U.S. DEPARTMENT OF ENERGY
Office of Science

USDA
United States Department of Agriculture

Genomic Science Contractors–Grantees Meeting XIV

USDA-DOE Plant Feedstock Genomics for Bioenergy Meeting

Abstract Book

March 6-9, 2016

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

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

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

Department of Energy
Washington, DC 20585
March 8, 2016

Dear Colleague:

On behalf of the Genomic Science Team as well as the entire Biological Systems Science Division (BSSD), welcome to the 2016 Genomic Science Annual PI Meeting! We look forward to another very interesting meeting this year and to discussing your latest scientific accomplishments in support of the program. This is the primary annual meeting for the Genomic Science program and seeks to bring together the principal investigators supported by the program to discuss new research results with each other, DOE program staff, representatives from elsewhere within the Department of Energy, and colleagues from other Federal Agencies.

I would like to especially welcome new PIs to the program. You will not find another meeting that is more directly relevant to your DOE research project than this annual meeting. We encourage you to take full advantage of this opportunity to enlarge your network and develop collaborations with scientists from the broad range of disciplines represented here to further enhance the innovation and success of your research efforts.

I would also like to extend a welcome to our colleagues from other DOE Offices and Federal Agencies. Please feel free to discuss your program(s) with researchers gathered at this meeting and working in areas of mutual interest.

The past year has been exciting and productive for the Genomic Science program. The agenda is packed with plenary and breakout session talks highlighting exceptional research results produced by individual investigators, interdisciplinary research teams and the Bioenergy Research Centers (BRCs).

The BRCs are now well into their second 5-year phase of operations and continue to be extremely productive, paving the way for the conversion of cellulosic biomass to biofuels. This year the BRC plenary session will be introspective as each center was asked to discuss their journey over the past 8+ years highlighting their past achievements and sharing exciting new bioenergy research results. FY 2016 will be a pivotal year for BER's efforts in bioenergy research as we anticipate the release of a new competitive FOA for this area of our portfolio.

The DOE Systems Biology Knowledgebase (KBase), a computational platform to integrate, analyze and share omics data and bioinformatics tools, is rolling out its next set of capabilities via the Software Development Kit (SDK). This new capability lowers the barriers and allows much faster integration of third-party tools and software into the KBase system. At this year's meeting the team will feature the "The KBase Developers' Workshop" that will demonstrate how researchers can integrate their command-line tools into the KBase platform. Also, at "The KBase Experience Hands-On Session" attendees can try out various KBase capabilities on their laptops by bringing their own microbial data to analyze, or using microbial and plant data already stored in the system. Members of the KBase team will be on hand to assist attendees at both sessions.

This meeting also provides an opportunity for researchers to learn about additional capabilities available at DOE user facilities such as the Joint Genome Institute (JGI), the Environmental Molecular Sciences Laboratory (EMSL) as well as capabilities at the DOE Synchrotron Light and Neutron Sources. JGI, EMSL and KBase will each have presentations during the plenary session and representatives from each of these groups will also be available at the poster sessions to meet with attendees and provide more information.

We are extremely pleased to welcome Dr. Victoria Orphan, Professor of Geobiology at the California Institute of Technology as our keynote speaker. Environmental microbiology continues to be a foundational component of the Genomic Science program as we seek to uncover the myriad ways that microorganisms adapt and thrive in the environment and then use that knowledge to help solve challenging bioenergy and environmental problems. Dr. Orphan's research is very much at the forefront of interdisciplinary science. Her investigations of anaerobic methane oxidation are an excellent mix of classic microbiology combined with all the modern day–omics technologies and the latest analytical imaging techniques to very clearly unravel the detailed physiology of this very enigmatic process. Dr. Orphan's deft application of a broad range of system biology approaches and analytical techniques is an excellent example of the multidisciplinary approach to fundamental science that we seek to foster

within the program. We are confident you will enjoy her presentation and find it to be an excellent catalyst for scientific discussion throughout the meeting.

As 'omics technologies continue to rapidly advance, the production of very large and complex (and diverse) datasets poses new computational challenges to efficiently derive biological meaning from this flood of data. This year one of our plenary sessions highlights a few new approaches to analyze information-rich omics data sets while another session highlights the use of omics data for metabolic engineering. This latter session highlights the Biosystems Design projects and will report on the latest advances in metabolic engineering and redesign of biological processes for biofuel production.

This year's breakout sessions will cover a broad range of BER mission-relevant topics from the subcellular details of metabolic pathways to the complexity of eukaryotic systems and microbial communities. One of our breakout sessions this year is focused on imaging technologies for systems biology and we are happy to welcome a few representatives from our new bioimaging technology development effort within the Division. These new projects focus on developing new integrative imaging capabilities for plant and microbial research. We are also excited this year to have presentations from several projects in our new Sustainable Bioenergy effort. Funded last year, these projects will provide brief introductions on how each are approaching the challenge of understanding complex microbe-plant-soil ecosystems for bioenergy production.

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

I know I speak on behalf of the entire staff when I say we are privileged to work with each and every one of you. This is an exceptionally strong program for DOE and that is due to your scientific creativity, insights and hard work! Thank you for making the program the success that it is. We look forward to an excellent meeting.

Sincerely,

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

Table of Contents

Droplet Microfluidic Platform for High Throughput Screening and Synthetic Biology Applications

Phillip Gach [1,2], Steve C.C. Shih[1,2], Garima Goyal[1,3], Peter Kim[1,2], Manasi Raje[1,2], Joshua Heinemann[1,3], Todd Duncombe[1,2], Daniel Arlow[1,4], Kai Deng[1,2], Kosuke Iwai[1,2], Seema Singh[1,2], Kenneth Sale[1,2], , Blake Simmons[1,2], Jay Keasling[1,3,4], Paul Adams[1,3,4], Trent Northen[1,3], Nathan Hillson[1,3], **Anup K. Singh**[1,2] **(aksingh@sandia.gov)**

[1]Joint BioEnergy Institute, Emeryville, CA; [2]Sandia National Laboratories, Livermore, CA; [3]Lawrence Berkeley National Laboratory, Berkeley, CA; [4]University of California, Berkeley, CA

http://www.jbei.org

Project Goal:
The JBEI mission is to conduct basic and applied research to enable economically-viable conversion of lignocellulosic biomass into transportation biofuels.

Abstract:

Microfluidic assays and devices have attracted a significant attention for performing biochemical reactions and analysis as they provide significant improvements over their macroscale counterparts with respect to speed, throughput, and multiplexing. We are involved in developing innovative microfluidic assays and integrated devices for many biofuel research applications including enzyme screening, enzyme evolution and synthetic biology. Currently, these experiments are done manually using fairly large amount of costly reagents per experiment making the process very expensive, extremely slow and irreproducible. We have developed a platform that uses droplets as discrete reaction chambers to integrate and automate the processes of reagent dispensing, addition, incubation and screening by mass spectrometry. The microscale platform allows these reactions to be performed faster using minimal manual intervention while consuming 10-100-fold lower reagents. Integration with mass spectrometry enables high sensitivity detection.

One example application is automation of synthetic biology experiments. Optimization of pathways can involve very large number of experiments as multiple variants are available for each gene. Currently, these experiments are done manually using fairly large amount of costly reagents per experiment making the process very expensive, extremely slow and irreproducible. Droplet platform can integrate and automate the processes of DNA assembly, transformation, and cell culture in one device. The hybrid chip combines droplets-in-flow and digital microfluidic (DMF) formats to take advantage of the high throughput nature of droplets-in-flow and the precise control over droplet manipulation offered by the DMF. We show that the platform is capable of accurate DNA assembly, efficient transformation, and cell culture and is compatible with many cloning methods (e.g., Golden Gate and Gibson) and chassis organisms (e.g., bacteria, yeast and fungus).

We are also integrating this platform with mass spectrometry to allow sensitive, label-free detection of chemicals and biofuels produced by the engineered cells.

Acknowledgements
The DOE Joint BioEnergy is supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's Nuclear Security Administration under contract DE-AC04-94AL85000.

References
1. Shih SC, Gach PC, Sustarich J, Simmons BA, Adams PD, Singh S, Singh AK. A droplet-to-digital (D2D) microfluidic device for single cell assays. Lab Chip. 2015 Jan 7;15(1):225-36. doi: 10.1039/c4lc00794h.

2. Shih, S. C., Goyal, G., Kim, P. W., Koutsoubelis, N., Keasling, J. D., Adams, P. D., Hillson, N. J., & Singh, A. K. (2015). "A Versatile Microfluidic Device for Automating Synthetic Biology". ACS Synth Biol. doi, 10.1021/acssynbio.5b00062.

3. Gach, Philip; Shih, Steve; Sustarich, Jess; Keasling, Jay; Hillson, Nathan; Adams, Paul; Singh, Anup, 2015, A Droplet Microfluidic Platform for Automating Genetic Engineering, ACS Synth Biol., Accepted.

4. Chieh Chang, Rajiv Bharadwaj, Aarthi Chandrasekaran, Paul D. Adams, Anup K. Singh, Droplet-Based Microfluidic Platform for Heterogeneous Assays: High-Throughput Screening of Cellulases for Biomass Hydrolysis, Lab On A Chip, 2013.

One-pot integrated process for the production of ethanol from lignocellulosic biomass

Jian Sun, Feng Xu, NVSN M. Konda, Tanmoy Dutta, Parthasarathi Ramakrishnan, Corinne Scown, Blake Simmons, Seema Singh

Deconstruction Division, Joint BioEnergy Institute, Emeryville, CA.

Biological and Engineering Sciences Center, Sandia National Laboratories, Livermore, CA, USA.

