methanol and for glucose in solution. The coarse graining has also involved development and application of Boltzmann inversion techniques and of the "REACH" (Realistic Extension Algorithm via Covariance Hessian) methodology, which maps results obtained from atomistic MD simulations onto models for larger-scale, coarse-grained MD.

The physical properties of lignocellulosic biomass derived using the multiscale methodologies serve as a basis for interpreting an array of biophysical experiments (4-5), and, in particular, the simulation models derived will be used to calculate and interpret a variety of neutron-scattering properties. To aid in the interpretation we have developed "dynamical fingerprinting" as a means of reconciling the multiple time scales accessed by experiment and simulation (6). This combination of simulation and experiment will eventually lead to a description of the physicochemical mechanisms of biomass recalcitrance to hydrolysis, and thus will aid in developing a strategy as to how rationally to overcome the resistance.

References

1. Lu B*, Cheng X*, Huang J, McCammon JA. AFMPB: An Adaptive Fast Multipole Poisson-Boltzmann Solver for Calculating Electrostatics in Biomolecular Systems. *Comput Phys Commun.* 2010, 181, 1150
2. Zhang B, Lu B, Cheng X, Huang J*, Pitsianis N, Sun X, and McCammon JA. Mathematical and Numerical Aspects of the Adaptive Fast Multipole Poisson-Boltzmann Solver. *Communications in Computational Physics* 2011, in press
3. Xu Z*, Cheng X* and Yang H. Treecode-based Generalized Born method. *J Chem Phys.* 2011 134(6):064107
4. Prinz, JH, Chodera, J, Pande V., Swope W, Smith JC and Noe F. Optimal use of data in parallel tempering simulations for the construction of discrete-state Markov models of biomolecular dynamics. *J. Chem. Phys.* 2011, 134, 244108
5. JH Prinz, M. Held, JC Smith and F. Noe. Efficient Computation, Sensitivity and Error Analysis of Committor Probabilities for Complex Dynamical Processes. *Multiscale Mod. & Sim.* 9, 545-567 (2011).
6. F. Noe, S. Doose, I. Daidone, M. Lollmann, M. Sauer, J.D. Chodera and JC Smith. Dynamical Fingerprints: Probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments. *Proc. Natl. Acad. Sci. U.S.A.* 2011, 108(12):4822-7
7. Klein H*, Cheng X*, Smith JC* and Shen T*. Transfer matrix approach to the hydrogen-bonding in cellulose Iα fibrils describes the recalcitrance to thermal deconstruction. *J Chem Phys.* 2011, 135, 085106
8. Srinivas G, Cheng X and Smith JC*. A Solvent-Free Coarse Grain Model for Crystalline and Amorphous Cellulose Fibrils. *J. Chem. Theo. Comp.* 2011, 7 (8), 2539–2548

This research is funded by the Genomic Science Program, Office of Biological and Environmental Research, and the Scientific Discovery through Advanced Computing program, U. S. Department of Energy, currently under FWP ERKJE84. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 and of the National Center for Computational Sciences at Oak Ridge National Laboratory, which is supported by a DOE INCITE award from the Office of Science of the U.S. Department of Energy.

250

A Multiscale Approach to the Simulation of Lignocellusic Biomass

Monica H. Lamm[1,2]* (mhlamm@iastate.edu), John Baluyut,[1,2] Ajitha Devarajan,[1] Yana Kholod,[1] Sergiy Markutsya,[1] Theresa L. Windus,[1,2] and **Mark S. Gordon**[1,2]

[1]Ames Laboratory, Ames, Iowa and [2]Iowa State University, Ames

Project Goals: Multiscale methods in theoretical chemistry and molecular simulation are applied to obtain an understanding of the structure, dynamics, and degradation pathways of cellulosic biomass. Using methods that range from accurate quantum chemical calculations to all-atom and coarse-grained molecular dynamics simulation, this project aims to use theory and simulation as a guide for overcoming the recalcitrance to hydrolysis in the production of fuel from biomass.

Cellulosic ethanol production is a two-stage process that involves the hydrolysis of cellulose to form simple sugars and the fermentation of these sugars to ethanol. Hydrolysis of cellulose is the rate-limiting step, and there is a great need to characterize the process with theoretical chemistry and molecular simulations to better understand the complex mechanisms that are involved. The ultimate goal is to generate accurate coarse-grained molecular models that are capable of predicting the structure of lignocellulose before and after pretreatment so that subsequent *ab initio* calculations can be performed to probe the degradation pathways.

Current computational studies include: 1) determining the energy barrier to rotation of free hydroxyl groups in cellulose Iα, 2) characterizing the interfragment and interchain interaction energies in cellulose Iβ with fragment molecular orbital (FMO) calculations of, 3) developing coarse-grained models for crystalline and amorphous cellulose fibers from all-atom molecular dynamics simulations using classical force fields and the effective fragment potential (EFP).

This research is supported under FWP AL-08-330-039 by the Genomic Science Research Program, Office of Biological and Environmental R esearch and the Scientific Discovery through Advanced Computing program in the U.S. Department of Energy Office of Science.

251

Combining Whole Cell Stochastic Simulations with Systems Biology Approaches

Elijah Roberts,[1] Piyush Labhsetwar,[2] John Cole,[3] Nathan Price,[1,2] and **Zaida Luthey-Schulten**[1,2,3]* (zan@illinois.edu)

[1]School of Chemical Sciences, [2]Center for Biophysics and Computational Biology, and [3]Department of Physics, University of Illinois, Urbana

In this work we compute the stochastic reaction-diffusion dynamics of selected biochemical pathways to show how individual cells vary expression of a set of genes in response to an environmental signal. The whole cells simulated under *in vivo* conditions include ribosomes, DNA, and large protein complexes which take up 30-50% of the cell volume and are placed according to data from cryoelectron tomography and proteomics. Using GPU processors, we simulate the dynamics for an entire cell cycle and compare the mRNA/protein distributions to those observed in single molecule experiments. We show how such distributions can be used to derive additional kinetic parameters and integrate effects of cell to cell variations into flux balance analysis of genome scale models of metabolic networks. The distribution of growth rates calculated for a colony of bacteria are analyzed and correlated to changes in fluxes through the metabolic network.

Publications

1. "Long time-scale simulations of *in vivo* diffusion using GPU hardware", E. Roberts, J. Stone, L. Sepulveda, Wen-mei Hwu, and Z. Luthey-Schulten, in *Proceedings of the IEEE International Symposium on Parallel and Distributed Processing*, 2009.
2. "Noise contributions in an inducible genetic switch: A whole cell simulation study", E. Roberts, A. Magis, J. Ortiz, W. Baumeister, and Z. Luthey-Schulten, *Plos Comput. Biol.* 7(3), e1002010 March (2011).
3. "Determining the stability of genetic switches: Explicitly accounting for mRNA noise" Michael Assaf, Elijah Roberts, and Zaida Luthey-Schulten, *Phys.Rev.Lett.* 24, 248102, (2011).

252

Toolbox Model of Evolution of Metabolic Pathways on Networks of Arbitrary Topology

Tin Yau Pang[1,2] and **Sergei Maslov**[1]* (maslov@bnl.gov)

[1]Biology Department, Brookhaven National Laboratory, Upton, N.Y. and [2]Department of Physics and Astronomy, Stony Brook University, N.Y.

Project Goals: The biological functioning of a living cell involves coordinated actions of its metabolic and regulatory networks. Metabolic networks are composed of many semi-autonomous functional units – metabolic pathways. These pathways are routinely controlled by dedicated transcription factors, and the activities of individual pathways need to be well coordinated with each other. The project goal is to investigate general principles behind such coordination in prokaryotic genomes. To this end we carry out dynamical and evolutionary modeling of the integrated network encompassing metabolic and regulatory interactions.

It has been previously reported[1] for prokaryotic genomes that the number of Transcription Factors (TFs) is proportional to the *square* of the total number of genes. As a consequence of this trend the fraction of TFs (the so-called "regulatory overhead") is less than 0.5% in small (<500 genes) bacterial genomes, while in large genomes (~10,000 genes) it can be as high as 10%. We recently proposed[2] a general explanation of this empirical scaling law and illustrated it using a simple model in which metabolic and regulatory networks co-evolve together. In our model prokaryotic organisms acquire new metabolic functions by the virtue of horizontal gene transfer of entire co-regulated metabolic pathways from a shared gene pool (the "universal metabolic network" or bacterial metabolic pan-genome). This transfer is followed by removal of redundant enzymes and assignment of a dedicated TF regulating the newly acquired pathway[3]. The whole process can be compared to a homeowner buying sets of tools from a hardware store and later returning duplicate items. We view the full repertoire of metabolic enzymes (or more generally all non-regulatory proteins) encoded in the genome of an organism as its collection of tools. Adapting to a new environmental condition (e.g. learning to utilize a new nutrient source) involves acquiring new tools as well as reusing some of the tools that are already encoded in the genome. As the toolbox of an organism grows larger, it can reuse its existing tools more often and thus needs to acquire fewer new enzymes to master each new functional task. From this argument it follows that, in general, the number of metabolic pathways and their regulators should always scale faster than linearly with the total number of genes in a genome. The empirically observed quadratic scaling between these two numbers can be mathematically derived for a broad range of universal network topologies[4]. Furthermore, the sizes of evolutionary conserved pathways in our model have a long-tailed power-law distribution that agrees with empirical observations. This offers a conceptual explanation for the empirically observed broad distribution of regulon sizes or TFs out-degrees in regulatory networks.

References

1. E van Nimwegen, "Scaling laws in the functional content of genomes", Trends Genet 19, 479-84 2003
2. S Maslov, S Krishna, T Y Pang, K Sneppen, "Toolbox model of evolution of prokaryotic metabolic networks and their regulation", PNAS 106, 9743-9748 2009
3. J Grilli, B Bassetti, S Maslov, and M Cosentino Lagomarsino, "Joint scaling laws in functional and evolutionary categories in prokaryotic genomes", Nucleic Acids Research doi: 10.1093/nar/gkr711 2011

4. TY Pang, S Maslov, "Toolbox model of evolution of metabolic pathways on networks of arbitrary topology" PLoS Comp. Bio 8, e1001137 2011.

Support of this work was provided by the DOE Systems Biology Knowledgebase project "Tools and Models for Integrating Multiple Cellular Networks."

253

Enabling the Use of Externally-Built Alignments and Trees in ARB for Evolutionary Analysis

Steve Essinger,[1] Erin Reichenberger,[1] **Chris Blackwood**,[2] and **Gail Rosen**[1]* (gailr@ece.drexel.edu)

[1]Electrical and Computer Engineering Department, Drexel University and [2]Biology Department, Kent State University

In order to investigate gene evolution, gene sequences from various organisms are commonly aligned to form a phylogenetic tree. Besides viewing the taxonomic information on the tree, a user may want to visually inspect how the gene product and KEGG pathway with the associated sequence has evolved, giving greater power to evolutionary hypothesis testing. A software package, such as ARB, has the power to pool this information from Genbank records, but ARB uses the local computer resources to perform the alignment. Therefore, a user may want to use external resources (such as the CIPRES portal on Terragrid), to perform the alignment and tree construction, and then import and link that information back into ARB to manipulate the data.

To accomplish this, we have created a pipeline that integrates external alignment and de novo tree construction for an arbitrary protein family (even one that contains over 10,000 member sequences). We have developed custom python scripts and an ARB import filter to extract meta-data from Genbank records and import this info with an externally-built alignment and phylogenetic tree. Using our scripts, a custom database, that includes all of the sequences and associated meta-data in the study, is imported into an ARB database using uniqueIDs. The user can then use the ARB suite of tools to manipulate the phylogenetic tree and display the associated metadata.

We demonstrate the use of our tool by examining a protein family of interest to the "Tracking down the cheaters" project. All code will be made available on our website that will allow other groups to view custom fields extracted from Genbank records on phylogenetic trees using externally-built trees and alignments.

254

Numerical Optimization Algorithms and Software for Systems Biology: A Globally Convergent Algorithm for Computing Stable Non-Equilibrium Steady State Concentrations in Genome-Scale Networks

Ronan M.T. Fleming[1]* (ronan.mt.fleming@gmail.com), Ines Thiele,[1] and **Michael A. Saunders**[2]

[1]Center for Systems Biology, University of Iceland, Iceland and [2]Department of Management Science and Engineering, Stanford University

http://www.hi.is/~rfleming

Project Goals: 1. Develop algorithms for solving optimization problems involving large stoichiometric matrices. (a) Extend existing sparse linear programming algorithms to enable the solution of such systems, in which the matrix coefficients represent reactions at multiple timescales and thus vary over many orders of magnitude. (b) 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. (c) Implement a parallel convex optimizer in to enable sampling of the thermodynamically feasible set. (d) Disseminate software to the systems biology community. 2. Investigate cyclic dependency between metabolic and macromolecular biosynthetic networks. (a) Predict the material and energy cost of macromolecular synthesis in an integrated metabolic, transcriptional and translational model of *Escherichia coli*. (b) Reconstruct and analyze the macromolecular synthesis network of *Thermotoga maritima*. 3. Quantify the significance of thermodynamic constraints on prokaryotic metabolism. (a) Simultaneous prediction of metabolic fluxes and concentrations in *Escherichia coli*. (b) Validate and interpret flux and concentration prediction in *Escherichia coli* and *Thermotoga maritima*. (c) Predict thermodynamically favorable pathways for hydrogen production by *Thermotoga maritima* on a range of substrates. (d) Numerically sample mass conserved, thermodynamically feasible steady state fluxes and concentrations in *Escherichia coli*.