Abstract

There is a clear and unmet need for a robust biomass conversion technology that can process a wide-range of sustainable feedstocks and produce high yields of fermentable sugars and biofuels with minimal intervention between unit operations. The integration of ionic liquid (IL) pretreatment with enzymatic saccharification and microbial fermentation is challenging due to the toxicity of the ILs currently used for pretreatment, requiring extensive water washes or the development of engineered IL tolerant enzymes and microbes. We demonstrate a one-pot, integrated process for the production of ethanol directly from lignocellulose without removal of IL or any other separation or post-treatment operations prior to saccharification and fermentation. This is achieved through the screening, identification and use of a biocompatible IL, cholinium lysinate, and using carbon dioxide to reversibly control the pH mismatch that historically prevented integration of IL pretreatment with downstream saccharification and fermentation operations into a single unit operation. Also high gravity biomass processing, including IL pretreatment, enzymatic saccharification, and yeast fermentation, was developed and optimized for high-titer cellulosic ethanol production (over 40 g L^{-1}) using a one-pot approach.

These technologies represent a significant development in IL biomass conversion into biofuels by 1) reducing the usage of ionic liquid and water; 2) using CO_2 to reversibly control the pH mismatch that historically prevented integration of IL pretreatment with downstream saccharification and fermentation operations into a single unit operation; and 3) providing

economic benefits for ethanol production at using high biomass loading (25 wt%). Next steps are to work with the Fuels Synthesis Division to incorporate advanced biofuel hosts.

This work aims to overcome the economic and sustainability challenges associated with current ILs based bioprocessing, and was conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

A Strategy to Understand and Improve Microbial Lignin Bioconversion through Solubilization and Separation of the Constituents of the Biopolymer

Chijioke J. Joshua[1,2]* (cjjoshua@lbl.gov), Jijiao Zeng[1,3], Blake A. Simmons[1,2], Ken Sale[1,3] and **Steven W. Singer[1,2]**

[1]Joint BioEnergy Institute, Emeryville, California 94608, [2]Biological and Systems Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, [3]Biological and Engineering Sciences Center, Sandia National Laboratories, Livermore, CA 94551

http://www.jbei.org/research/

Project Goals: We are seeking to enhance the depolymerization and conversion of lignin into advanced biofuels and high-value co-products by developing strategies and analytical tools for understanding the factors responsible for the recalcitrance of this abundant biopolymer. Lignin valorization is an emerging part of Joint Bioenergy Institute's (JBEI) mission to convert lignocellulosic plant biomass into renewable biofuels and biochemicals. The focus of this study is to gain a fundamental understanding of physico-chemical properties that are the origins of the recalcitrance of lignin, and to then use this knowledge to biologically depolymerize the macromolecule. This study complements the extensive research and technology development for deconstructing and converting the polysaccharide component of plant biomass into next generation biofuels that has been developed at JBEI.

Abstract

Lignin depolymerization and conversion to high-value renewable bio-products (valorization) have gained enormous attention in recent years because of the abundance and high energy content of these naturally derived aromatics macromolecules[1]. Lignin valorization is generally limited by the recalcitrance and complexity of the heterogeneous polyphenolic component of plant cell wall[2] and inadequate analytical tools and strategies for rapid and routine characterization of the macro-molecule. Lignin recalcitrance is often attributed to complex C-C or C-O ether bonds[3] that irregularly linked it primary *p*-hydrophenyl (H), guaiacyl (G) and syringyl (S) moieties. However, our studies indicated that physical interactions between the amphiphilic lignin molecules significantly influence the recalcitrance of the macromolecule to chemical and biological depolymerization. Therefore, this study focused on developing strategy for reducing the physical complexity of 'non-derivatized' lignin extracts such as Klason and organosolv lignin extracts. We report the deconvolution and separation of the primary components of Klason lignin from switchgrass by manipulating the amphiphilic properties of the lignin moieties at ambient temperature. Our result revealed that the lignin extract consists of 5 – 8 distinct types of moieties with varying solubility in water at neutral pH. These distinct lignin moieties were also observed in organosolv lignin and lignin extracts from ionic-lignin pretreated and enzymatically saccharified plant biomass. Approximately 80 – 90% of the total lignin in Klason lignin extract from switchgrass was soluble in alkaline solution. The residual alkali-

insoluble lignin moieties consist primarily of very high molecular weight aromatic polymers. A significant fraction of the alkali-soluble lignin fractions (30 – 50%) consist of polydispersed and water-soluble moieties (PW-L02) ranging from ≤ 1 kDa to ≥100 kDa. The PW-L02 fraction was used as substrate to characterize the ability of microbial species to utilize polymeric lignin.

References

1. R. H. White, *Wood and fiber science*, 1987, 19, 446-452.
2. R. Vanholme, B. Demedts, K. Morreel, J. Ralph and W. Boerjan, *Plant physiology*, 2010, 153, 895-905.
3. A. J. Ragauskas, G. T. Beckham, M. J. Biddy, R. Chandra, F. Chen, M. F. Davis, B. H. Davison, R. A. Dixon, P. Gilna, M. Keller, P. Langan, A. K. Naskar, J. N. Saddler, T. J. Tschaplinski, G. A. Tuskan and C. E. Wyman, *Science*, 2014, 344.

This work was funded by DOE Joint BioEnergy Institute (JBEI) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research (DE-AC02-05CH11231)

The Biodegradation of Lignin: harnessing the power of enzymes

Amanda C. Kohler[1,2*] (ACKohler@lbl.gov), Jijiao Zeng[1,2], Michael S. Kent[1,2], Jose Henrique Pereira[1,3], Paul D. Adams[1,3,4], Blake A. Simmons[1,2], and **Kenneth L. Sale**[1,2]

[1]Joint BioEnergy Institute, Emeryville, CA, USA; [2]Sandia National Laboratories, Livermore, CA, USA; [3]Lawerence Berkeley National Laboratory, Berkeley, CA, USA; [4]Department of Bioengineering, University of California Berkeley, Berkeley, CA, USA

http://www.jbei.org

Project Goals: Our objective is to characterize and optimize key lignin degrading enzymes for the development of multi-component enzyme mixtures aimed at producing defined, utilizable end products. Enzymes of particular interest for this research include fungal lignin-acting peroxidases as well as bacterial lignin metabolic enzymes. Thus, two goals of this work are to engineer a versatile peroxidase with enhanced temperature and pH stability, and to functionally and structurally characterize a key *Sphingobium* sp. SYK-6 metabolic enzyme, *O*-demethylase LigM.

Lignin is a key structural component of plant cell walls and one of the most abundant natural polymers on earth. A complex network of aromatic subunits, lignin serves as a rich and renewable source of valuable aromatic compounds, which can be used as precursors for the synthesis of pharmaceuticals, plastics, fuels, and other organic products. However, due to the recalcitrance of lignin, current cellulosic ethanol production methods fail to constructively extract lignin-derived compounds from plant biomass, thus leaving this important commodity untapped. Biodegradation of lignin using lignin-degrading enzymes offers a sustainable method of increasing cellulose availability and releasing low molecular weight aromatic molecules, thus increasing the profitability of biofuel production. We are studying lignin-degrading enzymes, and synergies among them, from a variety of fungi and bacteria that degrade and metabolize lignin. A combination of biochemical analyses, structural biology, and protein engineering techniques are presently being employed to characterize and optimize native enzymes along this pathway. The work presented here focuses on a fungal versatile peroxidase from *Pleurotus ostreatus*, which is influential in the first steps of lignin depolymerization, and on an *O*-demethylase, LigM from the lignin metabolizing bacteria *Sphingobium* sp. SYK-6, involved in late stage lignin degradation – the opening of lignin-derived aromatic rings. LigM is predicted to have a novel fold, as it shares minimal sequence homology with proteins of known structure. Thus, we are characterizing LigM using X-ray crystallography and HPLC-based kinetic assays and currently have LigM protein crystals that diffract to 1.75Å.

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. Crystallographic work was performed at the Advanced Light Source beam line 8.2.2 and was supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Development of Robust, Reproducible, High-Throughput Proteomic Assays for Cellulosic Biofuel Applications

Yan Chen,[1,2] Leanne Jade G. Chan,[1,2] Melissa Nhan,[1,2] Huu Tran,[1,2] Susana M. González Fernández-Niño,[1] Joshua L. Heazlewood,[1] Paul D. Adams,[1-3] and **Christopher J. Petzold**[1,2,*] (cjpetzold@lbl.gov)

[1]Joint BioEnergy Institute, Emeryville, California; [2]Biological Systems Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California; [3]Department of Bioengineering, University of California, Berkeley

http://www.jbei.org

Project Goals:

Over the past 10 years, the bioenergy field has realized significant achievements centered on biosynthetic production of fuel-like compounds. Key to the success of these efforts has been transformational developments in metabolic engineering of biofuel-producing microbes. To aid these efforts, we have developed proteomic methods based on standard flow UHPLC-MS to characterize and quantify complex samples. By using standard flow UHPLC-MS, we were able to routinely identify nearly 800 proteins from *E. coli* samples; while for samples from *Arabidopsis thaliana*, over 1,000 proteins could be reliably identified[1]. To increase sample throughput, we developed automated sample preparation protocols and fast chromatography methods for quantitative targeted proteomic experiments[2]. We shortened the time and minimized the effort to target new proteins by implementing retention time prediction method that allow direct transfer from shotgun proteomics data to short gradient targeted proteomics methods. We demonstrated this workflow on a variety of *E. coli* single gene knockout mutants and other hosts relevant to bioenergy applications.

References

1. González Fernández-Niño, S. M., Smith-Moritz, A. M., Chan, L. J. G., Adams, P. D., Heazlewood, J. L., Petzold, C. J. "Standard flow liquid chromatography for shotgun proteomics in bioenergy research", *Front. Bioeng. Biotechnol.*, 3:44; doi: 10.3389/fbioe.2015.00044.
2. Batth, T. S., Singh, P., Ramakrishnan, V. R., Sousa, M. M. L., Chan, L. J. G., Luning, E. G., Pan, E. Y., Vuu, K. M., Keasling, J. D., Adams, P. D., Petzold, C. J. "A targeted proteomics toolkit for high-throughput absolute quantification of Escherichia coli proteins", *Metab. Eng.* 2014, 26, 48-56. doi: 10.1016/j.ymben.2014.08.004.