Concerning project goal 3(a): At the core of computational systems biology lies a paradox. All of the currently available genome-scale modeling methods can only model chemical reaction rates, but not the abundance (or concentration) of the molecules involved in these reactions. At the same time, the vast majority of experimental omics data are measures of the abundance of some molecule, rather than the rate. The reason for this paradox is that modeling steady state reaction kinetics has been limited to small systems of chemical reactions as the inherently nonlinear systems of equations at the core of such models have been intractable to solve. We present the first globally convergent algorithm for simultaneously computing stable non-equilibrium steady state molecule concentrations and reaction rates. We leverage this algorithm to simultaneously predict stable steady state concentrations and reaction rates in *E. coli*, which are numerical consequences of various hypotheses regarding the manner in which evolved kinetic parameters may be optimal with respect to optimization of various cellular system objectives.

submitted post-press

259

Development of Predictive Software Tools to Construct and Analyze Large Dynamical Networks for Systems Biology Knowledgebase

Ravishankar R. Vallabhajosyula* (rrv@cfdrc.com), B. Prabhakarpandian, and Kapil Pant

CFD Research Corporation, Huntsville, Ala.

http://www.cfdrc.com

Project Goals: Recent biotechnological advances have accelerated the generation of 'omic' data. This has driven the development of computational tools to model the biological systems by inferring mechanisms responsible for response to external stimuli. However, lack of kinetic information for most biochemical interactions limits the predictive capabilities of these tools. CFD Research Corporation (CFDRC) is developing predictive modeling toolkit to overcome this limitation, thereby facilitating rapid and accurate characterization of the effects of the environment on phenotypes. In particular, our toolkit will enable (1) identification of significant biological features from omic datasets, (2) construction of a comprehensive network model of cellular pathways, and (3) simulation of this pathway model using a kinetics-free algorithm to predict the altered phenotypes when selected targets in the network are modified. This methodology is being validated using well-characterized organisms (e.g., yeast) as well as selected microbe-based biosystems of DoE interest (e.g., identification of higher quantity and quality biofuel yielding algal strains).

Recent developments in genetic engineering and biotechnology have enabled the modification of genes in an organism or the introduction of genes from other organisms towards achieving the desired phenotypes. However, these experimental procedures are often carried out without adequate systems-level knowledge of the cellular biology, which can lead to unexpected outcomes. Well-designed computational methodologies can be used to prevent such scenarios with the aid of predictive software tools. A key goal of the DoE Systems Biology Knowledgebase (Kbase) is to facilitate analysis of vast omic datasets for characterizing the response of organisms to various environmental stimuli towards predicting phenotypes. For example, such tools will be able to identify algal strains with improved attributes of biofuel production, while simultaneously overcoming slow growth rates associated with some of these strains. Such computational approaches should be based on a comprehensive understanding of the cellular biology of the organisms of interest, and will be significantly aided by the adaptation and application of novel algorithms and software that can analyze multi-omic data related to the observed response to various external stimuli.

Under DOE sponsored research, CFD Research Corporation (CFDRC) is currently developing predictive computational tools to address the goals of Kbase towards characterizing the response biological organisms to environmental stimuli that serve as inputs and predicting phenotypes most likely to be observed. Figure 1 shows a schematic of the framework being developed by CFDRC. Drawing upon available databases, our approach relies on the construction of mechanistic Systems Biology based and data-driven models of the differentially regulated cellular pathways. The complex pathway models are then analyzed without requiring information on the kinetics of various biochemical interactions. This enables the discovery and ranking of targets (for example genes, proteins or metabolites) for potential modification and the prediction of their response when these modifications are implemented. This approach thus offers the potential to inform experiments for the development of strains efficient at generating the desired phenotype such as algae strains that can produce biofuels at a higher rate.

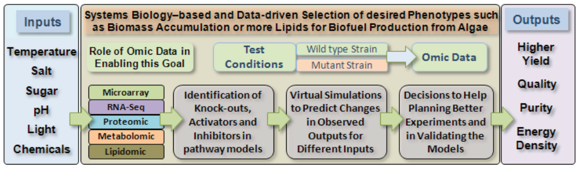

Inputs
Temperature
Salt
Sugar
pH
Light
Chemicals
Systems Biology–based and Data-driven Selection of desired Phenotypes such as Biomass Accumulation or more Lipids for Biofuel Production from Algae
Role of Omic Data in Enabling this Goal
Test Conditions
Wild type Strain
Mutant Strain
Omic Data
Microarray
RNA-Seq
Proteomic
Metabolomic
Lipidomic
Identification of Knock-outs, Activators and Inhibitors in pathway models
Virtual Simulations to Predict Changes in Observed Outputs for Different Inputs
Decisions to Help Planning Better Experiments and in Validating the Models
Outputs
Higher Yield
Quality
Purity
Energy Density

Figure 1: Schematic Detailing the use of Omic Data to Identify Targets towards Predicting Phenotypes

As part of the ongoing Phase I study, we are developing a prototype of the software toolkit using transcriptional data to construct and analyze complex pathway networks in an extensible SBML format (Hucka et al., 2003) that will be enhanced to analyze other omic data types in future. Development of these tools will enable researchers to analyze pathways that play important roles in sensing and responding to the external conditions in an integrated manner. These tools are important to understand the organism's behavior in the modified environment including its survival and in predicting the associated phenotypes. Towards this goal, we are studying the yeast environmental stress response to various external conditions as a test case to test and validate the model. We are also in active discussions with different organizations to demonstrate the technology for microbial systems of DoE interest e.g., identification of targets for genetic engineering of algal strains for higher quantity and quality biofuel production.

References

1. Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., et al., The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network models, *Bioinformatics*, **19**(4): 524-531, 2003.

* Presenting author

This work is being supported through the DOE Office of Biological and Environmental Research under an SBIR Phase I grant (DE-SC0006190).

submitted post-press

262

The GreenCut Resource, a Phylogenomically Derived Inventory of Proteins Specific To the Plant Lineage

Rikard Fristedt,[1] HiroakiYamasaki,[1] Steven Karpowicz,[1] Arthur Grossman,[2] and **Sabeeha Merchant**[1]* (merchant@chem.ucla.edu)

[1]Department of Chemistry and Biochemistry and Institute for Genomics and Proteomics, University of California, Los Angeles and [2]Department of Plant Biology, Carnegie Institution for Science, Stanford, Calif.

http://www.chem.ucla.edu/dept/Faculty/merchant.html

Project Goals: (see abstract)

The plastid is a defining structure of photosynthetic eukaryotes and houses many plant specific processes, including the light reactions, carbon fixation, pigment synthesis, and other primary metabolic processes. Identifying proteins associated with catalytic, structural, and regulatory functions that are unique to plastid-containing organisms is necessary to fully define the scope of plant biochemistry. We performed phylogenomics on 20 genomes to compile a new inventory of 597 nucleus-encoded proteins conserved in plants and green algae but not in non-photosynthetic organisms. At the time of analysis, 286 of these proteins were of known function, whereas 311 are not characterized. This inventory was validated as applicable and relevant to diverse photosynthetic eukaryotes using an additional eight genomes from distantly related plants (including Micromonas, Selaginella, and soybean). Manual curation of the known proteins in the inventory established its importance to plastid biochemistry. To predict functions for the 52% of proteins of unknown function, we used sequence motifs, subcellular localization, co-expression analysis, and RNA abundance data. About 18% of the proteins in the inventory have functions outside the plastid and/or beyond green tissues. Although 32% of proteins in the inventory have homologs in all cyanobacteria, unexpectedly, 30% are eukaryote-specific. Finally, 8% of the proteins of unknown function share no similarity to any characterized protein and are plant lineage-specific. We have initiated functional analyses of the eukaryote-specific proteins and we present phenotypes for loss of function mutations in some of the unknown GreenCut genes.

Communication and Ethical, Legal, and Societal Issues

255

ELSI Pilot: Assessing and Mitigating the Risks of Large-Scale Metabolic Engineering

J. Geller,[1] S. Lee,[1] D. Tarjan,[1] C. Wu,[1] T. Torok,[1] T.C. Hazen,[1] A.P. Arkin,[1,2,3] and **N.J. Hillson**[1,2]* (njhillson@lbl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]Joint BioEnergy Institute, Emeryville, Calif.; and [3]University of California, Berkeley

Project Goals: Establish methodologies for assessing and mitigating the risks of future large-scale metabolic engineering microbial projects, including those extending beyond the bioreactor.

The DOE EERE-funded Advanced Biofuel Process Development Unit (ABPDU) houses two 300 liter microbial fermentation tanks. In the extremely unlikely event that the ABPDU's post-fermentation microbicidal protocol (e.g. base treatment and neutralization) should catastrophically fail, broth harboring viable genetically engineered micro-organisms could be purged directly to downstream waste-water treatment processes. This Ethical, Legal and Social Implications (ELSI) pilot study seeks to quantitate the risks associated with this scenario, by measuring the viability of the engineered microbes (and perhaps more importantly, their embedded genes) in mock sewage reactors that mimic the conditions and microbial communities found in real-world waste water treatment plants. Furthermore, investigation of how differential genetic backgrounds (e.g. gene deletions) impact survival and gene transmission to sewage sludge communities will guide subsequent forward-engineering efforts to further reduce risk. This pilot study establishes methodologies (leveraging only recently available technologies) for assessing and mitigating the risks of future large-scale metabolic engineering microbial projects, including those extending beyond the bioreactor.

256

Managing the Risks of Synthetic Biology: Assessing the U.S. Regulatory System

Sarah R. Carter[1]* (scarter@jcvi.org), Michael Rodemeyer,[2] and **Robert Friedman**[3]

[1]J. Craig Venter Institute, Rockville, Md.; [2]University of Virginia, Charlottesville; and [3]J. Craig Venter Institute, San Diego, Calif.

Project Goals: (see below)

The Coordinated Framework for the Regulation of Biotechnology was established in 1986 as a "comprehensive federal regulatory policy for ensuring the safety of biotechnology research and products." (51 FR 23302) This framework has evolved over time (CEQ/OSTP, 2001), both as the Federal government has gained experience with biotechnology products and as the technology has advanced. However, with the advent of synthetic biology and other new technologies, new questions arise about the applicability of these rules and regulations to future biotechnology products (Rodemeyer, 2009). Synthetic biology refers to a set of techniques that together provide scientists and engineers with far greater capabilities to modify organisms than current techniques allow. The term "synthetic" comes from the relatively new ability to synthesize long pieces of DNA from chemicals, increasing both the power and precision of genetic engineering. Both the departure from older genetic engineering techniques and the broader type and scale of genetic changes may create challenges for the regulatory system.

The goal of this project is to assess how well the current Federal regulatory framework for biotechnology applies to the anticipated products of synthetic biology, and to provide options for addressing any gaps or shortcomings. This will include an analysis of the authorities that are used by regulatory agencies (primarily USDA, EPA, and FDA) as well as the risk assessment challenges that the agencies are likely to face. This is a two-year project that includes two workshops as well as multiple consultations with experts both within and outside the Federal government. The final report should be available by late 2012.

The first workshop will be held in January, 2012, and will focus on assessing the regulatory framework for likely synthetic biology products based on a case study approach. The four product case studies will be: cyanobacteria and microalgae for biofuel production; microbes for chemical production or for bioremediation; microbes for use as drugs or cosmetics; and modified plants for use as alcohol-fuel feedstock. By bringing together outside experts and Federal regulators, we hope to get a better understanding of the agencies' regulatory authorities, their capabilities to perform risk assessments, and where any gaps in the regulatory framework may occur.

References

1. CEQ/OSTP Assessment: Case Studies of Environmental Regulation for Biotechnology (2001). 66 FR 7905.
2. OSTP: Coordinated Framework for Regulation of Biotechnology (1986). 51 FR 23302.
3. Rodemeyer M. (2009). New Life, Old Bottles: Regulating First-Generation Products of Synthetic Biology. Washington, D.C.: Woodrow Wilson International Center for Scholars.

* Presenting author

This project is funded by the Biological and Environmental Research program within the Office of Science in the Department of Energy.

257

Choices and Challenges in Translating Science and Technology from Concepts to Realities

Amy K. Wolfe[1]* (ami@ornl.gov), David J. Bjornstad,[1,2] W. Christopher Lenhardt,[1] Barry Shumpert,[1] and Stephanie Wang[1]

[1]Oak Ridge National Laboratory and [2]University of Tennessee

Knowledge generated by modern science and technology (S&T) must be "translated" from individual ideas and discoveries to organized production blueprints and business plans if the fruits of knowledge generation are to lead to socially valuable products and processes. Whether by institutional design or the efforts of individual actors, this translation process must support a confluence of inputs—science, technology, technology transfer and entrepreneurship activities—and take place at a number of stages during the science-to-product chain. This poster describes our preliminary research findings into the process by which large S&T centers can develop programs and procedures to help convert the scientific and technical concepts they develop into the uses that entrepreneurs can nurture for societal benefit.

Our research is beginning to identify the implications of different ways that S&T institutions organize the conduct of science, both for how research is conducted and for research translation. A number of frameworks have been used to describe how information and products flow from laboratory toward use. We draw elements from these frameworks and from data we have collected to propose a different conception that we term "ushering." Ushering entails purposeful actions designed to move the information and products of S&T toward use, effectively extending the point at which scientists or organizations typically consider their work "done." Thus, institutions purposefully would create an environment that not only facilitates, but expects flows of information to occur within and between organizations. Within the organization, information flows may link fundamental to basic activities, basic to fundamental activities, or may involve other activities, such as technology transfer. Outside the organization, information flows may provide assistance to downstream entrepreneurs or organizations that help them gain access to needed scientific and technical knowledge in an organized manner or to other types of information exchange.

The scientists, science managers and administrators, and personnel involved in intellectual property, technology transfer, and commercialization we have queried typically agree on the importance of translation. However, they also recognize that their participation in the ushering process is governed by a variety of internal business practices and external drivers. These practices and drivers can provide information, mechanisms, and incentives to participate in the translation process, but they can also create barriers that discourage participation. Our inquiries have identified instances where practices created to achieve other goals have the effect, perhaps inadvertently, of competing or conflicting with research translation goals.

The research described in this poster is being carried out by the ORNL Ethical, Legal, and Social Issues Scientific Focus Area. It focuses, in part, on identifying the components of and resources needed for the ushering process and on analyzing the implications of alternative organizational rules and practices for science and research translation. Data for our translation work are drawn from structured discussions with individuals and from a day-long workshop involving 26 individuals from diverse areas of science and technology, management, and technology transfer. Our initial project efforts are focused on translating S&T toward use, within and beyond the community of scientists. Later work will expand the current set of ushering topics into broader types of support, mechanisms for providing this support, and experiences at other S&T institutions.

258

Biological and Environmental Research Information System: A Multifaceted Approach to DOE Systems Research Communication

Jennifer L. Bownas, Kris S. Christen, Holly L. Haun, Brett Hopwood, 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: Develop and distribute programmatic materials to help build the multidisciplinary community needed to advance systems research for DOE energy and environmental missions. The Biological and Environmental Research Information System group works with program managers and the scientific community to help develop and communicate key scientific and technical concepts for scientific community and public discourse. Ideas are welcome to extend program integration and improve communications 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 Department of Energy (DOE) research projects, progress, and significance to science and society; and (3) respond to outreach and information exchange needs of related DOE projects.

The Biological and Environmental Research Information System has focused on presenting all facets of genomics research for DOE's Office of Science (22 years). 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 exhibits at more than 10 scientific meetings each year and maintain the searchable BER Research Highlights database (public.ornl.gov/hgmis/bernews/). We also assist with the outreach efforts of DOE grantees—especially the Bioenergy Research Centers, Joint Genome Institute, Environmental Molecular Sciences Laboratory, and Atmospheric Radiation Measurement Climate Research Facility—to help increase their reach and impact.

Biological Systems Science Division—completed and ongoing projects include:

- Genomic Science program website http://genomicscience.energy.gov
- DOE Genomic Science Awardee Meeting X, February 26–29, 2012 (this abstracts book)
- Applications of New DOE National User Facilities in Biology Workshop Report (February 2012)
- Switchgrass Research Group: Progress Report (January 2012)
- Biosystems Design: Draft Report from the July 2011 Workshop (Web HTML, January 2012)
- Plant Feedstock Genomics for Bioenergy Joint Awards (August 2011)
- Revealing the Role of Microbial Communities in Carbon Cycling (July 2011)
- Projects Underpinning Knowledgebase Development (May 2011)
- Biological Systems Science Division Overview (revised May 2011)
- Genomic Science Program brochure (May 2011)
- DOE User Facilities: Advanced Technologies for Biology, Structural Biology brochure (May 2011)
- DOE BER Joint Genome Institute brochure (revised May 2011)
- Joint Meeting 2011: Genomic Science Awardee Meeting IX and USDA-DOE Plant Feedstock Genomics for Bioenergy Awardee Meeting, April 10–13, 2011, abstracts book, 238 pp., April 2011

Climate and Environmental Sciences Division (CESD)—completed and ongoing projects include:

- Subsurface Biogeochemical Research website (in development)
- Subsurface Biogeochemical Research Contractor-Grantee Workshop, April 30–May 2, 2012, abstracts book (in development)
- Subsurface Biogeochemical Research brochure (October 2011)
- Terrestrial Ecosystem Science brochure (October 2011)
- Climate and Environmental Sciences Division Overview (October 2011)
- GOAmazon2014 Workshop Report summary brochure (October 2011)
- GOAmazon2014 Workshop Report (September 2011)
- DOE BER Environmental Molecular Sciences Laboratory overview brochure (revised May 2011)
- DOE BER ARM Climate Research Facility overview brochure (revised May 2011)

We also continuously update and enhance websites, paying particular attention to navigation and increasing functionality and accessibility. These sites include:

- Genomic Science website (genomicscience.energy.gov). In addition to describing program research, the site provides information on how to access DOE user facilities and the DOE Systems Biology Knowledgebase.
- BER image gallery (public.ornl.gov/site/gallery/)
- BER Research Highlights Database (public.ornl.gov/site/bernews/)
- Subsurface Biogeochemical Research (in development)

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.

Appendix 1: Participants

As of March 8, 2012

Paul Adams
Lawrence Berkeley Laboratory

Michael Adams
University of Georgia

Dawn Adin
U.S. Department of Energy

Andrew Allen
J. Craig Venter Institute

Steven Allison
University of California, Irvine

Daniel Amador-Noguez
Princeton University

Gordon Anderson
Battelle - Biological Separations & Mass Spectrometry

Todd Anderson
U.S. Department of Energy

Maciek Antoniewicz
University of Delaware

Dionysios Antonopoulos
Argonne National Laboratory

Ludmilla Aristilde
Princeton University

Adam Arkin
Lawrence Berkeley National Laboratory

Emma Aronson
University of California, Irvine

Erin Baker
Battelle - Biological Separations & Mass Spectrometry

Scott Baker
EMSL, Pacific Northwest National Laboratory

Nitin Baliga
Institute for Systems Biology

Anindita Banerjee
Washington University in Saint Louis

Jill Banfield
University of California

Richard Baran
Lawrence Berkeley National Laboratory

Paul Bayer
U.S. Department of Energy

Alexander Beliaev
Pacific Northwest National Laboratory

Kelly Bender
Southern Illinois University

Matthew Benedict
University of Illinois at Urbana-Champaign

David Benton
University of Wisconsin - Madison, Great Lakes Bioenergy Res

Bertram Berla
Washington University in St. Louis

Renaud Berlemont
University of California, Irvine

Sean Berthrong
Cornell University

Dave Bjornstad
Oak Ridge National Laboratory

Christopher Blackwood
Kent State University

Robert Blake, II
Xavier University of Louisiana

Jeffrey Blanchard
University of Massachusetts, Amherst

Hans Blaschek
University of Illinois

Sara Blumer-Schuette
North Carolina State University

Paul Bohn
University of Notre Dame

Gregory Bokinsky
Lawrence Berkeley National Laboratory / Joint BioEnergy Inst

Harvey Bolton
Pacific Northwest National Laboratory

Yannick Bomble
National Renewable Energy Laboratory

Daniel Bond
University of Minnesota

Gregory Bonito
Duke University

Adam Book
Great Lakes Bioenergy Research Center (GLBRC)

Brian Bothner
Montana State University

Peter Bottomley
Oregon State University

Jeremy Bougoure
Lawrence Livermore National Laboratory

Ben Bowen
Lawrence Berkeley National Laboratory

James Brainard
National Renewable Energy Laboratory

Thomas Brettin
Oak Ridge National Laboratory

Eoin Brodie
Lawrence Berkeley National Laboratory

Yury Bukhman
Great Lakes Bioenergy Research Center, University of Wiscons

Robert Burnap
Oklahoma State University

Gareth Butland
Lawrence Berkeley National Laboratory

Stephen Callister
Pacific Northwest National Laboratory

Ashley Campbell
Cornell University

Allison Campbell
Pacific Northwest National Laboratory

William Cannon
Pacific Northwest National Laboratory

Shane Canon
Lawrence Berkeley National Laboratory

Sarah Carter
J. Craig Venter Institute

Ron Caspi
SRI International

Javier A. Ceja-Navarro
Lawrence Berkeley National Laboratory

Benli Chai
Michigan State University

John-Marc Chandonia
Berkeley National Laboratory

Kent Chapman
University of North Texas

Jay Chen
Oak Ridge National Laboratory

Xiaoliang Cheng
Lawrence Berkeley National Laboratory / Joint BioEnergy Inst

Xiaolin Cheng
Oak Ridge National Laboratory

Gang Cheng
Sandia National Laboratories

Jer-Ming Chia
Cold Spring Harbor Laboratory

Vincent Chiang
North Carolina State University

Dylan Chivian
Lawrence Berkeley National Laboratory

Karuna Chourey
Oak Ridge National Laboratory

Lon Chubiz
Harvard University

James Cole
Michigan State University

Dean Cole
U.S. Department of Energy

Frank Collart
Argonne National Laboratory

James Collins
Boston University

Cameron Cotten
University of Wisconsin-Madison / Great Lakes Bioenergy Rese

Robert Cottingham
Oak Ridge National Laboratory

Jennifer Couch
National Institutes of Health

Michael Crowley
National Renewable Energy Laboratory

Leland Cseke
University of Alabama in Huntsville

Alan Darvill
University of Georgia/ BioEnergy Research Center (BESC)

Maude David
Lawrence Berkeley National Laboratory

Mark Davis
National Renewable Energy Laboratory

Brian Davison
Oak Ridge National Laboratory

Paramvir Dehal
Lawrence Berkeley National Laboratory

Anne Dekas
California Institute of Technology

Adam Deutschbauer
Lawrence Berkeley National Lab

Patrik D'haeseleer
Lawrence Livermore National Laboratory

Stephen DiFazio
West Virginia University

Shi-You Ding
National Renewable Energy Laboratory

Gregory Ditzler
Drexel University

Mitchel Doktycz
Oak Ridge National Laboratory

Timothy Donohue
Great Lakes Bioenergy Research Center (GLBRC)

Sarah Douglas
Harvard University

Daniel Drell
U.S. Department of Energy

Inna Dubchak
Lawrence Berkeley National Laboratory

David Duncan
Great Lakes Bioenergy Research Center (GLBRC)

Mary Dunlop
University of Vermont

Ali Ebrahim
University of California, San Diego

Joseph Ecker
The Salk Institute

Sarah Eisenlord
University of Michigan

Dwayne Elias
Oak Ridge National Laboratory

James Elkins
Oak Ridge National Laboratory

Marwa El-Sayed
North Carolina State University

Steve Essinger
Drexel University

Aymerick Eudes
Joint BioEnergy Institute

Barbara Evans
Oak Ridge National Laboratory

Gang Fang
Yale University

Adam Feist
Univeristy of California, San Diego

Richard Ferrieri
Brookhaven National Laboratory

Matthew Fields
Montana State University

Mary Firestone
University of California, Berkeley

Zoë Fisher
Los Alamos National Laboratory

Ronan Fleming
University of Iceland

Barry Flinn
Institute for Advanced Learning and Research

Jason Flowers
University of Washington

Ellen Ford
Joint BioEnergy Institute (JBEI)

Cheryl Foust
Oak Ridge National Laboratory

Marvin Frazier
The Institute of Theoretical Science

Jim Fredrickson
Pacific Northwest National Laboratory

Robert Friedman
j. Craig Venter Institute

Andrei Gabrielian
National Institutes of Health

Leslie Galloway
University of Tennessee

David Gang
Washington State University

Hector Garcia Martin
Lawrence Berkeley National Laboratory / Joint BioEnergy Inst

Aaron Garoutte
Michigan State University

George Garrity
NamesforLife, LLC / Michigan State University

Mark Gerstein
Yale University

Maria Ghirardi
National Renewable Energy Laboratory

Richard Giannone
Oak Ridge National Laboratory

Jack Gilbert
Argonne National Laboratory

David Gilbert
Joint Genome Institute (JGI)

Paul Gilna
Oak Ridge National Laboratory / BESC

Carol Giometti
Argonne National Laboratory

Ee-Been Goh
Lawrence Berkeley National Laboratory / Joint BioEnergy Inst

Amita Gorur
Lawrence Berkeley National Laboratory

Joseph Graber
U.S. Department of Energy

Robin Graham
Argonne National Laboratory

Jean Greenberg
University of Chicago

Susan Gregurick
U.S. Department of Energy

Sol Gruner
Cornell University

Robert Gunsalus
University of California, Los Angeles

Adam Guss
Oak Ridge National Laboratory

Masood Hadi
Sandia National Laboratories

Rembrandt Haft
University of Wisconsin -Madison / Great Lakes Bioenergy Rese

Michael Hahn
University of Georgia / BioEnergy Science Center (BESC)

Kenneth Hammel
University of Wisconsin

William Harcombe
Harvard University

Samuel Hazen
University of Massachusetts

Terry Hazen
University of Tennessee/Oak Ridge National Laboratory

Shaomei He
Lawrence Berkeley National Laboratory

Zhili He
University of Oklahoma

Joshua Heazlewood
Lawrence Berkeley National Laboratory

Eric Hegg
Michigan State University

Richard Heins
Joint BioEnergy Institute (JBEI)

Christopher Henry
Argonne National Laboratory

Robert Hettich
Oak Ridge National Laboratory

Meytal Higgins
Princeton University

Nathan Hillson
Lawrence Berkeley National Laboratory / Joint BioEnergy Inst

Mike Himmel
National Renewable Energy Laboratory

Roland Hirsch
U.S. Department of Energy

Suzanne Hodgkins
Florida State University

Keith Hodgson
SLAC National Accelerator Laboratory

Susan Holladay
Oak Ridge National Laboratory

Hoi-Ying Holman
Lawrence Berkeley National Laboratory

John Houghton
U.S. Department of Energy

Adina Howe
Michigan State University

Bo Hu
University of Washington

Greg Hurst
Oak Ridge National Laboratory

William Inskeep
Montana State University

Roberto Izaurralde
Pacific Northwest National Laboratory

Dina Jabbour
Michigan State University-DOE Plant Research Laboratory

Janet Jansson
Lawrence Berkeley National Laboratory

Pouya Javidpour
Joint BioEnergy Institute (JBEI)

Thomas Jeffries
Great Lakes Bioenergy Research Center (GLBRC)

Jerry Jenkins
HudsonAlpha Institute

Yongqin Jiao
Lawrence Livermore National Laboratory

Yong-Su Jin
University of Illinois at Urbana-Champaign

Andrzej Joachimiak
Argonne National Laboratory

Shawn Jones
University of Delaware

Nicholas Justice
University of California, Berkeley

Justin Kaffenberger
University of Minnesota

Edward Kaleikau
U.S. Department of Agriculture

Amrita Kar
Boston University

Ulas Karaoz
Lawrence Berkeley National Laboratory

Peter Karp
SRI International

Tatiana Karpinets
Oak Ridge National Laboratory

Arthur Katz
U.S. Department of Energy

Alexey Kazakov
Lawrence Berkeley National Laboratory

Jay Keasling
Lawrence Berkeley National Laboratory / Joint BioEnergy Inst

Kevin Keegan
Argonne National Laboratory

Kenneth Keegstra
Michigan State University / Great Lakes Bioenergy Research

Martin Keller
Oak Ridge National Laboratory

Kimberly Keller
University of Missouri

Robert Kelly
North Carolina State University

Ken Kemner
Argonne National Laboratory

Seonhwa Kim
Institute for Advanced Learning and Research

Won Chan Kim
Michigan State University

Young-Mo Kim
Pacific Northwest National Laboratory

Christine Kirmaier
Washington University in St Louis

Daniel Klein-Marcuschamer
Joint BioEnergy Institute (JBEI)