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

A Cas9-Based Toolkit to Program Expression Context in *Saccharomyces cerevisiae*

Amanda Reider Apel[1,2]* (arreiderapel@lbl.gov), Leo d'Espaux[1,2], Maren Wehrs[1,2], Daniel Sachs[1,2], Megan Garber[1,2], Nathan Hilson[1,2], Seema Singh[1,2,3], Blake Simmons[1,2,3], **Jay D. Keasling**[1,2,4,5], and **Aindrila Mukhopadhyay**[1,2]

[1] Joint BioEnergy Institute, 5885 Hollis St, Emeryville, CA 94608, USA; [2] Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA; [3]Biological and Materials Science Center, Sandia National Laboratories, 7011 East Avenue, Livermore, CA 94551, USA; [4]Department of Bioengineering, University of California, Berkeley, CA 94720, USA; and [5]Department of Chemical & Biomolecular Engineering, University of California, Berkeley, CA 94720, USA

http://www.jbei.org

Project Goals: The Joint BioEnergy Institute (JBEI) is focused on providing clean, renewable transportation fuels identical to gasoline, diesel and jet fuel. In order to produce fuels at high titers, we need to have well-characterized expression "parts" so that we can easily and quickly modulate pathway genes. Additionally, for this process to be renewable, our engineered microbes must also be able to grow and produce fuels from biomass hydrolysates. To address these problems, we employed an integrated research approach, which is illustrated below.

Engineering of biological systems can be difficult to predict due to the vast complexity of living cells. In order to control chemical production, it is necessary to regulate a number of variables including DNA copy number, transcriptional timing, transcript stability, protein solubility and localization. *Saccharomyces cerevisiae* is an excellent organism for industrial production of biological molecules, and has many well-characterized tools available for genetic engineering. While the potential for the use of these technologies is immense, specific applications are often hampered by slow development times and unpredictability. Here we report on developing a Cas9-based toolkit to quickly institute genetic changes in yeast and to program heterologous gene expression. We characterize gene expression "parts" using a fluorescent reporter protein, exploring how chromosomal integration locus and promoter affect expression levels and dynamics. We further develop protein tags to control gene expression level, solubility, and sub-cellular localization. In parallel, we develop the organism as a microbial production platform for chemicals with potential uses as biofuels. We focused on the fatty acid biosynthetic pathway that produces large, hydrophobic molecules similar to modern fuels. Titers of biofuels produced by our engineered yeasts approach the g/L range from simple sugars. Our yeast strains

are also capable of high-level production in biomass hydrolysates, a renewable production platform.

The Joint BioEnergy Institute is a U.S. Department of Energy Bioenergy Research Center supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

Development of Plants with Multiple Traits for High Yield of Fermentable Sugars

Aude Aznar[1], Camille Chalvin[1,2], Aymerick Eudes[1], Khanh Vuu[1], Mi Yeon Lee[1], Patrick Shih[1], Vibe Marie Gondolf[1,3], Berit Ebert[1,3], Devon Birdseye[1], Dominique Loqué[1], **Henrik Vibe Scheller**[1,4] *(hscheller@lbl.gov)

[1]Feedstocks Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, California 94720; [2]Ecole Normale Supérieure de Cachan, 94230 Cachan, France; [3]University of Copenhagen, Department of Plant and Environmental Sciences, Frederiksberg, DK-1871, Denmark; [4]University of California, Department of Plant and Microbial Biology, Berkeley, CA, 94720

http://www.jbei.org

Project Goals: Development of bioenergy crops with increased ratio between hexoses and pentoses, low lignin content and low recalcitrance.

Second-generation biofuels produced from biomass can help to decrease dependency on fossil fuels, which would have many economical and environmental benefits. To make biomass more suitable for biorefinery use we need a better understanding of plant cell wall biosynthesis. Increasing the ratio of C6 to C5 sugars in the wall is an important target for engineering of plants that are more suitable for downstream processing for second-generation biofuel production. Likewise, decreasing the content of lignin is an important goal. We have studied the basic mechanisms of cell wall biosynthesis and identified genes involved in biosynthesis of pectic galactan including the GALS1 galactan synthase[1] and the URGT1 UDP-galactose transporter[2]. We have applied these findings to engineer plants that have a more suitable biomass composition and have developed synthetic biology and gene stacking tools to achieve this goal. Plants were engineered to have up to three-fold increased content of pectic galactan in stems by expressing GALS1, URGT1 and a UDP-glucose epimerase. Furthermore, the increased galactan was engineered into plants that were already engineered to have low xylan content by restricting xylan biosynthesis to vessels where this polysaccharide is essential[3]. Finally, the high galactan and low xylan traits were stacked with low lignin obtained by expressing the *QsuB* gene encoding dehydroshikimate dehydratase[4]. By targeting the transgene expression to specific cell types, we could substantially improve saccharification while avoiding adverse effects on plant growth and development.

References

1. Liwanag, A.J., et al., Pectin biosynthesis: GALS1 in Arabidopsis thaliana is a beta-1,4-galactan beta-1,4-galactosyltransferase. Plant Cell, 2012. 24(12): p. 5024-36.

2. Rautengarten, C., et al., The Golgi localized bifunctional UDP-rhamnose/UDP-galactose transporter family of Arabidopsis. Proc Natl Acad Sci U S A, 2014. 111(31): p. 11563-8.

3. Petersen, P.D., et al., Engineering of plants with improved properties as biofuels feedstocks by vessel-specific complementation of xylan biosynthesis mutants. Biotechnology for Biofuels, 2012. 5(1): p. 84.

4. Eudes, A., et al., Expression of a bacterial 3-dehydroshikimate dehydratase reduces lignin content and improves biomass saccharification efficiency. Plant Biotechnol J, 2015.

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

Engineering Mannan Biosynthesis in Plants

Thea Pick[1], Amir Mahboubi[1], Vy Ngo[1], Soe Htwe[1], Jeemeng Lao[1], **Jenny C. Mortimer**[1]* (jcmortimer@lbl.gov)

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA

http://www.jbei.org

The goal of this project is to increase the accessible hexose content of the plant cell wall. This will increase the fermentation efficiency of lignocellulosic biomass into renewable transportation fuels.

Plant biomass serves as an abundant renewable source for biofuel and bioenergy production. As energy demand steadily increases the competing priorities between energy production needs and food supply will grow significantly. Therefore, it is critical that biomass is sourced from sustainable, non-food sources - namely lignocellulosic biomass. However, fermentation of lignocellulosic material is currently costly, both economically and environmentally. One approach to improve fermentation of lignocellulosic feedstocks is to increase the hexose sugar content in plant cell walls, since the majority of microorganisms favor hexoses over pentoses for fermentation. Plant cell walls are composed of cellulose, hemicellulose, and lignin, of which only cellulose and a small fraction of the hemicelluloses contain C6 sugars. In cellulose these sugars are packed in crystalline arrays and are least accessible to microbial fermentation. The C6 sugars in hemicelluloses like mannans are more accessible to enzymatic hydrolysis. Our goal is to increase the content of mannans in the plant cell wall to produce plants with enhanced bio refinery feedstock characteristics, meaning increased biofuel production through improved fermentation.

The Cellulose Synthase-like A (CslA) proteins have been identified as mannan synthesizing enzymes in plants. However, overexpression of theses enzymes *in planta* resulted in only a minor increase in cell wall mannan. Therefore, we hypothesize that we may need to alter other aspects of mannan biosynthesis, such as substrate pools, mannan solubility and enzyme efficiency in order to achieve a meaningful boost in cell wall hexose content.

We are using a systems approach to understand the factors which control mannan biosynthesis. Here we will present our approach to targeting different aspects of mannan biosynthesis, some preliminary data, and our future plans to integrate these into a single plant.

Funding statement.

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

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

The glycosylation of plant sphingolipids affects cellulose crystallization, plant defense signaling and nitrogen fixation ability

Lin Fang[1]* (linfang@lbl.gov), Toshiki Ishikawa[2], Emilie A. Rennie[1], William Moore[1], Jeemeng Lao[1], Heidi Wipf[1], Edward Baidoo[1], Jay Keasling[1], Taku Demura[3,4], Maki Kawai-Yamada[2], Henrik V. Scheller[1] and **Jenny C. Mortimer[1,3]**

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, California; [2]Saitama University, Japan; [3]Cellulose Production Research Team, BMEP, CSRS, RIKEN Yokohama, Japan; and [4]NAIST, Japan

http://www.jbei.org

Project goal: The aim of this project was to characterize plant Golgi glycosyltransferases (GTs), as the majority of these proteins are involved in cell wall biosynthesis. The understanding of the plant cell wall is critical to understanding feedstock recalcitrance and enabling the predictive engineering of biomass. However, the GT we characterized turned out to have a role in sphingolipid glycosylation, and in turn to affect cellulose crystallinity. This project has enabled us to discover an unexpected role for membrane lipids in cellulose biosynthesis.

The plasma membrane acts as an interface between the plant cell wall and the inside of the cell. It is the site of cellulose biosynthesis and a range of signal transduction complexes. Glycosylinositol phosphorylceramides (GIPCs) are a class of glycosylated sphingolipids found only in plants, fungi and protozoa. They are extremely abundant in the plant plasma membrane, estimated to form ~25% of the total lipids. Little is known about the glycosylated headgroup, but two recent papers have indicated a key role in plant signaling and defense, and shown that it is synthesized in the Golgi. The Golgi apparatus is also the site of cell wall polysaccharide biosynthesis, with the exception of cellulose and callose, but it is not clear how pools of substrates are directed to different glycans. Here, we identify a Golgi-localized *Arabidopsis thaliana* glycosyltransferase, GIPC MANNOSYL TRANSFERASE1 (AtGMT1), and demonstrate that it is a GIPC-specific mannosyl-transferase.