Niels Klitgord
Lawrence Berkeley National Laboratory

Allan Konopka
Pacific Northwest National Laboratory

Konstantinos Konstantinidis
Georgia Institute of Technology

Wayne Kontur
University of Wisconsin-Madison

Galina Kovaleva
Lawrence Berkeley National Laboratory

Helen Kreuzer
Pacific Northwest National Laboratory

Akhil Kumar
Pennsylvania State University

Rajeev Kumar
University of California, Riverside

Cheryl Kuske
Los Alamos National Laboratory

Deborah Kuzmanovic
Grinnell College

Nikos Kyrpides
Lawrence Berkeley National Laboratory

Jimmy Lafontaine Rivera
University of California, Los Angeles

Philip Laible
Argonne National Laboratory

K. V. Lakshmi
Rensselaer Polytechnic Institute

Monica Lamm
Iowa State University

Kemin Lan
Drexel University

Robert Landick
University of Wisconsin - Madison / Great Lakes Bioenergy R

Paul Langan
Oak Ridge National Laboratory

Alejandra Lara-Chavez
The Institute for Advanced Learning and Research

Peter Larsen
Argonne National Laboratory

Mario Latendresse
SRI International

Mark Le Gros
Lawrence Berkeley National Laboratory

Taek Soon Lee
Lawrence Berkeley National Laboratory

Kimberly Lemmer
University of Wisconsin-Madison / Great Lakes Bioenergy Rese

David Lesmes
U.S. Department of Energy

Caleb Levar
University of Minnesota

Semen Leyn
Sanford–Burnham Medical Research Institute

James Liao
University of California

Mary Lidstrom
University of Washington

Po-Heng Lin
University of California

Stephen Lindemann
Pacific Northwest National Laboratory

Mary Lipton
Pacific Northwest National Laboratory

Hanbin Liu
Joint BioEnergy Institute

Frank Loeffler
University of Tennessee / Oak Ridge National Laboratory

Dominique Loque
Lawrence Berkeley National Laboratory / Joint BioEnergy Inst

Zaida Luthey Schulten
University of Illinois

Pin-Ching Maness
National Renewable Energy Laboratory

Reinhold Mann
Brookhaven National Laboratory

Betty Mansfield
Oak Ridge National Laboratory

Costas Maranas
Pennsylvania State University

Christopher Marx
Harvard University

Sergei Maslov
Brookhaven National Laboratory

Lakshmi Matukumalli
National Institute of Food and Agriculture / USDA

Konstantinos Mavrommatis
Joint Genome Institute (JGI)

Ryan McAndrew
Joint BioEnergy Institute (JBEI)

Gwendolyn McClung
U.S. Environmental Protection Agency

Lee Ann McCue
Pacific Northwest National Laboratory

Ann McDermott
Columbia University

Shawn McGlynn
California Institute of Technology

Gail McLean
U.S. Department of Energy

Darcy McRose
Monterey Bay Aquarium Research Institute

David Mead
Lucigen

John Meeks
University of California, Davis

Chuansheng Mei
Institute for Advanced Learning and Research

Matthew Melnicki
Pacific Northwest National Laboratory

Sabeeha Merchant
University of California, Los Angeles

Barbara Methe
J. Craig Venter Institute

Noelle Metting
U.S. Department of Energy

Thomas Metz
Pacific Northwest National Laboratory

Folker Meyer
Argonne National Laboratory

Birte Meyer
University of Washington

Jonathan Mielenz
Oak Ridge National Laboratory

Lisa Miller
Brookhaven National Laboratory

Benjamin Miller
Great Lakes Bioenergy Research Center (GLBRC)

Jeffrey H. Miller
University of California, Los Angeles

Marissa Mills
Oak Ridge National Laboratory

Caroline Milne
University of Illinois

Todd Mockler
Donald Danforth Plant Science Center

Debra Mohnen
University of Georgia - Complex Carbohydrate Research Center

Rhiannon Mondav
University of Queensland

James Moran
Pacific Northwest National Laboratory

Annika Mosier
University of California, Berkeley

Wellington Muchero
Oak Ridge National Laboratory

Thomas Mueller
Pennsylvania State University

Arijit Mukherjee
University of Wisconsin, Madison

Aindrila Mukhopadhyay
Lawrence Berkeley National Laboratory

Biswarup Mukhopadhyay
Virginia Bioinformatics Institute, Virginia Tech

Thomas Murray
The Hastings Center

David Myrold
Oregon State University

Harish Nagarajan
University of California, San Diego

Leandro Neves
University of Florida

Lee Newberg
Wadsworth Center, New York State Department of Health

B Tracy Nixon
Pennsylvania State University

Angela Norbeck
Environmental Molecular Sciences Laboratory, PNNL

Trent Northen
Lawrence Berkeley Laboratory

Jaspreet Notey
North Carolina State University

Pavel Novichkov
Lawrence Berkeley National Laboratory

Mari Nyyssonen
Lawrence Berkeley National Laboratory

Edward O'Brien
University of California--San Diego

Paula Olsiewski
Alfred P. Sloan Foundation

Daniel Olson
Dartmouth College

Hugh O'Neill
Oak Ridge National Laboratory

Tullis Onstott
Princeton University

Victoria Orphan
California Institute of Technology

Kanchana Padmanabhan
North Carolina State University

Himadri Pakrasi
Washington University in St. Louis; I-CARES

Chongle Pan
Oak Ridge National Laboratory

Eleftherios (Terry) Papoutsakis
University of Delaware

Byung Park
Oak Ridge National Laboratory

Junyoung Park
Princeton University

Sara Patterson
University of Wisconsin

Dale Pelletier
Oak Ridge National Laboratory

Loukas Petridis
Oak Ridge National Laboratory

Jennifer Pett-Ridge
Lawrence Livermore National Laboratory

Brian Pfleger
University of Wisconsin-Madison

Tommy Phelps
Oak Ridge National Laboratory

George Phillips
University of Wisconsin-Madison

Madeleine Pincu
University of California, Irvine

Sai Venkatesh Pingali
Oak Ridge National Laboratory

Mircea Podar
University of Tennessee Knoxville and Oak Ridge National Lab

Nathan Price
Institute for Systems Biology

Endang Purwantini
Virginia Bioinformatics Institute, Virginia Tech, USA

Wei-Jun Qian
Pacific Northwest National Laboratory

Pablo Rabinowicz
U.S. Department of Energy

Arthur Ragauskas
Georgia Institute of Technolgy

Priya Ranjan
University of Tennessee

Dmitry Ravcheev
Sanford-Burnham Medical Research Institute

Kevin Redding
Arizona State University

Erin Reichenberger
Drexel University

David Reiss
Institute for Systems Biology

David Rejeski
Woodrow Wilson International Center for Scholars

Ruth Richardson
Cornell University

Kathryn Richmond
Great Lakes Bioenergy Research Center

Wayne Robinson
Creative Discovery Museum

Errol Robinson
Environmental Molecular Sciences Laboratory (EMSL)

Dmitry Rodionov
Sanford-Burnham Medical Research Institute

Robin Rohwer
Pennsylvania State University

Margaret Romine
Pacific Northwest National Laboratory

Cathy Ronning
U.S. Department of Energy

Jocelyn Rose
Cornell University

Gail Rosen
Drexel University

Amelia-Elena Rotaru
University of Massachusetts Amherst

Edward Rubin
Lawrence Berkeley Laboratory / Joint Genome Institute (JGI)

Rajib Saha
The Pennsylvania State University

Jennifer Salazar
Argonne National Laboratory

Bushra Samad
University of California Berkeley

Robert Sanford
University of Illinois at Urbana-Champaign

Sanjaya Sanjaya
Great Lakes Bioenergy Research Center

Vinay Satish
Lawrence Berkeley National Laboratory

Trey Sato
University of Wisconsin / Great Lakes Bioenergy Research Cen

Chris Schadt
Oak Ridge National Laboratory

Amy Schaefer
University of Washington

Henrik Scheller
Lawrence Berkeley National Laboratory / Joint BioEnergy Inst

Jonathan Schilling
University of Minnesota

Jeremy Schmutz
Joint Genome Institute (JGI) - HudsonAlpha

Israel Scott
University of Georgia

Daniel Segre
Boston University

Rajandeep Sekhon
University of Wisconsin-Madison / Great Lakes Bioenergy Rese

Jeremy Semrau
The University of Michigan

Margrethe Serres
Marine Biological Laboratory

Joao Setubal
University of Sao Paulo

Manoj K. Sharma
Joint BioEnergy Institute (JBEI)

Maxim Shatsky
Lawrence Berkeley National Laboratory

Henry Shaw
Lawrence Livermore National Laboratory

Hui Shen
The Noble Foundation

Louis Sherman
Purdue University

Shengjing Shi
University of California, Berkeley

Pravin Shrestha
University of Massachusetts, Amherst

Blake Simmons
Joint BioEnergy Institute (JBEI)

David Sims
JGI - HudsonAlpha Institute for Biotechnology

Steven Singer
Lawrence Berkeley National Laboratory

Seema Singh
lawrence Berkeley National Laboratory

Raghuveer Singh
University of Nebraska-Lincoln

Gary Siuzdak
The Scripps Research Institute

Kimmen Sjolander
University of California Berkeley

David Skinner
Lawrence Berkeley National Laboratory

Steve Slater
Great Lakes Bioenergy Research Center

Mark Smith
Massachusetts Institute of Technology

Jeremy Smith
Oak Ridge National Laboratory

Richard Smith
Pacific Northwest National Laboratory

Michael Souza
University of California, Berkeley

Thomas Squier
Pacific Northwest National Laboratory

Prem Srivastava
U.S. Department of Energy

Robert Stack
U.S. Department of Energy

David Stahl
University of Washington

Joshua Steele
California Institute of Technology

Rick Stevens
Argonne National Laboratory/ University of Chicago

Marvin Stodolsky
formerly U.S. Department of Energy

Yuekai Sun
Institute for Computational and Mathematical Engineering

Patrick Suthers
Pennsylvania State University

Mustafa Syed
Oak Ridge National Laboratory

Taichi Takasuka
Great Lakes Bioenergy Research Center / University of Wiscon

Li Tan
University of Georgia - Complex Carbohydrate Research Center

Tracy Teal
Michigan State University

Michael Teresinski
U.S. Department of Energy

Astrid Terry
Lawrence Berkeley National Laboratory

Michael Thelen
Lawrence Livermore National Laboratory

Brian Thomas
University of California, Berkeley

David Thomassen
U.S. Department of Energy

James Tiedje
Great Lakes Bioenergy Research Center/Michigan State Univers

Ming Tien
Pennsylvania State University

Crystal Tobin
Argonne National Laboratory

Trudy Torto-Alalibo
Virginia Tech

Andrey Tovchigrechko
J. Craig Venter Institute

Michael Travers
SRI International

Susannah Tringe
Joint Genome Institute (JGI)

Timothy Tschaplinski
Oak Ridge National Laboratory

Serdar Turkarslan
Institute for Systems Biology

Gerald Tuskan
Oak Ridge National Laboratory

Gene Tyson
The University of Queensland

Edward Uberbacher
Oak Ridge National Laboratory

Clifford Unkefer
Los Alamos National Laboratory

Breeanna Urbanowicz
University of Georgia

Sagar Utturkar
University of Tennessee

Brian Vaccaro
University of Georgia

Ravishankar Vallabhajosyula
CFD Research Corporation

Marijke van Baren
Monterey Bay Aquarium Research Institute

Patanjali Varanasi
Joint BioEnergy Institute (JBEI)

Fernando Vega
U.S. Department of Agriculture

Miguel Vega-Sanchez
Joint Bioenergy Institute (JBEI)

Nathan VerBerkmoes
Oak Ridge National Laboratory

Alex Verseput
Pennsylvania State University

Rytas Vilgalys
Duke University

Peter Walian
Lawrence Berkeley National Laboratory

Judy Wall
University of Missouri

Yun Wang
Northwestern University

Zeng-Yu Wang
The Noble Foundation

David Ward
Montana State University

Doreen Ware
Cold Spring Harbor Laboratory

Thomas Wasylenko
Massachusetts Institute of Technology

Sharlene Weatherwax
U.S. Department of Energy

Peter Weber
Lawrence Livermore National Laboratory

Drew Weisenberger
Thomas Jefferson National Accelerator Facility

David Weston
Oak Ridge National Laboratory

Janet Westpheling
University of Georgia

Elizabeth White
U.S. Department of Energy

Derrick White
University of Nebraska-Lincoln

Steven Wiley
Pacific Northwest National Laboratory

Halina Witkowski
University of California, San Francisco

Amy Wolfe
Oak Ridge National Laboratory

Aaron Wright
Pacific Northwest National Laboratory

Cathy Wu
University of Delaware

Liyou Wu
University of Oklahoma

J.H. David Wu
University of Rochester

Yifan Xu
Princeton University

Xiaohan Yang
Oak Ridge National Laboratory

Haw Yang
Princeton University

Song Yang
University of Washington

Steven Yannone
Berkeley National Laboratory

Kelsey Yee
Oak Ridge National Laboratory

Suzan Yilmaz
Sandia National Laboratories

Yanbin Yin
University of Georgia

Lydia Zeglin
Oregon State University

Karsten Zengler
University of California, San Diego

Sarah Zerbs
Argonne National Laboratory

Yaoping Zhang
University of Wisconsin-Madison / Great Lakes Bioenergy Rese

Jizhong Zhou
University of Oklahoma

Yimin Zhu
University of Wisconsin-Madison

Zhuoyun Zhuang
Boston University

Angela Ziebell
National Renewable Energy Laboratory

Ali Zomorrodi
Pennsylvania State University

Jeffrey Zurawski
North Carolina State University

Appendix 2: Websites

Genomic Science Program Websites

- Genomic Science program http://genomicscience.energy.gov
- Office of Biological and Environmental Research Genomic Science http://science.energy.gov/ber/research/bssd/genomic-science/
- This book http://genomicscience.energy.gov/pubs/2012abstracts/

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

- Jackson Lab Grassland Ecology
  http://agronomy.wisc.edu/jackson
- BESC KnowledgeBase
  http://besckb.ornl.gov
- BioCyc Database Collection
  http://biocyc.org
- Center for Advanced BioEnergy Research (CABER)
  http://bioenergy.illinois.edu
- Pathway Tools Information Site
  http://brg.ai.sri.com/ptools
- Center for Molecular Biophysics
  http://cmb.ornl.gov/
- Computation of Microbial Ecosystems in Time and Space (COMETS)
  http://comets.bu.edu
- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA)
  http://enigma.lbl.gov/
    - ENIGMA and Hydrogen Regulation
      http://baliga.systemsbiology.net/enigma
- FunGene Functonal Gene Pipeline and Repository
  http://fungene.cme.msu.edu/
- Genegrabber
  http://genegrabber.berkeley.edu/
- KBASE DOE Systems Biology Knowledgebase
  http://kbase.us/
    - KBase Blog http://outreach.kbase.us
- PhyloFacts
  http://makana.berkeley.edu/phylofacts/
- Microbial ENergy processes Gene Ontology Project (MENGO)
  http://mengo.vbi.vt.edu
- MetaCyc
  http://metacyc.org
- MicrobesOnline
  http://microbesonline.org
- Plant-Microbe Interfaces
  http://PMI.ornl.gov
- Ribosomal Database Project
  http://rdp.cme.msu.edu/
- RegPrecise
  http://regprecise.lbl.gov/
- NamesforLife
  http://www.namesforlife.com