Previously, we identified a GIPC-specific GDP-mannose transporter mutant *gonst1*, which had dwarfed stature and a constitutive defense response[1]. We obtained three alleles of *gmt1*, which displayed a very similar phenotype to *gonst1*. Sphingolipid analysis revealed that *gmt1* almost completely lacked hexosylated GIPCs. *gmt1* has elevated production of salicylic acid and H_2O_2, suggesting that GIPC sugar decoration plays a role in plant defense signaling. Unexpectedly, we also found a reduction in crystalline cellulose content in the *gmt1* plants, suggesting significant misglycosylation can impact cellulose crystallinity. We are now investigating the role of these proteins in other species, and have found that a homolog of AtGMT1 in *Medicago truncatula*

may have a role in nodulation and nitrogen fixation. Our future work will focus on understanding how GIPC glycosylation is able to control these varied phenotypes.

References

1.Mortimer, J.C., Yu, X., Albrecht, S., Sicilia, F., Huichalaf, M., Ampuero, D., Michaelson, L.V., Murphy, A.M., Matsunaga, T., and Kurz, S. (2013). Abnormal Glycosphingolipid Mannosylation Triggers Salicylic Acid–Mediated Responses in Arabidopsis. The Plant Cell Online 25, 1881-1894.

Funding statement.

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

Fungal Biotechnology: How can we improve enzyme production?

Morgann C. Reilly[1]* (mcreilly@lbl.gov), Saori Amaike Campen[1], Jinxiang Zhang[1], Joonhoon Kim[1], Junko Yaegashi[1], Yvette Tran[1], James Kirby[1], Jeffrey M. Skerker[2], Wendy S. Schackwitz[3], Joel Martin[3], Jon M. Jacobs[4], Chee-Hong Wong[3], Blake A. Simmons[1], Scott E. Baker[1,4], John M. Gladden[1], and **Jon K. Magnuson[1]**

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA; [2]University of California-Berkeley and Energy Biosciences Institute, Berkeley, CA; [3]Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, CA; and [4]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA

Project Goals: Development of recombinant protein production platforms in fungi for the production of ionic liquid-tolerant lignocellulolytic enzymes at high titers.

Efficient and economical deconstruction of biomass is critical for the success of lignocellulosic biorefineries. Biomass pretreatment with ionic liquids (ILs) tackles this issue as it increases biomass saccharification efficiency at lower cellulase loadings. However, some ILs inhibit the activity of commercial cellulases and must first be removed from the biomass, a costly mitigation.

To overcome this issue, the Microbial Communities and Enzyme Optimization teams at JBEI have identified cellulases that can function in the presence of ILs. These enzymes were expressed in *Escherichia coli,* which is a suitable host for lab-scale enzyme characterization but not for industrial-scale production of enzymes.

Filamentous fungi have been widely utilized for enzyme production in industry and therefore the Fungal Biotechnology team has focused on developing *Aspergillus niger* as a heterologous enzyme production host. Several approaches have been utilized to increase enzyme production in the fungus: 1) reverse genetics, 2) forward genetics and mutagenesis, 3) genetics parts development, and 4) characterization of a wide variety of heterologously expressed enzymes. We are also expanding our efforts to include fungal lignin deconstruction and conversion into advanced bioproducts.

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

The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

A portion of the research was performed using EMSL, a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory.

A robust gene stacking method utilizing yeast assembly for plant synthetic biology

Patrick Shih[1]* (pmshih@lbl.gov), Nasim Mansoori,[1] Khanh Vuu,[1] Leïla Ayad,[1] Katherine Louie,[2] Ben Bowen,[2] **Trent Northen**,[1,2] **Dominique Loqué**[1]

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA

Project Goals:
Plant synthetic biology efforts have been hampered by a dearth of versatile transformation vectors, DNA part libraries, and efficient assembly strategies. Here, we present a strategy utilizing *in vivo* yeast homologous recombination to assemble multiple gene cassettes to facilitate plant metabolic engineering, which we have named jSTACK. In doing so, we have also generated a library of DNA parts consisting of promoters, genes, and terminators that will be publicly available as a resource to the plant synthetic biology community. We demonstrate how this method can facilitate pathway engineering of molecules of pharmaceutical interest, production of potential biofuels, and shuffling of disease resistance traits between crop species. *In vivo* homologous recombination has been leveraged for the large-scale DNA assembly of synthetic chromosomes and genomes in microbes. Likewise, our approach extends this technology to plants, providing a powerful alternative to conventional approaches for stacking traits and genes to address many impending agricultural challenges.

This work was part of the DOE Early Career Award and the DOE Joint BioEnergy Institute supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research; and U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility through contract DEAC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

Exploiting The Substrate Promiscuity of Hydroxycinnamoyl-CoA:shikimate Hydroxycinnamoyl Transferase to Reduce Lignin

Aymerick Eudes[1*] (ageudes@lbl.gov), Jose H. Pereira,[1] Sasha Yogiswara,[1,2] George Wang,[1] Veronica Teixeira Benites,[1] Edward E. K. Baidoo,[1] **Taek Soon Lee**,[1] **Paul D. Adams,**[1], **Jay D. Keasling,**[1,2] **and Dominique Loqué**[1]

[1]Joint BioEnergy Institute, Emeryville, California; [2]University of California, Berkeley

http://www.jbei.org/

Project Goals: The most abundant organic material on earth is lignocellulosic biomass or non-food plant fiber. JBEI's mission is to convert biomass to biofuels. The goal is to provide the nation with clean, renewable transportation fuels identical to gasoline, diesel and jet fuel. Inside JBEI's Emeryville laboratories, researchers are using the latest tools in molecular biology, chemical engineering, computational and robotic technologies, and pioneering work in synthetic biology to transform biomass sugars into energy-rich fuels. One of the goals of the cell wall engineering team in JBEI's Feedtstocks Division is to engineer lignin in plants for reducing biomass recalcitrance.

Lignin poses a major challenge in the processing of plant biomass for agro-industrial applications. For bioengineering purposes, there is a pressing interest in identifying and characterizing the enzymes responsible for the biosynthesis of lignin [1]. Hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase (HCT, EC 2.3.1.133) is a key metabolic entry point for the synthesis of the most important lignin monomers: coniferyl and sinapyl alcohols [2]. In this study, we investigated the substrate promiscuity of HCT from a bryophyte (Physcomitrella) and from five representatives of vascular plants (Arabidopsis, poplar, switchgrass, pine, and Selaginella) using a yeast expression system. We demonstrate for these HCTs a conserved capacity to acylate with *p*-coumaroyl-CoA several phenolic compounds in addition to the canonical acceptor shikimate normally used during lignin biosynthesis. Using either recombinant HCT from switchgrass (PvHCT2a) or Arabidopsis and switchgrass stem protein extracts, we show evidence of the inhibitory effect of these phenolics on the synthesis of *p*-coumaroyl shikimate in vitro, which presumably occurs via a mechanism of competitive inhibition. Structural study of PvHCT2a confirmed the binding of a non-canonical acceptor in a similar manner as shikimate in the active site of the enzyme. Finally, we exploited in Arabidopsis the substrate flexibility of HCT to reduce lignin content and improve biomass saccharification by engineering transgenic lines that overproduce one of the HCT non-canonical acceptors. Our results demonstrate conservation of HCT substrate promiscuity and provide support for a new strategy for lignin reduction in the effort to improve the quality of plant biomass for forage and cellulosic biofuels.

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

Integrated BioCAD Toolchain Enables Search for Experimentally Validated Components

Hector A. Plahar,[1,2,3] Mark Forrer,[1,2,4] Jacob Coble,[1,2,3] William Morrell,[1,2,4] Jay D. Keasling,[1,3] Paul D. Adams,[2,5] Hector Garcia Martin,[1,3] and **Nathan J. Hillson**[1,2,3,6]* (njhillson@lbl.gov)

[1]Fuels Synthesis and [2]Technology Divisions, DOE Joint BioEnergy Institute, Emeryville, CA; [3]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [4]Applied Biosciences and Engineering Center, Sandia National Laboratory, Livermore, CA; [5]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [6]DOE Joint Genome Institute, Walnut Creek, CA.

Project Goals:

Biological computer-aided design (bioCAD) tools are gaining widespread adoption in commercial, academic, and government research settings. As more and more tools emerge, there is an increasingly urgent need to integrate them into a larger toolchain that aggregates functionality and data. Researchers at JBEI engineer plants, enzymes, and microbes to produce biofuels. To enable them to design biological systems without knowing the intricate details of the constituent components, we have integrated several of our bioCAD tools (DeviceEditor, DIVA, ICE, EDD Analytics, and EDD). The specifications of these biological components are experimentally validated to meet particular design criteria (e.g., promoters with high transcriptional levels in *E. coli*).

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

Systematic Metabolic Flux Modeling Techniques and Applications at the Joint BioEnergy Institute

David Ando[1]* (david.ando@lbl.gov), Amit Ghosh,[1] Weerawat Runguphan,[1] Charles Denby,[1] Chris Shymansky,[1] Yasuo Yoshikuni,[2] Jay D Keasling,[1] and **Hector Garcia Martin**[1]

[1]Joint BioEnergy Institute, LBNL, Emeryville, CA; [2]Joint Genome Institute, LBNL, Walnut Creek, CA

Project Goals: The Joint BioEnergy Institute is focused on developing advanced second generation biofuels which can replace gasoline, diesel and jet fuels while providing significant environmental benefits in the form of reduced global dependence on crude oil and minimizing CO2 production. Our goal is to apply genome scale metabolic flux analysis and other modeling techniques to assist metabolic engineering efforts in fatty acid-derived biofuel production. Our analysis provides the necessary insights into cellular metabolism required for systematic genetic engineering efforts, currently increasing fatty acid production by over 40%. Similarly, our goal is to complete a joint project with JGI to study how substituting the enzymes responsible for most of NADPH production by their equivalents producing NADH in an isopentenol (from the mevalonate pathway) producing strain of *S. cerevisiae* S288c affects targeted production. We find that an optimal set of cofactor swaps is predicted to increase targeted production more than 10x that of the reference strain.