Author Index

Indexed by Page Number

A

Abraham, Paul 39
Adams, Michael W.W. 43, 75, 167, 170, 171
Adams, Paul D. 3, 7, 8, 9, 12, 13, 153, 155, 156, 157, 158, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172
Adams, Rachel 39
Adams, Rachel M. 150
Adkins, Joshua N. 117, 118
Agnew, Daniel E. 26
Alahuhta, Markus 46
Allahyari, Mehdi 33
Allan, J. 85
Allen, Andrew E. 107, 108
Allen, Eric E. 107
Allen, Lisa Zeigler 107
Allen, Sara M. 150
Allen, Simon 153, 156, 171
Allgaier, Martin 9
Allison, Steven D. 87, 88
Alm, Eric J. 159, 161, 162
Alsin, Fredrik 125
Alves, Pedro 195
Amador-Noguez, Daniel 49, 50
Anderson, Brian J. 118
Anderson, Gordon A. 117, 118, 119, 136
Ané, Jean-Michel 28
Ansong, Charles 100
Antoniewicz, Maciek R. 50
Arbeleaz, P. 157
Arbing, M. 104
Argyros, Aaron 45
Aristilde, Ludmilla 50
Arkin, Adam A. 153, 155
Arkin, Adam P. 9, 11, 13, 14, 154, 157, 158, 159, 160, 161, 162, 164, 165, 166, 167, 168, 169, 170, 171, 172, 175, 176, 177, 178, 179, 180, 187, 207
Aronson, Emma L. 87
Arora, Rohit 6
Arp, Daniel 110
Aryal, Uma K. 69, 121
Asano, Keiji G. 150
Atamodjo, Melani 37
Auberry, Kenneth J. 117
Auer, Manfred 5, 7, 153, 157

B

Badger, Jonathan H. 107
Baidoo, Edward 168
Baker, Erin S. 118
Baker, John O. 36
Baker, Scott E. 121
Balan, Venkatesh 22, 24
Balasch, Monica Moya 95
Baliga, Nitin S. 155, 157, 158, 159, 160, 161, 167, 168
Ball, D.A. 157
Baluyut, John 200, 201
Balzergue, Sandrine 4
Bandow, Nathan 112
Bandyopadhyay, Anindita 69
Banerjee, Goutami 23, 24
Banfield, Jillian F. 79, 80, 180, 190
Baran, Richard 153, 155
Bare, J. Christopher 158
Barnard, Romain 93
Barnes, Stephen 151
Bar-Peled, Maor 32, 33
Barry, Kerrie 114, 115
Bartley, Laura 4
Bastian, Caleb 50
Basu, Rahul 180
Bates, John T. 158
Batth, Tanveer 12
Baumohl, Jason 158, 161, 180
Baxter, D.J. 122
Beam, Jacob 136, 137
Bebout, Brad M. 68
Beck, David A.C. 48, 99
Begemann, Matthew B. 26
Beliaev, Alexander S. 64, 65, 71, 130
Beller, Harry R. 13
Belnap, Jayne 96
Belov, Mikhail E. 117
Bender, Kelly 167
Benedict, Matthew N. 104
Benke, Peter I. 4
Bennett, P. 85
Benning, Christoph 27
Benton, David 180
Berim, Anna 126
Berla, Bertram M. 70
Berlemont, Renaud 47
Bernstein, Hans 137
Best, Aaron A. 189, 194
Bhardwaj, Nitin 195
Bianchetti, Christopher M. 24
Bickford, J. 99
Bigelow, Diana J. 64, 130
Biggin, Mark D. 153, 156, 157, 167, 168, 171
Bing, Jap 168
Binkowski, Andrew 173
Biswas, Ranjita 45
Bjornstad, David J. 208
Blaby-Haas, Crysten E. 77
Blackwell, Meredith M. 67
Blackwood, Christopher B. 103, 203
Blake II, Robert C. 114
Blanchard, Jeffrey L. 51
Blanchette, Robert A. 52
Blanch, Harvey W. 3, 9
Blaschek, Hans P. 52
Blumer-Schuette, Sara E. 43
Blum, P.H. 73
Bohn, P.W. 138
Bokinski, Greg 11
Bomble, Yannick J. 36, 46
Bond, Daniel R. 124
Bonito, Gregory M. 143, 147
Book, Adam J. 16, 17
Borglin, Sharon 10
Borglin, Sharon E. 162, 166
Bothfeld, William 22
Bothner, Brian 138
Bottomley, Peter J. 90, 92, 110
Boucher, N. 73
Bougoure, Jeremy 93, 94
Boundy-Mills, Kyria 24
Bowen, Benjamin 8, 79, 80, 153, 155, 160, 177, 180
Bownas, Jennifer L. 208
Boyum, Julie 21
Bracho-Garrillo, Rosvel 81, 83
Brady, John W. 46
Brenner, Steven E. 156, 160, 161, 171, 177
Brettin, Thomas 179, 180
Brewer, Heather M. 117
Brileya, Kristen A. 169
Bristow, Jim 114, 115
Brodie, Eoin L. 67, 88, 93, 94
Brown, C. Titus 17, 18, 181
Browning, James 7
Brown, Page 7
Brown, Roslyn 118, 136, 137
Brown, Steven D. 44, 75, 144, 145, 146, 162, 164
Brown, Titus 92

Brown, Virginia 38
Brumm, Philip 21
Brunecky, Roman 46
Brutnell, Thomas 63
Bryant, Donald A. 64, 129, 133, 136
Buckley, Daniel H. 80
Buell, Robin 16
Bukhman, Yury 24, 180
Burnum, Kristin E. 117, 118
Burow, Luke C. 68
Butland, Gareth P. 153, 154, 168, 171

C

Cady, Nathaniel C. 199
Callister, Stephen J. 69, 99, 100, 113, 117, 118
Campbell, Ashley N. 80
Campbell, Elsie 72
Campbell, James H. 164
Cannon, William R. 122, 129
Canon, Shane 179
Cao, Peijian 4
Cardenas, Erick 18
Carlile, Melissa 34
Carlson, Ross 137
Carroll, Andrew 4
Carter, Sarah R. 207
Caspi, Ron 194
Castro, Hector F. 143, 146
Ceja-Navarro, Javier A. 67
Cha, Min 45
Chabout, Salem 4
Chadwick, Grayson 102
Chai, Benli 181
Chain, Patrick 18, 92
Chakraborty, Romy 162, 166, 172
Chan, R. 9
Chandonia, John-Marc 154, 156, 157, 161, 171, 177
Chang, Changsoo 173
Chang, Christopher H. 33
Chang, Jeff 110
Chang, Roger L. 197
Chanton, Jeffrey 84, 86
Chapman, Jarrod 114
Chapman, Kent D. 59
Chee-Sanford, Joanne 123
Chen, Baowei 130
Chen, Fang 14, 40
Chen, J. 10
Chen, Jay 140, 143
Chen, Mo 46
Chen, Z. 9
Cheng, Chao 195
Cheng, Gang 6, 7
Cheng, K. 85
Cheng, Lei 81, 83
Cheng, Xiaoliang 8
Cheng, Xiaolin 200
Chhabra, Swapnil R. 9, 11, 12, 153, 154, 157, 167, 168, 171, 172
Chia, Jer-ming 178
Chiang, Vincent 58
Childs, Kevin 16
Chistoserdova, L. 99
Chiu, Wan-Ling 72
Chivian, Dylan 9, 11, 14, 175, 176, 177, 180
Cho, Byung-Kwan 198
Cho, Kevin 155
Choi, Megan 171
Chou, Katherine 76
Chou, Wen-Chi 32
Chourey, Karuna 85, 102, 143, 150
Christen, Kris S. 208
Christman, Harry 72
Chundawat, Shishir P.S. 38
Chung, Daehwan 45
Ciesielski, Peter N. 36
Clarke, Declan 195
Clark, Nivedita 103
Clauss, Therese R.W. 117, 118, 136
Cole, James R. 18, 81, 83, 92, 181
Cole, John 202
Collart, Frank 110, 134, 139, 151, 173
Conrad, Mark 166
Copié, Valérie 138
Corrier, Kristen 92
Cotten, Cameron 22, 26
Cottingham, Robert W. 43, 175, 176, 177, 178, 179
Crill, Patrick M. 84, 86
Crowell, Kevin L. 117, 118
Crowley, Michael F. 36, 46
Cseke, Leland J. 139, 151
Csencsits, R. 157
Cuff, Marianne 173
Curlevski, Nathalie J.A. 80
Currie, Cameron R. 16, 17

D

Dai, Ziyu 121
Dale, Bruce E. 22, 24, 38
Danielson III, William F. 118
da Rocha, Ulisses Nunes 94
Darvill, Alan 37
Datta, Ruchira 185
Datta, Suchitra 64
Datta, Supratim 7
David, Maude M. 90, 92
Davis, Mark 34, 40
Davison, Brian H. 30, 31, 32, 35, 42, 60, 61, 175
DeAngelis, Kristen 10
DeBolt, Seth 63
Decker, Steve 14, 34
Deerinck, Tomas J. 102
DeGiovanni, Andy 3, 13
Dehal, Paramvir S. 158, 161, 166, 167, 168, 172, 175, 176, 177
DeJongh, Matt 194
Dekas, Anne E. 101
DeLeon, Kara 162
De Leon, K.B. 163
de Leon, Natalia 16
DeMartini, Jaclyn D. 38
Deng, Kai 8
Deng, Ye 81, 83, 165, 184
Deng, Yu 45
Dennis, Glass 200
Densmore, D. 10
Desai, Narayan 176, 179
Deshpande, Shweta 29
Deutschbauer, Adam M. 75, 153, 154, 155, 156, 157, 167, 168, 172
Devarajan, Ajitha 200, 201
Deviod, Scott 179
Devoid, Scott 194
D'haeseleer, Patrik 9, 10
Di, Yannming 110
Dietrich, Jeffrey 11
DiFazio, Steve P. 38
Ding, Shi-You 36, 62, 63
DiSpirito, Alan A. 112
Disz, Terrence 179, 194
Dixon, Richard A. 14, 40, 42
Dmytriv, Z. 10
Doeppke, Crissa 34
Dohnalkova, Alice 131, 137
Doktycz, Mitchel J. 140, 141, 142, 143, 144, 145, 146, 147, 149, 150
Dong, Ming 168, 171
Donohoe, Bryon S. 36
Donohue, Timothy J. 29
Doucette, Christopher D. 49
Downing, Kenneth H. 153, 157
Drake, Meghan M. 164
Driscoll, C.M. 138
Dubchak, Inna 159, 161, 187
Dubey, Manish 7
Dubini, Alexandra 68
Ducoste, Joel 58
Dunbar, John 95, 96
Duncan, David 20
Duncan, Shona M. 52

Dunlop, Mary J. 47
Dunn, John 116
Dupont, Chris L. 108
Dupont, Christopher L. 107, 108
Durrett, Timothy 27
Dyer, John M. 59
Dykstra, Andrew 39

E

Eberhard, Stefan 37
Ebert, Berit 5
Ebrahim, Ali 198
Ecker, Joseph R. 57
Edwards, Ivan P. 89
Edwards, Rob 194
Eisenlord, Sarah D. 89
Elbourne, Liam D.H. 107
Elias, Dwayne A. 162, 164, 167, 168, 170
Elkins, James G. 43, 44
Ellinger, James 22
Elliott, Nicholas 168
Elliott, S. 85
Ellisman, Mark H. 101, 102
El-Sayed, M. 73
Elsbree, N. 10
Elsen, Nathaniel L. 24
Elvitigala, Thanura 70
Embree, Mallory 98
Emory, Joshua F. 43
Engelhardt, Barbara E. 160
Engle, Nancy L. 42, 43, 44, 149
Entwistle, Elizabeth E. 89
Esker, Alan 7
Essinger, Steve 203
Eudes, Aymerick 4
Evans, Barbara R. 60, 61
Evans, R. David 96

F

Falbel, Tanya 15
Fan, Yao 173
Fang, Gang 178, 195, 196
Farkas, Joel 45
Faybishenko, Boris A. 162, 163, 164, 166
Federowicz, Stephen A. 198
Feist, Adam M. 98
Felnagle, Elizabeth 45
Fels, Samuel 168
Feng, Xiao-jiang 49
Feng, Xueyang 69
Fernie, Alisdair 108
Fields, Matthew W. 156, 157, 162, 163, 165, 166, 167, 169
Filley, Timothy 52
Firestone, Mary 93, 94
Fischer, Curt 80
Fisher, Susan J. 153, 156, 171
Fish, Jordan 18, 92, 181
Fleming, Ronan M.T. 197, 203
Flinn, Barry 151
Flores, Dan Carlo 96
Flowers, Jason 168
Foley, Karen 158
Fortney, Julian L. 162, 166, 172
Foster, Cliff 15, 16
Foston, Marcus B. 34, 43, 60, 61
Fowler, S. 30
Fox, Brian G. 8, 16, 17, 24
Francoleon, D. 104
Fredrickson, James K. 110, 127, 130, 131, 137
Friedland, Greg D. 9, 11
Friedman, Robert 108, 207
Friedman, Yonatan 159, 161
Fristedt, Rikard 205
Frock, A.D. 73
Frolking, Steve 84, 86
Fu, Chunxiang 40, 41, 42
Fu, Na 64, 130
Fuhrman, Jed 107

G

Gallegos-Graves, La Verne 95, 96
Galli, Mary 57
Galloway, Leslie 31
Galloway, Michael 179
Gang, David R. 126
Gao, Ruili 14
Gao, Xiaoli 118
Garcia, David 12
Garcia-Martin, Hector 169
Gardner, Jeffrey G. 22
Garoutte, Aaron 18
Garrity, George M. 183
Gasch, Audrey P. 24
Gelfand, Mikhail S. 189
Geller, Jil T. 156, 157, 162, 166, 170, 171, 207
George, Anthe 4, 11
George, Kevin W. 6, 12
Gerber, Robert B. 60
Gerstein, Mark 178, 195, 196
Ghirardi, Maria L. 68
Giannone, Richard J. 39, 43, 44, 138
Gilna, Paul 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46
Gjersing, Erica 34, 40
Gladden, John M. 8, 9
Glaeser, Robert 153, 157
Glass, Elizabeth M. 175
Glavina del Rio, Tijana 91
Godin, Stephen 51
Goglio, Frances 15
Goh, Ee-Been 13
Gold, Barbara 171
Goltsman, Daniela Aliaga 79
Gomez-Alvarez, Vicente 19
Gonnerman, Matthew C. 104
Goodstein, David 114
Goodwin, Lynne 21
Gordon, Mark S. 200, 201
Gorur, Amita 157
Gottel, Neil 143
Gottel, Neil R. 146
Goulden, Michael L. 88
Grass, Jeffrey A. 22
Greenberg, E. Peter 146
Griffin, J. 30
Griff, Megan N. 114
Grimwood, Jane 4, 114, 115
Grossman, Arthur 205
Gunsalus, R.P. 104
Gunter, Lee E. 38, 140, 149
Guo, Jiarong 18
Guo, Jim 140
Gurtowski, James 178
Guss, Adam 45

H

Hadi, Masood Z. 3, 8, 13
Haft, Dan H. 108
Haft, Rembrandt J.F. 22
Hahn, Michael G. 14, 37, 38
Halbert, Candice 7
Hall, Steve 153
Hall, Steven C. 156, 171
Halpern, Aaron L. 108
Halstead, Adam 24
Ham, T.S. 10
Hamilton-Brehm, Scott D. 41, 43, 44
Hamilton, Choo Y. 41, 42
Hammel, Kenneth E. 54
Han, Kyung-Hwan 15
Hansen, Sara Fasmer 3
Hao, Zhao 5, 67
Harcombe, William 199
Hardin, Frank 40
Harrison, Mary 47
Harwood, Caroline S. 145, 146
Hatcher, P. 85
Haun, Holly L. 208
Hauser, Loren 43

Hazen, Samuel P. 56
Hazen, Terry C. 8, 9, 156, 157, 162, 163, 164, 165, 166, 169, 170, 171, 172, 207
He, Chuan 116
He, Jennifer 154
He, Juhong 60
He, Shaomei 91
He, Zhili 81, 83, 93, 94, 165, 168, 172, 184
Headman, Jennifer 25
Heazlewood, Joshua L. 4, 5
Heins, Richard 8
Heller, William T. 61
Hellsten, Uffe 38, 114
Helmich, Kate E. 24
Hendrix, William 190
Henry, Christopher S. 175, 177, 178, 194
Herman, Donald 93, 94
Herring, Chris 45
Hettich, Robert L. 39, 43, 44, 79, 80, 85, 90, 92, 102, 138, 150, 190
Higbee, Alan 25
Hill, Eric A. 64, 65, 71, 130
Hillesland, Kristina L. 156, 168
Hillson, Nathan J. 8, 10, 12, 207
Himmel, Michael E. 36, 46
Hinchman, Li 24
Hintze, Arend 17, 18
HiroakiYamasaki 205
Hisano, Hiroshi 40
Hixson, Kim K. 117
Hoang, Linh 155
Hodge, David B. 24
Hodgkins, Suzanne B. 84, 86
Hoehler, Tori 68
Holladay, Susan 31, 142
Holman, Hoi-Ying N. 5, 67, 172
Holmes, Bradley M. 4, 11
Hoover, Spencer W. 26
Hopwood, Brett 208
Houtman, Carl J. 54
Howe, Adina 17, 18, 92
Hsia, Chih-Hao 53
Hsu, Shifang 103
Hu, Bo 48
Huang, H. 50
Huber, Dale 7
Hughes, Robert E. 125
Hugo, A. 122
Hunt, Christopher G. 54
Hura, Gregory L. 153, 157
Hurst, Gregory B. 143, 148, 150
Hurt, Richard A. 162, 164
Hutcheson, Steven W. 7
Hu, Zeping 118
Hwang, Mona 63
Hyatt, Doug 31
Hyduke, Daniel R. 197

I

Ibrahim, Yehia M. 118
Inskeep, William P. 127, 136, 137
Isik, Fikret 58
Ivanova, Natalia 29
Izaurralde, R.C. 28

J

Jabbour, Dina 23
Jablin, Michael 7
Jackson, Randall 20
Jansson, Janet K. 18, 90, 92
Jap, Bing K. 156
Jaramillo, Sebastian 129
Javidpour, Pouya 11
Jawdy, Sara S. 149
Jay, Zack 136, 137
Jeffries, T.W. 25
Jenkins, Jerry 4, 114, 115
Jennings, Ryan 127, 136, 137
Jin, Jian 171
Jin, Mingjie 22
Jin, Yong-Su 52
Jo, Ji Hye 76
Joachimiak, Andrzej 173
Joachimiak, Marcin P. 158, 159, 161, 167, 168, 172
Johnson, Courtney M. 144
Jordan, Michael I. 160
Joshi, Hiren 4
Joyner, Dominique C. 162, 166
Juba, Thomas R. 154, 157, 168, 171
Jumpponen, Ari 90, 92
Jurgens, Joel 52
Justice, Nicholas 79

K

Kaeppler, Shawn 16, 28
Kaffenberger, Justin 52
Kalisiak, Ewa 155
Kalluri, Udaya C. 140, 143, 150
Kalyanaraman, A. 122
Kalyuzhnaya, M.G. 99
Kandemkavil, Sindhu 38
Kar, Amrita 199
Karpinets, Tatiana V. 31, 33, 141, 142, 143, 145, 150
Karpowicz, Steven 205
Karp, Peter 193, 194
Karve, Abhijit A. 149
Kataeva, Irina 43
Katahira, Rui 34
Katsnelson, Jacob 7
Kazakov, Alexey E. 159, 161, 187
Kazanov, Marat D. 188
Keasling, Jay D. 4, 6, 8, 9, 10, 11, 12, 13
Keating, David H. 22
Keegan, Kevin 176
Keller, Keith 158, 161, 176, 177, 180
Keller, Kimberly L. 75, 167
Kelly, Peter 113
Kelly, Robert M. 43, 73
Kennedy, David W. 131
Kent, Michael S. 6, 7
Kerley, Marilyn 143, 145, 146
Khairy, Marjon 171
Kholod, Yana 200, 201
Khudyakov, Jane 10
Kidd, M. 85
Kiebel, Gary R. 117
Kieliszewski, Marcia 37
Kiley, Patricia J. 22
Kim, Edward 8
Kim, Helen 151
Kim, Hoon 14, 15
Kim, Jin Sun 4, 5
Kim, Jong-Seo 121
Kim, Joonhoon 26
Kim, Joo-Yeol 15
Kim, Seonhwa 151
Kim, U. 104
Kim, Won-Chan 15
Kim, Young-Mo 127
Kirst, Matias 56
Kitin, Peter 54
Klatt, Christian 133, 136
Kleeson, Sabrina 108
Klein-Marcuschamer, Daniel 3
Klingeman, Dawn M. 144, 145
Klitgord, Niels 125, 199
Ko, Jae-Heung 15
Koech, Phillip K. 135
Konopka, Allan E. 64, 65, 71, 131
Konopka, M.C. 99
Konstantinidis, Konstantinos 81, 83, 123
Kontur, Wayne S. 29
Koppenaal, Dave W. 69
Kora, Guruprasad 31, 141, 142, 143, 180
Korripally, Premsagar 54
Kozubal, Mark 127, 136, 137
Kral, S. 30
Kreuzer, Helen 127, 137
Kronewitter, Scott R. 118

Kruglyak, Leonid 50
Kucek, Leo A. 65, 71
Kuehl, Jennifer V. 75, 154, 168
Kumar, Akhil 192
Kumari, Sunita 178
Kumar, Sunil 171
Kumar, Vinay Satish 13, 14
Kuske, Cheryl R. 89, 95, 96

L

Labbé, Jessy L. 140, 143, 148, 149
Labhsetwar, Piyush 202
LaButti, Kurt 29
Lacayo, Catherine I. 63
Lafontaine Rivera, Jimmy G. 196
Lai, Bergeman 8
Lam, Bonita R. 156, 157, 162, 171
LaMarche, Brian L. 117
Lamarche-Gagnon, G. 85
Lamendella, Regina 92
Lamm, Monica H. 200, 201
Lancaster, W. Andrew 170
Land, Miriam L. 43, 144, 179
Landick, Robert 22
Landorf, Elizabeth 134
Langan, Paul A. 60, 61
Lange, B. Markus 126
Lankford, Patricia K. 145, 150
Lao, Jeemeng 5
Lapierre, Catherine 4
Lappala, Colin R. 146
Lara-Chavez, Alejandra 151
Larimer, Frank W. 43
Larsen, Peter E. 139, 151
Larson, David 7
Lasken, Roger S. 108
Latendresse, Mario 193
Lau, C.Y.M. 85
Lau, Jane 5
Law, Cody 34
Lawrence, Charles E. 55
Layton, A. 85
LeDuc, Richard 20, 180
Lee, Kelvin H. 50
Lee, S. 207
Lee, Scott J. 56
Lee, Taek-Soon 6, 11, 12
Lee, Yongjin 165
Lemmer, Kimberly C. 29
Lenhardt, W. Christopher 208
Lennen, Rebecca M. 26
Lerman, Joshua A. 197
Leung, C.M. 157
Levar, Caleb 124
Lewin, Gina R. 17
Lewis, Michael P. 64
Leyn, Semen A. 187, 188
Li, Changsheng 84, 86
Li, Chenlin 6
Li, Dejun 81, 83
Li, Ting 148, 150
Li, Zhou 150
Liao, James C. 45, 196
Lidstrom, Mary E. 48, 99
Light, Yooli 164
Lin, Hening 113
Lin, Paul P. 45
Lindemann, Stephen R. 127, 131
Lindquist, Erika 10
Link, R.P. 28
Linshiz, G. 10
Liphardt, Jan 157, 167
Lipscomb, Gina L. 75
Lipton, Mary S. 76, 117, 118, 119, 120, 121, 136, 137
Liu, Changjun 116
Liu, Fanghua 96
Liu, Feifei 184
Liu, Haichuan 153, 156, 171
Liu, Hanbin 6
Liu, Kuan-Liang 95
Liu, Nancy L. 154, 171
Liu, Peng 164
Liu, X. 85
Liu, Ying 124
Liu, Yu-San 36
Liu, Zelin 7
Liu, Zhenfeng 133
Lo, Jon 45
Löffler, Frank 123
Logan, Tyler 84, 86
Lombardo, Mary-Jane 108
Long, Tanya M. 25
Loo, J. 104
Loo, R. 104
Loque, Dominique 4, 5
Lovley, Derek R. 96, 98
Lowman, Scott 151
Lu, Fachuang 14, 15
Lu, Tse-Yuan S. 144, 145, 149
Luning, E. 167
Lunin, Vladimir V. 46
Luo, Chengwei 81, 83
Luo, Yiqi 81, 83
Luthey-Schulten, Zaida 202
Lynd, Lee R. 45
Lyons, Catherine 183

M

Ma, Ping 104
Ma, Qin 43
Ma, Suzanne 11, 12
Mack, Jamey C. 173
Majewski, Jaroslaw 7
Majors, Paul D. 131
Malfatti, Stephanie 91
Malik, J. 157
Maness, Pin-Ching 76
Manowitz, D.H. 28
Mansfield, Betty K. 208
Mansfield, Shawn 15
Mao, Xizeng 32
Maranas, Costas D. 50, 70, 191, 192, 196
Maravelias, Christos 26
Markillie, Meng L. 120
Markutsya, Sergiy 200, 201
Marlow, Jeff J. 102
Marshall, Ian P.G. 68
Martin, Francis 148, 150
Martin, Hector Garcia 13, 14
Martin, J. 38
Martin, Joel 29
Martin, Madhavi Z. 42
Martin, Michael C. 172
Martin, Sheryl A. 31, 208
Martin, Stanton L. 144
Martiny, Adam C. 47, 88
Martiny, Jennifer B.H. 88
Marx, Christopher J. 199
Maslov, Sergei 175, 176, 177, 178, 179, 195, 202
Masyuko, R. 138
Mathews, Ian 6
Mavromatis, Konstantinos 92
Mayali, Xavier 67, 94
McAndrew, Ryan 3, 13
McCalley, Carmody K. 84, 86
McCarthy, James 108
McCrow, John P. 107
McCue, Lee Ann 55, 113
McCurdy, James 24
McGarrell, Donna 181
McGill, W.B. 28
McGlynn, Shawn E. 102
McInerney, Peter 3
McKenzie, Heather L. 34, 43
McManus, J. Barry 86
McRose, Darcy 99
Mead, David 21
Meagher, Robert J. 164
Meeks, John C. 72
Mehta, Pankaj 199
Mei, Chuansheng 151
Melnicki, Matthew R. 64, 65, 71
Melton, Sarah 140

Mendez-Perez, Daniel 26
Meng, Da 117
Menon, Angeli Lal 170
Merchant, Sabeeha 205
Merchant, Sabeeha S. 77
Metcalf, William W. 104
Metz, Thomas O. 117, 118, 127
Meyer, Birte 167, 168
Meyer, Folker 175, 176, 179
Meyers, Blake C. 50
Michalska, Karolina 173
Mielenz, Jonathan R. 40, 41, 42
Millen, Jonathan 76
Miller, Christopher 79
Miller, Jeffrey 38
Miller, Lisa M. 173
Mills, Marissa D. 208
Milne, Caroline B. 52
Mishra, Sujata 69
Mistry, Dhaval B. 43
Mitra, Prajakta 4, 5
Moberly, James G. 164, 167
Mohammadi, Allison 14
Mohnen, Debra 33, 37
Mondav, Rhiannon 84, 86
Monroe, Matthew E. 117, 118
Montiel, Daniel 53
Moore, Ronald J. 117, 118, 136
Moran, James J. 127, 131, 137
Morlen, Ryan 146
Morrell-Falvey, Jennifer L. 140, 149
Mosher, Jennifer J. 162, 164, 170
Mosier, Annika 79
Mouille, Grégory 4
Moulton, Steve 179
Muchero, Wellington 38, 115, 140, 143, 148
Muddiman, David 58
Mueller, Daniel B. 146
Mueller, Thomas J. 70
Mukerjee, Purba 4
Mukherjee, Arijit 28
Mukherjee, Supratim 51
Mukhopadhyay, Aindrila 4, 12, 167, 168
Mukhopadhyay, Biswarup 186
Mullen, Robert T. 59
Murnane, Mollie 96
Murphy, J. 85
Murton, Jaclyn K. 7
Myers, Richard M. 115
Mykytczuk, N. 85
Myneni, S. 85
Myrold, David D. 90, 92