The Joint BioEnergy Institute is a DOE Research Center focused on developing advanced second generation biofuels which can replace gasoline, diesel and jet fuels while providing significant environmental benefits in the form of reduced dependence on crude oil and minimizing CO2 production. In order to produce economically and industrially feasible biofuels it is desirable to be able to systematically determine genetic modifications in biofuel producing organisms which may improve yield. Fluxes describe the flow of molecules through a metabolic pathway, with flux-based metabolic modeling being particularly well suited to providing the insights required in finding genetic modifications which increase biofuel yields since it can provide a global view on how carbon flows from feed to final biofuel product in an organism, and where potential bottlenecks lie. Flux Balance Analysis (FBA) has previously been used successfully for this purpose, which obtains fluxes by using a network of cellular metabolism which includes all reactions coupled with a linear programming assumption that metabolism is tuned, due to evolutionary pressure, to maximize growth rate (or other evolutionary assumptions can be used). Two scale 13C Metabolic Flux Analysis (2S-13C MFA) improves on FBA by retaining the genome scale metabolic network but drops the evolutionary assumption in favor of 13C

constraints from cellular metabolites measured experimentally and applied to the metabolic core of the model[1].

We have applied 2S-13C MFA towards improving production of fatty acids through a biosynthetic pathway developed earlier by Runguphan et al[2]. First we performed 13C tracer experiments and used 2S-13C MFA to measure fluxes both before and after boosting acetyl-CoA production via the addition of the ATP citrate lyase enzyme (ACL). Although acetyl-CoA is the substrate for fatty acid production the introduction of ACL resulted in only a small gain in fatty acid production. 2S-13C MFA identified the most significant sink of acetyl-CoA after the introduction of ACL to be Malate synthesis (MALS). Further downregulating MALS resulted in a more significant increase in fatty acid production of roughly 30%. Finally, as fatty acid production increased as we further engineered for higher fatty acid production 2S-13C showed that the Glycerol-3-phosphate dehydrogenase pathway, which competes for carbon with the acetyl-CoA production pathway, had a carbon flux which similarly increased. We downregulated cytoplasmic Glycerol-3-phosphate dehydrogenase in our engineered strains so more carbon flux would be available for fatty acid production and as expected these strains showed an increase in fatty acid production.

Additionally, we have used Flux Balance Analysis and Minimization of Metabolic Adjustment to determine how different possible combinations of cofactor swaps among a set of NADPH reductases affects isopentenol (from the mevalonate pathway) production in S. cerevisiae strain S288c. In a joint project between JBEI and JGI, we are substituting the enzymes responsible for most of NADPH production by their equivalents producing NADH, with the expectation that the excess NADH will increase isopentenol production due to the reengineered mevalonate pathway's dependence on NADH. Our metabolic modeling simulations support this expectation, with an optimal set of cofactor swaps being predicted to increase targeted production more than 10x that of the reference strain.

References

1. Héctor García Martín, Vinay Satish Kumar, Daniel Weaver, Amit Ghosh, Victor Chubukov, Aindrila Mukhopadhyay, Adam Arkin, Jay D. Keasling. A Method to Constrain Genome-Scale Models with 13C Labeling Data. PLOS Computational Biology 11, e1004363 (2015).
2. Weerawat Runguphan, Jay D. Keasling. Metabolic engineering of Saccharomyces cerevisiae for production of fatty acid-derived biofuels and chemicals. Metabolic Engineering 21, 103–113 (2014).

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

Advances in Ionic Liquid Pretreatment Technologies: Impact on the Economics of Cellulosic Biofuel Production

N.V.S.N. Murthy Konda[1,2*] (MurthyKonda@lbl.gov), Feng Xu,[1,3] Jian Sun,[1,3] Seema Singh,[1,3] Blake A. Simmons,[1,3] and **Corinne D. Scown**[1,2]

[1]Joint BioEnergy Institute, 5885 Hollis St. Emeryville, California 94608

[2]Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720

[3]Sandia National Laboratories, Biological and Material Science Center, Livermore, California 94551

[*]corresponding author

http://www.jbei.org/

Project Goals: One of the main goals of JBEI's Deconstruction Division is to develop efficient, scalable, and economically viable pretreatment technologies. Towards this goal, the technoeconomic analysis (TEA) group has been developing technoeconomic models for lignocellulosic biorefineries to understand the technical potential and cost impacts of novel ionic liquid (IL) pretreatment process configurations currently under development at JBEI.

Abstract.

Our previous work (Konda et al., 2014) has shown that the traditional IL pretreatment configuration – referred to as the water-wash (WW) route – presents environmental challenges (due to excessive water usage during the water-wash step) and economic challenges (due to wastewater treatment costs). In an effort to move away from this WW configuration, the Pretreatment group (Shi et al., 2014) developed a novel process in which the water-wash step was eliminated. This simplification was possible with the use of an IL-tolerant enzyme mixture (JTherm), which eliminated the need for IL removal prior to hydrolysis, and is referred to as JTherm process in this study. In a previous study (Konda et al., 2014), we showed that the JTherm process, while successful in reducing water demand, required further improvements to achieve significant cost savings relative to the WW route. This was largely due to the lack of economically viable technologies to extract sugars from the resulting hydrolysate. To address this challenge, JBEI's Pretreatment group has recently developed a novel one-pot (OP) process, using biocompatible ILs, that does not require any separation steps (i.e., water-wash or sugar extraction) between pretreatment, hydrolysis, and fermentation. Furthermore, this process is capable of handling 'high gravity' conditions (i.e., >20% solids loading) and thus referred to as the OP-HG process. In this work, we have evaluated the economic viability of this OP-HG

process if it were to be scaled up to industrial throughputs (e.g., to process 2000 MT/day of dry biomass). The economic analysis utilizes parameters consistent with the NREL study (Humbird et al., 2011) and the minimum ethanol selling price was (MESP) computed based on a detailed cash flow analysis using a 10% internal rate of return (IRR). Two scenarios ('current' and 'projected') were constructed to quantitatively evaluate the economic potential in terms of MESP. Based on the modeled scenarios, the OP-HG process was found to be a promising configuration, capable of lowering the MESP below $3/gal. We found that the main cost bottleneck for the OP-HG system is currently the fact that xylose sugars are not being fermented, reducing overall ethanol yield – co-utilization of both the glucose and xylose sugars is necessary to realize full potential of this system. In future work, the TEA group will conduct more rigorous process engineering and feasibility assessment for downstream processes in the OP-HG system.

References:

1. Humbird et al., (2011) Process Design and Economics for Biochemical Conversion of Lignocellulosic Biomass to Ethanol Dilute-Acid Pretreatment and Enzymatic Hydrolysis of Corn Stover. NREL.
2. Konda, NVSN Murthy, et al. "Understanding cost drivers and economic potential of two variants of ionic liquid pretreatment for cellulosic biofuel production." Biotechnol Biofuels 7.1 (2014): 86.
3. Shi et al. (2013) One-pot Ionic Liquid Pretreatment and Saccharification of Switchgrass. Green Chemistry, 2579-2589.

Funding statement.

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

Isoprenoid Biofuels Research at JBEI

Aram Kang (akang@lbl.gov) *, Jorge Alonso-Gutierrez, Daniel Mendez-Perez, Kevin W. George, Edward Baidoo, Christopher J. Petzold, Paul Adams, Jay D. Keasling and **Taek Soon Lee**

Fuels Synthesis Division & Technology Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, California

http://www.jbei.org

Project Goals: The Joint BioEnergy Institute (JBEI) aims to produce a chemically diverse suite of biofuels from lignocellulosic biomass. Among the biofuels compounds, isoprenoids have low freezing temperature and high octane number, which makes them promising biofuels. Mevalonate (MVA) pathway is one of the major biosynthetic pathways of isoprenoids biofuels, and the engineering of this pathway is a key approach to achieve higher production of these biofuels. JBEI's approaches to improve isoprenoids biofuel production involves understanding pathway enzymes, identifying bottlenecks and optimizing the pathway, which has been accelerated by development of various engineering strategies, analytical tools and mathematical models.

Isoprenoids are the largest and most diverse group of natural products. They are commonly used for medicinal purposes but their high energy density and low freezing point (due to methyl branching and cyclic structure often found in their carbon skeleton) also make them good candidates for gasoline, diesel and jet fuel replacements. Various isoprenoids are all synthesized from the two universal C_5 building blocks: isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). Using an engineered *Escherichia coli* strain for the overproduction of IPP and DMAPP via the mevalonate pathway, we have produced isopentenol (C_5), limonene, pinene (C_{10}), cineole (C_{10}) and bisabolene (C_{15}), achieving more than 2 g/L of isopentenol, 1.1 g/L of bisabolene, and 0.5-0.6 g/L of limonene and 1,8-cineole. These improved titers were achieved by various strategies and analytical tools including omics analysis (e.g. proteomics and metabolomics to quantify key enzymes and intermediate metabolites), mathematical models to interpret the data. In addition, modified IPP-bypassing MVA pathways have been developed to reduce intrinsic energy demands of the MVA pathway and subsequently the aeration cost in larger scale fermenters, which occupies a significant portion of overall operational cost. We are currently developing Design-Build-Test-Learn (DBTL) cycle based on rational engineering approaches used for improving isoprenoid biofuels production in JBEI, and the optimized DBTL cycle will also be applied to optimize new target biofuel pathways and molecules.

Funding statement.