N

Nag, Ambarish 33
Nagarajan, Harish 96, 98
Nealson, Kenneth H. 107, 108
Nelson, David D. 86
Neves, Leandro Gomide 56
Newberg, Lee A. 55
Newcomer, D.R. 163
Nico, Peter 172
Nicora, Carrie D. 69, 117, 118
Nixon, B. Tracy 53
Nixon, William J. 75
Noguera, Daniel R. 29
Noll, K.M. 73
Norbeck, Angela D. 117
Northen, Trent R. 8, 79, 80, 93, 125, 153, 154, 155, 156, 157, 167, 168, 180
Notey, J.S. 73
Novichkov, Pavel 159, 161, 179
Novichkov, Pavel S. 158, 177, 187, 188, 189
Novotny, Mark 108
Nowak, Jerzy 151
Nuccio, Erin 93, 94
Nunes-Nesi, Adriano 108
Nyyssönen, Mari 67, 88

O

O'Brien, Edward J. 197
Ohlrogge, John 27
Oleskowicz-Popiel, Piotr 3
Olivar, Estefania 45
Olson, Andrew 178
Olson, Bob 179
Olson, Dan 179
Olson, Daniel G. 45
Olson, Robert 194
O'Neill, Hugh M. 60, 61
O'Neill, Malcolm 33
Onstott, T.C. 85
Orphan, Victoria J. 99, 101, 102
Orton, Daniel J. 117, 118
Osipiuk, Jureck 173
Osterman, Andrei L. 110, 135
Ouellet, Mario 12
Overbeek, Ross 177, 179, 194
Ozdemir, Inci 43

P

Padmanabhan, Kanchana 190
Page, Lawrence E. 70
Pakrasi, Himadri B. 69, 70
Palsson, Bernhard Ø. 98, 197, 198
Palumbo, Anthony V. 162, 164
Pan, Chongle 79, 80, 102, 150, 180, 190
Pandey, Gaurav 160
Pang, Tin Yau 202
Pant, Kapil 204
Papanek, Beth 52
Papoutsakis, E. Terry 50
Parang, Morey 31
Parello, Bruce 179, 194
Park, Byung 31, 141
Park, J.I. 9
Park, Ji-Young 15
Park, Joshua 9
Park, Junyoung O. 49
Park, Sunkyu 58
Parker, Charles T. 183
Parks, Jerry M. 42
Parsons, Harriet 4
Paša-Tolić, Ljiljana 69, 117
Passian, Ali 35
Pasternak, Shiran 178, 179
Paternak, Shiran 179
Pattathil, Sivakumar 37, 38
Patterson, Sara E. 15
Paulsen, Ian T. 107
Pearson, Timothy 14
Pell, Jason 17, 18
Pelletier, Dale A. 140, 141, 143, 144, 145, 146, 149, 150
Peña, Maria 33
Pennacchio, Christa P. 24, 38
Pennacchio, L. 38
Penton, Ryan 81, 83
Peralta-Yahya, Pamela P. 11, 13
Pereira, Jose H. 3, 9, 13
Perez, José 110
Persil-Cetinkol, Ozgul 4
Petersen, Pia Damm 5
Peters, Theodore 125
Petridis, Loukas 36, 61
Petrus, A. 73
Pett-Ridge, Jennifer 67, 68, 93, 94
Petzold, Christopher 9, 12, 153
Pfiffner, S. 85
Pfleger, Brian F. 26
Phelps, Tommy J. 85, 162, 164
Phillips Jr., George N. 17, 24
Phinney, Brett 72
Pietri, Ruth 173
Pinchuk, Grigoriy E. 65, 71, 130
Pincu, Madeleine 60
Pinel, Nicolas 168
Pingali, Sai Venkatesh 60, 61
Plahar, H.A. 10
Plett, Jonathan 148, 150

Podar, Mircea 138, 143, 146, 162, 164
Pollard, Mike 27
Pollard, W. 85
Pollet, Brigitte 4
Poole, Farris L. 170, 171
Poole II, Farris L. 43
Poovaiah, Charleson R. 40
Porat, Iris 75
Poretsky, Rachel S. 101
Porras-Alfaro, Andrea 95
Prabhakarpandian, B. 204
Prak, Lina 5
Pregitzer, Kurt S. 89
Price, Morgan N. 153, 154, 158, 160, 161, 167, 168
Price, Morgan R. 172
Price, Nathan 202
Price, Nathan D. 52, 104
Priya, Ranjan 140
Prufert-Bebout, Leslie 68
Pu, Yunqiao 42
Purvine, Samuel O. 100, 117
Purwantini, Endang 186

Q

Qian, Wei-Jun 64, 118, 121
Qin, Yujia 184
Qu, Yi 76, 121
Quest, Daniel 179

R

Rabinowitz, Joshua D. 49, 50
Rabitz, Herschel 49, 50
Ragauskas, Arthur J. 34, 42, 43, 60, 61
Rajeev, Lara 167, 168
Ralph, John 4, 14, 15, 54, 58
Ramsay, B.D. 163
Ranjan, Priya 38, 148, 178
Rapp-Giles, Barbara J. 75
Ravcheev, Dmitry A. 187, 189
Reaves, Marshall L. 49
Redding, Alyssa 9
Redding-Johanson, Alyssa 12
Reed, Jennifer L. 22, 26, 65
Reeves, Darryl 178, 195, 196
Regner, Matthew 14
Reichenberger, Erin 203
Reindl, Wolfgang 8
Reiss, David J. 158, 159, 161
Reistetter, Emily 99, 100
Remis, J.P. 157
Repo, Susanna 160
Resch, Michael G. 46
Reveco, Sonia A. 9, 154, 157, 171
Reyes-Ortiz, Vimalier 8
Rice, Charles W. 90, 92
Richardson, Ruth E. 113
Richter, R. Alexander 108
Rich, Virginia 86
Rich, Virginia I. 84
Riehl, William J. 199
Risser, Douglas 72
Ritalahti, Kirsti 123
Riveri, Sebastian Jaramillo 134
Roberts, Elijah 202
Robinson, W. 30
Rocha, Andrea M. 190
Rodemeyer, Michael 207
Rodgers-Melnick, E. 38
Rodionov, Dmitry A. 110, 135, 187, 188, 189
Rodriguez Jr., Miguel 41
Rohlin, L. 104
Rohwer, Robin 53
Rokhsar, Daniel 4
Rokhsar, Daniel S. 38, 114, 115
Rollefson, Janet 124
Romine, Margaret F. 127
Romine, Margie 110, 120, 134, 135, 137
Ronald, Pamela C. 4, 5, 7, 178
Roquet, Nathaniel 49
Rose, Jocelyn 63
Rosen, Gail 203
Rosengarten, Rafael 12
Rosengarten, R.D. 10
Rotaru, Amelia-Elena 96, 98
Rowe, Annette R. 113
Rubin, Edward M. 114
Rüegg, Thomas 10
Rusch, Douglas B. 107, 108
Rutoski, Thomas J. 24

S

Saarunya, G. 85
Sadler, Natalie C. 64
Saffold, Trina 33
Saha, Rajib 70
Salazar, Jennifer 175
Sale, Kenneth L. 6, 8, 9
Saleska, Scott R. 84, 86
Samad, Bushra 185
Samatova, Nagiza F. 190
Sam, Daniel 100
Samuel, Reichel 42
Samuels, Reichel 34
Sanders, R. 85
Sandoval, John D. 117
Sanford, Robert 123
Sangha, Amandeep K. 42
Sanjaya 27
Santoro, Nicholas 16
Santoro, Nick 15
Sapra, Rajat 8
Sato, Trey K. 24
Saunders, Michael A. 197, 203
Saxena, Abigail 101
Sayavedra-Soto, Luis 110
Scanlon, Michael 126
Schackwitz, Wendy 24, 38
Schadt, Christopher W. 141, 142, 143, 144, 145, 146, 147, 149, 164
Schaefer, Amy L. 145, 146
Schatz, Michael 178
Scheller, Henrik Vibe 3, 5
Schilling, Jonathan S. 52
Schmidt, Adam 126
Schmidt, Matthew C. 190
Schmidt, Thomas M. 19
Schmutz, Jeremy 4, 114, 115
Schulson, H. 30
Schut, Gerrit J. 75
Schuur, Edward A.G. 81, 83
Schwab, David J. 49
Schwab, D.E. 28
Schwalbach, Michael S. 22
Scott-Craig, John S. 23, 24
Scott, Israel 170
Scott, Robert A. 75
Seaver, Sam 178
Sederoff, Ron 58
Segrè, Daniel 199
Seiler, John 151
Sekhon, Rajandeep 16, 28
Selengut, Jeremy D. 108
Semrau, Jeremy D. 112
Seokwon, Jung 43
Serres, Gretta 120, 129
Serres, Margrethe (Gretta) 134
Setubal, Joao C. 186
Severins, Ryan 76
Shachar-Hill, Yair 27
Shackwitz, Wendy 29
Shah, Anuj 117
Shah, Manesh 92
Shakya, Migun 143, 146
Shanklin, John 116
Shao, Jason 24
Sharma, Manoj 4, 7
Sharma, Rita 7
Shatsky, Maxim 156, 161, 171
Shaw, A. Joe 45
Shen, Hui 40
Shen, Yaoqing 185
Shen, Yufeng 117
Sherman, Louis A. 69, 70

Sherwood, Robert 113
Shi, Shengjing 93, 94
Shi, Zhou 165
Shresta, Pravin Mala 96
Shrestha, Pravin 98
Shrout, J. 138
Shukla, Anil K. 117, 118
Shumpert, Barry 208
Shutthanandan, Janani I. 69
Simmons, Blake A. 3, 4, 6, 7, 8, 9, 10, 11, 24
Sims, David 115
Singer, Steven W. 8, 9, 68, 79
Singh, Anup K. 8, 162, 164, 168
Singh, R. 73
Singh, Seema 4, 6, 7
Singhal, Mudita 117
Siuzdak, Gary 153, 155
Sjölander, Kimmen 185
Slater, Steve 180
Slavov, G.T. 38
Sleighter, R.L. 85
Slysz, Gordon W. 117, 118, 136
Smilgies, Detlef 63
Smith, Chris 58
Smith, Jeremy C. 36, 42, 61, 200
Smith, Jessica 96
Smith, Mark B. 162
Smith-Moritz, Andreia Michelle 5
Smith, Richard 99, 100
Smith, Richard D. 69, 117, 118, 119, 121
Smith, Robert W. 24
Soderlund, Carol A. 126
Song, Rong 165
Soubigou-Taconnat, Ludivine 4
Sousa, Leonardo da Costa 22
Souza, Michael L. 160, 161
Spormann, Alfred M. 68
Squier, Thomas C. 64, 121, 130
Sreedasyam, Avinash 139
Srinivas, Goundla 200
Stackhouse, B. 85
Stahl, David A. 156, 162, 165, 167, 169
Standaert, Robert F. 42
Stavila, Vitalie 6
Steele, Joshua A. 101
Steen, Eric 11
Steinke, Laurey 136
Stephanopoulos, Greg 196
Steven, Blaire 96
Stevens, Rick 175, 176, 177, 178, 179, 194
Stewart Jr., Neal 40
Stöckel, Jana 69
Stolyar, Sergey M. 127, 130, 131, 133
Stuart, Joshua 99, 100
Su, Dian 64, 121
Su, Yi-Kai 25
Sukumar, Poornima 143, 150
Sun, Lan 4, 5, 7
Sun, Shulei 101, 102
Sun, Yanni 181
Sun, Yuekai 197
Suthers, Patrick F. 192
Suttangkakul, Anongpat 5
Sweedler, J.V. 138
Swithers, S. 73
Syed, Mustafa 31, 141, 142, 180
Sykes, Robert 34, 40
Synovec, Robert 48
Szakal, Evelin D. 153, 156, 171
Szmidt, Heather 12

T

Tainer, John 153, 157
Takasuka, Taichi E. 16, 17
Takasumi, Jennifer 45
Talkwalkar, Ameet 160
Tan, Kemin 173
Tan, Li 37
Tang, Yinjie 69
Tang, Yuhong 151
Tarjan, D. 207
Tauscher, A. 157
Tautenhahn, Ralf 155
Tavormina, Patricia L. 101
Taylor, Ronald C. 120
Teal, Tracy K. 19
Tetard, Laurene 35
Thelen, Michael P. 9, 10, 63
Thiagarajan, Mathangi 107
Thiele, Ines 203
Thomas, Brian C. 79, 180
Thomason, Jim 178
Thompson, William A. 55
Tice, Hope 29
Tiedje, James M. 17, 18, 20, 81, 83, 92, 181
Tien, Ming 53
Timberlake, Sonia 161
Titmus, Matt 178
Tobimatsu, Yuki 14
Tolić, Nikola 117
Tolmachev, Aleksey V. 117
Tomkins, Bruce A. 43
Tornqvist, Carl-Erik 15
Torok, Tamas 162, 207
Torto-Alalibo, Trudy 186
Tran, Huu 8
Travers, Mike 193
Tremaine, Mary 22
Tremblay, Pier-Luc 96
Treseder, Kathleen K. 88
Trimble, William 176
Tringe, Susannah 18, 92, 137
Tringe, Susannah G. 90, 91
Trivedi, Geetika 139
Tschaplinski, Timothy J. 40, 42, 43, 44, 75, 148, 149, 150
Tu, Qichao 165
Tullman-Ercek, Danielle 8, 11
Turkarslan, Serdar 158, 161
Turkaslan, Serdar 159
Turner, Geoff 34
Tuskan, Gerald A. 32, 38, 39, 43, 114, 115, 140, 141, 143, 147, 148, 149, 150
Tyler, Brett M. 186
Tyson, Gene W. 84, 86

U

Uberbacher, Edward 31, 141, 142, 143, 180
Udvardi, Michael 32
Ueda, Ken-ichi 180
Urbanowicz, Breeanna 33
Urban, Volker 60, 61
Utturkar, Sagar M. 144, 145, 146

V

Vacarro, Brian 170
Valas, Ruben 108
Vallabhajosyula, Ravishankar R. 204
van Baren, Marijke J. 100
Van Berkel, Gary J. 43
van der Veen, Douwe 45
Van Nostrand, Joy D. 81, 83, 162, 165, 184
Varanasi, Patanjali 7
Vega-Sanchez, Miguel E. 5, 7
Venter, J. Craig 107, 108
VerBerkmoes, Nathan C. 90, 92
Verhertbruggen, Yves 5
Vilgalys, Rytas 95, 143, 147
Vishnivetskaya, T. 85
Voigt, Christopher 11
Voit, Eberhard 32
Voordeckers, James 165
Vu, Trang 65

W

Waldrop, Mark 91
Walian, Peter J. 156, 168
Walker, Larry 63

Wall, Judy D. 75, 154, 157, 162, 165, 167, 168, 171, 172
Waltman, Peter 100
Walton, Jonathan D. 23, 24
Wanamaker, Shelly 57
Wang, Chao 7
Wang, Lu 121
Wang, Qiong 181
Wang, Shan 32
Wang, Stephanie 208
Wang, Wendy D.M. 76
Wang, Yi 52
Wang, Zeng-Yu 40, 41, 42
Ward, David M. 127, 133, 136
Ware, Doreen 175, 178
Warrier, Ajithkumar 185
Wasylenko, Thomas 196
Watson, Brian J. 7
Weaver, Daniel 13
Webb-Robertson, B.-J. 122
Weber, Carolyn F. 95
Weber, Peter K. 67, 68
Wecker, Matthew 68
Weerasinghe, Deepika 194
Wehr, Rick 84, 86
Weihe, Claudia 88
Weisenberger, Drew 117
Weitz, Karl K. 117, 118
Welsh, Allana 123
Welsher, Kevin 53
Weston, David J. 140, 149, 150, 175, 178
Westpheling, Janet 30, 45
Wetmore, Kelly 154, 168
Whitaker, Rachel J. 104
White, D. 73
Whitmore, Laura 127, 137
Whyte, L. 85
Wiley, H. Steven 120, 131
Wilhem, R. 85
Wilke, Andreas 176
Wilkening, Jared 176, 179
Wilkerson, Curtis 15
Wilkins, Michael 113
Williams, Cranos 58
Williams, J.R. 28
Williamson, Shannon J. 107
Willis, Laura B. 25
Windham-Myers, Lisamarie 91
Windus, Theresa L. 200, 201
Wingreen, Ned S. 49
Withers, Saunia 15
Witkowska, H. Ewa 153, 156, 171
Woebken, Dagmar 68
Wohlbach, Dana J. 24
Wolfe, Amy K. 208
Wong, April 8
Woo, Hannah 10
Worden, Alexandra Z. 99, 100
Woyke, Tanja 137
Wright, Aaron T. 64, 110, 117, 118, 121, 135
Wu, C. 207
Wu, Cathy H. 50
Wu, J.H. David 76
Wu, Liyou 81, 83, 165
Wu, Wei-Ju 158
Wyman, Charles E. 34, 38, 43
Wyrick, Judy M. 208

X

Xia, Fangfang 194
Xiao, Xirong 40
Xie, Enhai 24
Xie, Gang 95
Xie, X. Sunney 62
Xiong, Yijia 130
Xu, Qi 46
Xu, Xia 81, 83
Xu, Yi-Fan 49
Xu, Ying 31, 32, 43
Xue, Kai 81, 83

Y

Yan, Koon-Kiu 195
Yan, Shuangchun 99
Yang, Fan 4, 5
Yang, Haw 53
Yang, Lee 171
Yang, S. 99
Yang, Xiaohan 148, 150
Yang, Zamin 143, 146
Yannone, Steve 167
Yao, Qin 116
Ye, Chu-Yu 148
Yee, Kelsey 41
Yee-Greenbaum, Joyclyn 108
Yennamalli, Ragothaman M. 17
Yilmaz, Suzan 164, 168
Yin, Hengfu 148
Yin, Yanbin 32, 43
Yoo, Shinjae 178
Yoon, Sukhwan 123
Yooseph, Shibu 107, 108
York, William S. 33
Youens-Clark, Ken 178
Youngblood, Matt 125
Youngquist, J. Tyler 26
Yu, Dantong 178
Yu, Dantung 179
Yu, Hao 165
Yuan, Mengting 81, 83
Yunes, Jeffrey M. 160

Z

Zagursky, Robert J. 76
Zahniser, Mark S. 86
Zak, Donald R. 89
Zane, Grant M. 154, 167, 168
Zeglin, Lydia H. 90, 92
Zemla, Marcin 7
Zengler, Karsten 96, 98, 197, 198
Zeng, Lucy 156
Zeng, Yining 36
Zerbs, Sarah 173
Zhang, Fuzhong 11
Zhang, Haito 7
Zhang, Jin 81, 83
Zhang, Ling 5
Zhang, Ping 165
Zhang, Qibin 118
Zhang, Rui 117
Zhang, Sheng 113
Zhang, Xiaohui 69
Zhang, Yaoping 22
Zhang, Yuan 76
Zhao, Rui 117, 118
Zhou, Aifen 165, 168, 172
Zhou, Jizhong 81, 83, 93, 94, 162, 165, 172, 184
Zhou, Joe 168
Zhu, Xiang 37
Zhu, Yimin 14, 15
Zhu, Zhengjiang 155
Ziebell, Angela 34, 40
Zink, Erika M. 64
Ziolkowski, L. 85
Zomorrodi, Ali R. 191, 196
Zurawski, Jeffrey V. 43

Institutional Index

Indexed by Page Number

Institutions determined by affiliations provided by authors

Aerodyne Research Inc. 86
Ames Laboratory 201
Argonne National Laboratory 110, 134, 139, 151, 173, 175, 176, 177, 178, 179
BioEnergy Science Center 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 62, 180
Boston University 199
Brookhaven National Laboratory 116, 173, 175, 176, 177, 178, 179, 195, 202
Brown University, Providence 55
Buck Institute 125
C5•6 Technologies 21
California Institute of Technology 99, 101, 102
Carnegie Institution for Science 205
Center for Integrated Nanotechnologies 7
CFD Research Corporation 204
Chonnam National University, Gwangju 15
Cold Spring Harbor Laboratory 175, 178, 179
Cornell University 46, 63, 80, 113, 126
Creative Discovery Museum 30
Dartmouth College 45
Delaware Biotechnology Institute 50
DOE JGI HudsonAlpha Genome Sequencing Center 114, 115
DOE Joint Genome Institute 4, 10, 18, 24, 29, 38, 90, 91, 92, 114, 115, 137
Donald Danforth Plant Science Center 63
Drexel University 203
Duke University 95, 143, 147, 160
ENIGMA 75
Enivronmental Protection Agency ORD-NERL 19
Fellowship for Interpretation of Genomes 194
Florida State University, Tallahassee 84, 86
Forest Products Laboratory 25
Genebiologics, LLC 68
Genome Science and Technology Program 150
Georgetown University Medical Center 50
Georgia Institute of Technology 34, 42, 43, 60, 61, 81, 83, 123
Great Lakes Bioenergy Research Center 4, 14, 15, 16, 17, 20, 21, 22, 23, 24, 25, 26, 28, 29, 38, 180
Harvard University 62, 199
Hebrew University, Jerusalem 60
Hope College 189, 194
HudsonAlpha Institute of Biotechnology 4
IJPB-AgroParisTech 5
INRA-CNRS Evry 5
INRA-Nancy 148, 150
Institute for Advanced Learning and Research 151
Institute for Information Transmission Problems RAS 187, 188, 189
Institute for Systems Biology 52, 104, 155, 157, 158, 159, 160, 161, 167, 169
Iowa State University, Ames 112, 201
J. Craig Venter Institute, Rockville 107, 207
J. Craig Venter Institute, San Diego 107, 108, 207
Joint BioEnergy Institute 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 24, 180, 207
Kansas State University 90
Kansas State University, Manhattan 92
Kellogg Biological Station 19
Kent State University 103, 203
Kyung Hee University, Yongin 15
Lawrence Berkeley National Laboratory 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 18, 38, 67, 68, 76, 79, 80, 88, 90, 92, 93, 94, 125, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 175, 176, 177, 178, 179, 180, 187, 188, 189, 199, 207
Lawrence Livermore National Laboratory 63, 67, 68, 93, 94
Los Alamos National Laboratory 7, 18, 21, 85, 89, 95, 96
Louisiana State University 67
Lucigen Corporation 21
Macquarie University, Sydney 107
Marine Biological Laboratory 120, 129, 134
Mascoma Corporation 45
Massachusetts Institute of Technology 11, 159, 161, 162, 196
Max-Planck-Insitut für Molekulare Pfanzenphysiologie 108
McGill University 85
McMaster University 85
Michigan State University 27, 92
Michigan State University, East Lansing 15, 16, 17, 18, 19, 20, 22, 23, 24, 81, 83, 181, 183
Montana State University 127, 129, 156, 157, 163, 167, 169
Montana State University, Bozeman 127, 129, 133, 136, 137, 138, 162, 165, 166
Monterey Bay Aquarium Research Institute 100
Moscow State University 189
Mount Sinai School of Medicine 160
NamesforLife, LLC 183
NASA Ames Research Center 68
National Renewable Energy Laboratory 14, 33, 34, 36, 40, 46, 62, 63, 68, 76
North Carolina State University, Raleigh 43, 58, 73, 190
Oak Ridge National Laboratory 7, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 60, 61, 75, 79, 80, 85, 90, 92, 102, 114, 115, 123, 138, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 162, 163, 164, 165, 166, 167, 169, 170, 171, 172, 175, 176, 177, 178, 179, 180, 190, 208

Ohio University, Athens 37
Old Dominion University 85
Oregon State University 90
Oregon State University, Corvallis 92, 110, 186
Pacific Northwest National Laboratory 28, 55, 64, 65, 69, 71, 76, 100, 110, 113, 117, 118, 119, 120, 121, 122, 127, 129, 130, 131, 133, 134, 135, 136, 137, 163
Pennsylvania State University, University Park 50, 53, 64, 70, 129, 133, 136, 191, 192, 196
Princeton University 49, 50, 53, 85
Purdue University 69, 70
Purdue University, West Lafayette 52
Rensselaer Polytechnic Institute 55
Salk Institute for Biological Studies 57
Sandia National Laboratories 169
Sandia National Laboratories, Livermore 6, 7, 162, 164
San Diego State University 194
Sanford-Burnham Medical Research Institute 110, 135, 187, 188, 189
Scripps Center for Mass Spectrometry, La Jolla 153
Southern Illinois University 167
SRI International 193, 194
Stanford University 68, 197, 203
Stockholm University 84, 86
Stony Brook University, New York 202
Texas A&M University 28
The Samuel Roberts Noble Foundation, Inc. 14, 40, 41, 42, 151
The Scripps Center for Mass Spectrometry 155
Thomas Jefferson National Accelerator Facility 117
University at Albany 199
University of Alabama, Birmingham 151
University of Alabama, Huntsville 139, 151
University of Arizona, Tucson 84, 86, 126
University of British Columbia, Vancouver 15
University of California, Berkeley 5, 8, 79, 80, 93, 94, 156, 157, 160, 161, 171, 180, 185, 190, 207
University of California, Berkeley and San Francisco 8
University of California, Davis 4, 7, 24, 72, 178
University of California, Irvine 47, 87, 88
University of California, Los Angeles 45, 77, 104, 196, 205
University of California, Riverside 34, 38, 43
University of California, San Diego 101, 107
University of California-San Diego, La Jolla 96, 98, 102, 197, 198
University of California, San Francisco 153, 156, 171
University of California, Santa Cruz 100
University of Chicago 116, 194
University of Connecticut, Storrs 73
University of Copenhagen 5
University of Delaware, Newark 50
University of Florida, Gainesville 56, 81, 83
University of Georgia 14, 33, 37, 38, 75, 167, 171
University of Georgia, Athens 30, 31, 32, 33, 43, 45, 170
University of Guelph 59
University of Iceland 197, 203
University of Idaho, Moscow 89
University of Illinois, Urbana 52, 104, 202
University of Illinois, Urbana-Champaign 123, 138
University of Irvine 60
University of Kentucky 63
University of Maryland, College Park 7, 28
University of Massachusetts, Amherst 51, 56, 96, 98
University of Michigan, Ann Arbor 89, 112
University of Minnesota, Saint Paul 52
University of Minnesota, Twin Cities 124
University of Missouri, Columbia 75, 154, 157, 162, 165, 167, 168, 169, 171, 172
University of Natural Resources and Applied Life Sciences, Vienna 28
University of Nebraska, Lincoln 73
University of Nebraska Medical Center, Omaha 136
University of New Hampshire, Durham 84, 86
University of Northern British Columbia 28
University of North Texas 59
University of Notre Dame 138
University of Oklahoma 4
University of Oklahoma, Norman 81, 83, 93, 94, 162, 165, 169, 172, 184
University of Queensland 84, 86
University of Rochester 76
University of São Paulo, Brazil 186
University of Southern California, Los Angeles 107
University of South Florida 190
University of Tennessee 163, 208
University of Tennessee, Knoxville 31, 33, 40, 42, 85, 123, 138, 143, 144, 145, 146, 156, 162, 164, 165, 166
University of Texas, Austin 85, 146
University of Vermont, Burlington 47
University of Virginia, Charlottesville 207
University of Washington 145, 146, 156, 167
University of Washington, Seattle 48, 99, 145, 146, 162, 165, 169
University of Wisconsin, Madison 4, 8, 14, 15, 16, 17, 20, 21, 22, 24, 25, 26, 28, 29, 54, 58, 65, 180
USDA-ARS 123, 178
USDA-ARS-ALARC 59
U.S. Forest Products Laboratory 54
U.S. Geological Survey, Menlo Park, Calif. 91
U.S. Geological Survey, Moab, Utah 96
Virginia Commonwealth University, Richmond 72
Virginia Polytechnic Institute, Blacksburg 7, 151, 186
Wadsworth Center, New York State Department of Health 55
Washington State University, Pullman 96, 122, 126
Washington University, St. Louis 69, 70
Western Illinois University 95
West Virginia University, Morgantown 38
Xavier University of Louisiana 114
Yale University 178, 195, 196
Zhengzhou Tobacco Research Institute 4