This JBEI work was supported by the U.S. DOE under Contract No. DE-AC02-05CH11231

Fatty acid-related research at JBEI: methyl ketones and branched fatty acids in *E. coli*

Ee-Been Goh[1]* (EGoh@lbl.gov), Robert Haushalter,[1] Helcio Burd,[1] Taek Soon Lee,[1] Hector Garcia Martin,[1] Jay D. Keasling,[1] and **Harry R. Beller[1]**

[1]Joint BioEnergy Institute (JBEI), Emeryville, CA

http://www.jbei.org/research/divisions/fuels-synthesis/

Project Goals: The Joint BioEnergy Institute (JBEI) aims to produce a chemically diverse suite of biofuels from lignocellulosic biomass. Some JBEI fuels use fatty acids as precursors, as these biomolecules are highly reduced, aliphatic compounds that, when modified (e.g., decarboxylated), can have properties comparable to those of petroleum-derived fuel components. The goals of the project presented here is to (1) engineer *E. coli* to produce diesel-range methyl ketones in the gram-per-liter range with yields of at least 40% of maximum theoretical yield, and (2) improve the cold-temperature properties of these fatty-acid derived compounds by incorporating methyl-branching.

Accelerated research-and-development activity in biofuels in recent years has facilitated the development of metabolic pathways that enable biochemical conversion of fatty acids (and intermediates of fatty acid biosynthesis) to a range of industrially relevant compound classes, including aliphatic methyl ketones (Beller et al. 2015). Here, we describe two fatty acid-related projects at JBEI in which we engineered *Escherichia coli* for improved biofuel performance: (1) efforts to improve the titers, rates, and yields (TRY) of diesel-range methyl ketones (a class of fatty acid-derived molecules that have high cetane numbers), and (2) efforts to produce methyl-branched fatty acids, which would yield diesel fuels with better cold-temperature properties than straight-chain fatty acids.

Methyl ketones: We have engineered *E. coli* to overproduce aliphatic methyl ketones (MK) in the C_{11} to C_{15} (diesel) range; this group of MK includes 2-undecanone and 2-tridecanone, which have favorable cetane numbers and are also of importance to the flavor and fragrance industry. Overall, we have made specific improvements that resulted in more than 10,000-fold enhancement in MK titer relative to that of a fatty acid-overproducing *E. coli* strain. The first generation of engineered *E. coli* (Goh et al., 2012) produced ~380 mg/L of MK in rich medium and had modifications that included: (a) overproduction of β-ketoacyl-coenzyme A (CoA) thioesters achieved by modification of the β-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). We have subsequently made additional genetic modifications that included balancing overexpression of *fadR* and *fadD* to increase fatty acid flux into the pathway, consolidation of the pathway from two plasmids into one, codon optimization, and knocking out key acetate production pathways (Goh et al., 2014). These latest modifications have led to a MK titer of 1.4 g/L with 1% glucose in shake flask experiments, which represents

40% of the maximum theoretical yield, and also attained titers of 3.4 g/L after ~45 h of fed-batch glucose fermentation (the best values reported to date).

Using another approach for improving MK production, metabolic modeling was used to identify gene deletions that could improve flux through the MK pathway. One of the specified knockouts, Δ*scgE*, which was annotated as a homolog of Rpe (ribulose-5-phosphate epimerase) in the pentose phosphate pathway, improved MK production by >50% relative to the base strain (EGS1710). ^{13}C-glucose experiments have recently been performed on the knockout strain along with the DH1 wild type and base strain (EGS1710) to obtain more comprehensive metabolic flux profiles that will refine the genome-scale model and enable additional improvements in MK production.

Branched fatty acids: We have engineered *E. coli* to produce *iso*- and *anteiso*-branched fatty acids. From a biofuel perspective, the straight-chain fatty acids produced by wild-type *E. coli* are not ideal because they have relatively high melting points. Although *E. coli* generates unsaturated fatty acids, which have far lower melting points than their saturated analogs, the unsaturated fatty acids are more susceptible to oxidation during storage. Other bacteria, such as *Bacillus subtilis*, naturally produce branched fatty acids, rather than unsaturated fatty acids, to modulate membrane fluidity. We have retooled *E. coli* fatty acid biosynthesis with heterologous enzymes to produce > 20% *anteiso*-branched fatty acids, which exhibit lower melting points than straight-chain and *iso*-branched isomers (Haushalter et al., 2014). The ability to synthesize methyl-branched fatty acids in *E. coli* will be important for optimizing fatty acid-derived biofuels, such as methyl ketones, for use in cold climates.

References

1. Beller, H. R., Lee, T. S., & Katz, L. (2015). Natural products as biofuels and bio-based chemicals: fatty acids and isoprenoids. *Natural Product Reports*. **32**:1508-1526.
2. Goh, E.-B., Baidoo, E. E. K., Keasling, J. D., & Beller, H. R. (2012). Engineering of bacterial methyl ketone synthesis for biofuels. *Appl. Environ. Microbiol.* **78**:70–80.
3. Goh, E.-B., Baidoo, E. E. K., Burd, H., Lee, T. S., Keasling, J. D., & Beller, H. R. (2014). Substantial improvements in methyl ketone production in *E. coli* and insights on the pathway from *in vitro* studies. *Metabolic Engineering*. **26C**:67–76.
4. Haushalter, R. W., Kim, W., Chavkin, T. A., The, L., Garber, M. E., Nhan, M., … Keasling, J. D. (2014). Production of *anteiso*-branched fatty acids in *Escherichia coli*; next generation biofuels with improved cold-flow properties. *Metabolic Engineering*. **26C**:111–118.

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.

Structural Study of Hydroxycinnamoyl-CoA:shikimate Hydroxycinnamoyl Transferase

Jose H. Pereira[1,2,*] (jhpereira@lbl.gov), Aymerick Eudes[1,3], Sasha Yogiswara[1,3,4], George Wang[1,3], Veronica Teixeira Benites[1,3,5], Edward E. K. Baidoo[1,3], Taek Soon Lee[1,3], Jay D. Keasling[1,3,4], Dominique Loqué[1,3] and **Paul D. Adams[1,2]**

[1]Joint BioEnergy Institute, EmeryStation East, 5885 Hollis St, 4th Floor, Emeryville, CA 94608, USA. [2]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA. [3]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA. [4]Department of Bioengineering & Department of Chemical & Biomolecular Engineering, University of California, Berkeley, CA 94720, USA. [5]Graduate Program, San Francisco State University, San Francisco, CA 94132, USA.

Project Goals: Understand structurally the substrate promiscuity of Hydroxycinnamoyl-CoA:shikimate Hydroxycinnamoyl Transferase (HCT) from a *Panicum virgatum* (PvHCT) toward a series of benzene and benzoate derivatives. The high-resolution crystal structures of PvHCT-*p*-coumaroyl-CoA-shikimate and PvHCT-*p*-coumaroyl-CoA-protocatechuate confirmed the binding of a non-canonical acceptor in a similar manner as shikimate in the active site of the enzyme.

Lignin poses a crucial challenge in the processing of plant biomass for agro-industrial applications. For bioengineering purposes, there is a pressing interest in identifying and characterizing the enzymes responsible for the biosynthesis of lignin. Hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase (HCT) is a key metabolic entry point for the synthesis of the most important lignin monomers: coniferyl and sinapyl alcohols. In this study, we investigated structurally the substrate promiscuity of HCT from *Panicum virgatum*. The crystal structure of PvHCT consists of two domains, with the *p*-coumaryl-CoA and shikimate binding sites located between them. Domain I consists of N-terminal residues 1 to 199 and C-terminal residues 387 to 409. Domain II consists exclusively of C-terminal residues 200 to 386 and 410 to 446. The structure of PvHCT2 in complex with *p*-coumaroyl-CoA and shikimate reveals that the two molecules reacted during the soaking of the compounds into the crystal. Therefore, the product state was observed in the electron density map, given the ternary complex of PvHCT, free coenzyme A and *p*-coumaroyl shikimate. In contrast, the crystal structure of PvHCT-*p*-coumaroyl-CoA-protocatechuate is in a substrate state, i.e. the ternary complex consists of PvHCT, *p*-coumaroyl-CoA and protocatechuate.

The *p*-coumaroyl-shikimate contacts the PvHCT via the phenolic group and carbonyl group of the *p*-coumaroyl portion. The shikimate portion contacts PvHCT through both carboxyl and hydroxyl groups. The carbonyl group of the *p*-coumaroyl moiety directly contacts Trp384 and the phenolic moiety interacts via water-mediated hydrogen bonds with Ser38 and Tyr40. The carboxyl group of the shikimate moiety makes a salt bridge interaction with Arg369. While the C5 hydroxyl group contacts the catalytic residue His163, the C3 hydroxyl group contacts Thr382. PvHCT-*p*-coumaroyl-CoA-protocatechuate structure shows that the protocatechuate binds in a very similar manner as shikimate, with the carboxyl group making a tight salt bridge with Arg369, and the C3 (equivalent to C5 in shikimate moiety) hydroxyl group interacting with the nitrogen NE2 of His163. However, we see that the C5 hydroxyl (equivalent to C3 in shikimate moiety) interaction with Thr382 is lost, since this hydroxyl group is absent in protocatechuate[1].

References

1. Eudes, A., Pereira, J.H., Yogiswara, S. Wang, G., Benites, V.T., Baidoo, E.E.K., Lee, T.S., Adams, P.D., Keasling, J.D. and Loqué, D.Exploiting The Substrate Promiscuity of Hydroxycinnamoyl-CoA:shikimate Hydroxycinnamoyl Transferase to Reduce Lignin. *Plant & Cell Physiology*. IN PRESS

Funding

This work was part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. We thank Dr. Nigel W. Moriarty for helpful analysis of the ligands geometry and restraints. We are also grateful to the staff of the Berkeley Center for Structural Biology at the Advanced Light Source of Lawrence Berkeley National Laboratory. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Ultrastructure of plant cells by electron microscopy: towards increased biofuel production

Maria Mueller[1*] (mmueller@lbl.gov), Purbasha Sarkar[1,2], Melissa Roth[2], Salil Apte[2], Edgar Yap[2], Michael Kowalczyk[3], Pablo Guindos[4], **Krishna K. Niyogi[2], Paul Adams[1], Manfred Auer[1,2]**

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley; [2]UC Berkeley MBIB Division, Berkeley; [3]Edinburgh Napier University, Edinburgh, United Kingdom; [4]Fraunhofer Institute for Wood Research, Stuttgart, Germany

http://www.jbei.org/people/directors/manfred-auer/

Project goals: The Physical Analysis group of the Joint BioEnergy Institute focuses on the development and the application of methods for the physical characterization of plants and microbes at the nano- and meso-scale. We support the analysis of JBEI experiments in the Feedstocks, Deconstruction, Fuels Synthesis and Technology divisions by using faithfully preserved samples and a variety of electron microscopy techniques, including traditional SEM and TEM, and wide-field montages TEM, as well as advanced 3D imaging approaches, such as electron tomography, and focused ion beam SEM and serial block face SEM.

The plant biomass suitable for the production of bioenergy and high value chemicals is concentrated in the cell wall. While significant progress has been made in understanding the composition and to some degree the regulation of the cell wall synthesis, little is known about its 3D organization and thus the mechanical properties that result as a consequence from this organization. Any effort aiming to re-engineer cell walls for the purpose of increasing yield or facilitating deconstruction needs to take plant cell wall 3D organization into consideration. Using quantitative 3D analysis of ultrarapid frozen, cryo-sectioned *Arabidopsis thaliana* stems and computer-assisted design (CAD) model building, we have obtained the first accurate 3D model of the primary and secondary cell wall of mature xylem tracheary cells *in situ*. The 3D model shows an unexpected pattern of microfibril orientation with sparse yet stiff hemicellulose cross-linking of microfibrils. Comparing the mechanical simulation results of our experimental model with other proposed models for turgor pressure (compression) and shear forces (bending) reveals how elegantly cell walls are designed by evolution to withstand the mechanical forces imposed on plants, leading to rigid yet flexible mechanical strength.

Plant chloroplasts are the site of photosynthesis and carbon-fixation and, therefore, a foundational understanding of their function and maturation is key for effective biomass generation. To carry out this complex task, in both plants and algae, chloroplasts contain an elaborate architecture of complex lamellar membrane systems, also known as photosynthetic thylakoid membranes. We are focusing on understanding the interplay between the molecular processes of light harvesting, carbon-fixation and ultrastructural thylakoid network organization including changes under changing light conditions. The unicellular green algae *Chromochloris zofingiensis* is known for its

production of biodiesel and a high-value carotenoid and, therefore is an ideal model system for studying chloroplast development. It allows us to trap defined stages of chloroplast development by switching between autotrophy and heterotrophy. Currently, we are working on analyzing the ultrastructural data of chloroplasts at different time points during the photosynthetic breakdown and biogenesis using serial sectioning and electron microscopy. Our recently developed cryo-immobilization sample preparation techniques show superior near-native ultrastructural preservation without imposing any size limitations.

Funding statement:
This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DEAC02-05CH11231.

Rapid Determination of Enzyme Activities for Lignocellulose Deconstruction and Analysis of Biofuel Molecules Using Nanostructure-Initiator Mass Spectrometry

Kai Deng[1,2]*(kdeng@lbl.gov), Joshua Heinemann[1,3], Joel Guenther[1,2], Taichi Taksuka[4], Lai Bergeman[4], Aram Kang[1,3], Taek Soon Lee[1,3], Jijiao Zeng[1,2], Kenneth Sale[1,2], Ee-Been Goh[1,3], Harry Beller[1,3], Ben Bowen[3], Blake Simmons[1,3], Paul Adams[1,3,5], Anup Singh[1,2], Brian Fox[4] , **Trent Northen**[1,3]

[1]Joint BioEnergy Institute, Emeryville, CA; [2]Sandia National Laboratories, Livermore, CA; [3]Lawrence Berkeley National Laboratory, Berkeley, CA; [4]GLBRC, University of Wisconsin, Madison, WI; [5]University of California, Berkeley, CA

http://www.jbei.org

Project Goal:
The JBEI mission is to find a viable way to convert lignocellulosic biomass into next generation transportation biofuels.

Abstract:

Our goal is to develop a platform to rapidly characterize the activity of enzymes responsible for the deconstruction of cellulose, hemicellulose and lignin and to screen alcohol and ketone biofuel production. Lignocellulosic biomass is composed of carbohydrate polymers (cellulose, hemicellulose) and an aromatic polymer (lignin). The complexity of the biomass structure requires cost effective enzyme cocktails for its deconstruction. In addition, a robust method to screen biofuel-producing strains for desired products is needed to support development and optimization of strains with high titre productivity. In order to meet these crucial challenges, we are developing mass spectrometry based assays with high-throughput, small sample volume, good sensitivity and importantly, adaptability to automated workstations to facilitate study large enzyme or microbial library strain libraries.

Central to our approach is to use synthetic organic chemistry to prepare chemical probes that enhance nanostructure-initiator mass spectrometry (NIMS) based analysis. This includes model substrates suitable for screening the activities of cellulases, hemicellulases and lignases and a post-reaction products-tagging strategy. Together these enable quantitation of glycan product cascades from biomass deconstruction (to obtain enzyme kinetic parameters so as to help the development of enzyme cocktails), and also methyl ketone and alcohol products (1-butanol, 3-methyl-3-butenol et. al) from biofuel production strains.

For high-throughput analysis of biomass deconstruction, we have standardized a panel of 12 substrates to span the biomass linkages of interest for plant-based biofuel production. To test the value of this standard panel for our high-throughput platform, we characterized the activities of three engineered cellulases CelAcc-CBM3a, CelRcc-

CBM3a, CelEcc-CBM3a and their synergy of combination across a range of reaction conditions and enzyme concentrations. We anticipate that large-scale screening using the standardized platform and substrates will generate critical datasets to enable direct comparison of enzyme activities for cocktail design. Work is underway integrating microfluidics with NIMS and we anticipate this new high-throughput platform will greatly enhance our ability to study biomass deconstruction and biofuel production.

Acknowledgements

The DOE Joint BioEnergy Institute and DOE Great Lakes Bioenergy Research Center are supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 and through contract DE-FC02-07ER64494, respectively. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's Nuclear Security Administration under contract DE-AC04-94AL85000.

References

1. Deng, K. et al. *Front. Bioeng. Biotechnol.* 2015 Oct 13;3:153. doi:10.3389/fbioe.2015.00153
2. Deng, K. et al. *Front. Bioeng. Biotechnol.* 2015 Oct 27;3:165. doi: 10.3389/fbioe.2015.00165
3. Deng, K. et al. *ACS Chem. Biol.* **2014**, *7*, 1470-1479.

Genomic Selection and Genome-wide Association Analyses for Bioenergy Traits in Switchgrass

Guillaume P. Ramstein[1] (ramstein@wisc.edu), Joseph Evans[2,3], Shawn M. Kaeppler[1,4], Jeremy Schmutz[5], Robert B. Mitchell[6], Kenneth P. Vogel[6], C. Robin Buell[2,3], **Michael D. Casler**[1,7]

[1] Department of Agronomy, University of Wisconsin-Madison, Madison, WI 53706, USA

[2] Department of Energy Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI 48824, USA

[3] Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA

[4] Department of Energy Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI 53706, USA

[5] HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA

[6] Grain, Forage, and Bioenergy Research Unit, Agricultural Research Service, United States Department of Agriculture, University of Nebraska, Lincoln, NE 68583-0937, USA

[7] Agricultural Research Service, United States Department of Agriculture, Madison, WI 53706, USA

Website: http://www.ars.usda.gov/pandp/people/people.htm?personid=32258

Project goals: 1. Employ exome capture to assess SNP marker diversity across the entire range of switchgrass. 2. Initiate genomic selection and phenotypic selection (as a control) on two switchgrass populations. 3. Identify genes involved in flowering time regulation in switchgrass. 4. Conduct candidate gene and QTL analyses of genes related to recalcitrance in switchgrass. 5. Develop bioinformatic tools to support switchgrass research within and outside of GLBRC.

Abstract: Switchgrass, a relatively high-yielding and environmentally sustainable biomass crop, has been chosen by the USDA and the USDOE as one of the main sources of bioenergy in the US. However, further genetic gains in biomass yield and quality must be achieved to make it an economically viable bioenergy feedstock. Genomics-assisted selection methods are particularly promising for generating rapid genetic gains in switchgrass and meeting the goals of a substantial displacement of petroleum use with biofuels in the near future. Here, we report on two types of analyses supporting the use of genomics-assisted selection for switchgrass breeding: genomic selection (GS), i.e., the use of genome-wide marker information to directly predict performance in breeding programs, and genome-wide association studies (GWAS), i.e., the search for regions in the genome showing significant associations with the traits of interest. We assessed GS prediction procedures for biomass yield, plant height and heading date in breeding populations and achieved prediction accuracies which, we believe, should motivate the implementation of GS in switchgrass breeding programs. We are currently performing GWAS for morphological and quality traits in a diversity panel and, according to preliminary results, we should be able to identify several candidate genomic regions involved in the elaboration of important bioenergy traits. The results in GS and GWAS that we are presenting here will pave the way for upcoming breeding experiments which will compare genomics-assisted selection to traditional types of selection and generate new cultivars for sustainable bioenergy production.

Funding:
US Department of Energy Great Lakes Bioenergy Research Center, Office of Science BER DE-FC02-07ER64494, Project Number 1.4.4
US Department of Energy Joint Genome Institute, Office of Science Contract No. DE-AC02-05CH11231

Interannual Climate Variability Affects the Microbial Response to and Fermentability of Lignocellulosic Biofuel Crops

David M. Cavalier[1]* (cavalie8@msu.edu), Donna Bates,[2] Scott Bottoms,[2] Charles Donald, Jr.,[1] Dustin Eilert,[2] Alan Higbee,[2] Li Hinchman,[2] Margaret Krueger,[1] Mick McGee,[2] Lawrence G. Oates,[2] Jeff Piotrowski,[2] Edward Pohlmann,[2] Gregg Sanford,[2] Jose Serate,[2] Scott A. Smith,[3] Dan Xie,[2] **Rebecca Garlock Ong**,[1]and **Yaoping Zhang**[2]

[1]DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI;
[2]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI;
[3]RTSF Mass Spectrometry & Metabolomics Core, Michigan State University, East Lansing, MI

http://www.glbrc.org

Project Goals: To design and implement a "Feedstock-to-Fuel" analytical pipeline that can be used to investigate how variables such as agronomic conditions, feedstock variability, pretreatment chemistry, and microorganism growth and productivity interact with each other during biofuel production.

Fundamental studies on lignocellulosic biofuel production are often limited in their scope, either focusing on individual stages of the process and/or a single feedstock, or spanning only a few stages and variables. In these evaluations, the impacts of variables within an earlier stage (for example, agronomic practices) on the results obtained in later stages (such as fermentation yields and efficiency) are often inferred based on assumptions that are widely accepted, but have little experimental evidence to support their validity. Indeed, results from these studies are typically shortsighted and have limited use in guiding the improvement of upstream operations. There is a need for fundamental studies that also have a systems-wide scope, however, these studies are very challenging to implement. By taking advantage of the breadth of expertise and resources within the Great Lakes Bioenergy Research Center (GLBRC) we were able to design and implement a "Feedstock-to-Fuel" analytical pipeline that allows us to perform the types of analyses needed to investigate how upstream variables such as agronomic conditions, feedstock variability, and pretreatment chemistry interact with each other and influence the properties of the fermentation media (enzymatic hydrolysate) and ultimately the genetic response, growth and productivity of the fermentation microorganisms.

In the first of these studies we investigated the impact of interannual climate variability on downstream biofuel production processes. Corn stover and switchgrass were collected at the UW Ag Research Station in Arlington, WI during three years with significantly different precipitation profiles (2010, 2012, and 2013). These were chosen to represent a wet, dry, and an average year, respectively, when compared to the 30-year climate norms for Arlington, WI. During fermentation of AFEX-treated biomass hydrolysates, *Zymomonas mobilis* 2032 showed no major difference in response, with the exception of slightly lower cell growth in the switchgrass

hydrolysates. *Saccharomyces cerevisiae* Y128 likewise showed reduced growth in the switchgrass hydrolysates, but was completely unable to grow in hydrolysate generated from 2012 switchgrass (SG). Statistical analysis of the chemical composition of the hydrolysates (common organic acids, phenolic-derived inhibitors, and minerals) did not implicate any compounds for the poor performance. Chemical genomic analysis of the yeast response to these hydrolysates pointed to inhibition of ergosterol biosynthesis as a reason for the growth inhibition. Further investigation revealed the presence of high quantities of pyrazines and imidazoles in both the acetone extract and hydrolysate of the pretreated 2012 switchgrass. Both classes of compounds are derived from reactions of soluble sugars with ammonia and are known inhibitors of ergosterol biosynthesis. Of these compounds, 2-methylpyrazine was particularly inhibitory to yeast growth. Addition of these compounds to a synthetic hydrolysate at the same concentrations as in the actual hydrolysate also inhibited growth of *S. cerevisiae* Y128. During the drought year, soluble sugars accumulated in the switchgrass, possibly because of incomplete utilization of the sugars due to early senescence or as an osmotic stress response, and were ultimately degraded during pretreatment to chemicals that were severely inhibitory to yeast growth. These results show a complex interplay between the weather impacts on biomass crop composition, pretreatment chemistry, and the response of the microorganisms during fermentation.

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

.

Functional Metagenomic Comparison of Soil Microbiomes Associated with Six Cellulosic Biofuel Feedstock Production Systems

David Duncan[1,2] * (dsduncan@wisc.edu), **James Tiedje**[3,4], and **Randall Jackson**[1,2]

[1]University of Wisconsin, Madison; [2]Great Lakes Bioenergy Research Center, Madison, Wisconsin; [3]Michigan State University, East Lansing; [4]Great Lakes Bioenergy Research Center, East Lansing, Michigan

https://www.glbrc.org/research/sustainability
https://jacksonlab.agronomy.wisc.edu/
http://cme.msu.edu/tiedjelab/jtiedje.shtml

Project Goals: The Sustainability Research Area of the Great Lakes Bioenergy Research Center (GLBRC) addresses the sustainability of producing liquid transportation fuels from cellulosic materials, with an emphasis on evaluating and enhancing novel cropping systems that can provide cellulosic feedstocks. A key component of this work has been the study of soil microbes to better understand how plant-microbe interactions shape microbial communities and enhance plant productivity, and how microbial communities influences biogeochemical cycling. Here, we focused on how a diverse array of potential cellulosic feedstock production systems, including traditional annual crops, perennial crops, and assemblages of native species, reshape the functional metagenomic profile of their associated soil microbial communities in the years following cropping system.

Among the most exciting aspects of a cellulosic bioenergy economy is the potential to commercialize agroecosystems that provide valuable environmental services but that have historically been overlooked in favor of more conventional agricultural production systems (*1*). These systems include perennial natives such as switchgrass (*Panicum virgatum*), highly productive exotics such as miscanthus (*Miscanthus x giganteus*), and even diverse assemblages of native prairie species. Much of our understanding of many of these systems derives from conservation projects, which may not accurately reflect the management conditions for cellolosic feedstock production. It is thus imperative to understand how agronomic management emphasizing feedstock production impacts the agroecological properties of these highly promising systems.

There is growing awareness of the critical role soil microbes play in determining the ecological functionality of agroecosystems (*2*). Previous work from the GLBRC has demonstrated the tremendous influence cropping systems and their management exert on soil microbial community compositions (*3*), as well as the potential for those differences in composition to impact key agroecosystem processes such as soil respiration and methanotrophy (*4*). Our present study expands on these earlier findings, leveraging both the experimental infrastructure of the GLBRC's Bioenergy Cropping Systems Experiment (BCSE) and the recent, rapid development in use of high-throughput sequencing to characterize the functional metagenomics of complex microbial communities.

We studied six experimental cellulosic feedstock production systems: continuous, no-till corn (*Zea mays*), monocultures of switchgrass, miscanthus, and hybrid poplar (*Populus nigra x P. maximowiczii*), a mixture of five native grass species, and a mixture of 18 native prairie species, established in 2008 and sampled annually from 2010 through 2012. We extracted environmental genomic DNA from the unseparated bulk and rhizosphere soils from these systems, which was subjected to shotgun metagenomic sequencing by the DOE Joint Genome Institute. Our dataset consists of 164 annotated, individually-assembled metagenomes, 1 to 2 Gbp in size. We used clusters of orthologous group (COG) annotation to define the functional genetic composition of these communities, relativizing by COG model length and abundance of a suite of single-copy housekeeping genes to derive relative abundances of individual COGs.

We observed considerable interannual variability within cropping systems, particularly in the continuous corn. There were significant differences among cropping systems in their overall functional gene content, but contrary to our expectations, the corn system did not function as an outlier relative to the perennial systems. Similarly, there were very few differences among systems in the relative abundances of key denitrification genes, despite considerable differences in the emissions of nitrous oxide observed during this time (*5*). Our findings suggest the overall functional genetic profile of soil microbial communities may display a less consistent response to differences among cropping systems than is observed in microbial biomass. Alternatively, it is possible that there are system-driven changes in the identity of the organisms possessing a particular function which may not translate to differences in the relative frequency with which that function occurs in the overall population.

References

1. G. P. Robertson *et al.*, Sustainable biofuels redux. *Science.* **322**, 49–50 (2008).
2. M. G. A. van der Heijden, R. D. Bardgett, N. M. van Straalen, The unseen majority: Soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol. Lett.* **11**, 296–310 (2008).
3. C. Liang *et al.*, Soil microbial communities under model biofuel cropping systems in southern Wisconsin, USA: Impact of crop species and soil properties. *Appl. Soil Ecol.* **54**, 24–31 (2012).
4. U. Y. Levine, T. K. Teal, G. P. Robertson, T. M. Schmidt, Agriculture's impact on microbial diversity and associated fluxes of carbon dioxide and methane. *ISME J.* **5**, 1683–1691 (2011).
5. L. G. Oates *et al.*, Nitrous oxide emissions during establishment of eight alternative cellulosic bioenergy cropping systems in the North Central United States. *GCB Bioenergy* (2015), doi:10.1111/gcbb.12268.

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

Identification of Grass Stem Specific Promoters for Improvement of Biofuel Crops by Metabolic Engineering

Jacob K Jensen,[1,2]* (jensen58@msu.edu), Mingzhu Fan,[1,2] Sang Jin Kim,[1,3] Starla Zemelis,[1,3] Federica Brandizzi,[1,3] and **Curtis G Wilkerson**[1,2,4]

[1]Great Lakes Bioenergy Research Center (GLBRC), Michigan State University, East Lansing; [2]Department of Plant Biology, Michigan State University, East Lansing; [3]Plant Research Laboratory, Michigan State University, East Lansing; and [4]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing

https://www.glbrc.org/

Project Goals: The major goal of our project is to manipulate the levels and composition of hemicellulose in the cell walls of biofuel crops. We therefore seek to understand the regulation of hemicellulose biosynthesis in sufficient detail so we can manipulate these processes in the plant to optimize their use as biofuel feedstocks. Specifically we are studying the grass specific hemicellulose mixed-linkage glucan (MLG), which can accumulate to very high level in tissues such as the endosperm. We wish to replicate the very large accumulation of MLG seen in endosperms cell walls in stem pith parenchyma cells in order to increase the sink capacity of these cells and to provide a more digestible wall with a high C6 to C5 ratio.

Plants store a variety of compounds as reserves of energy and carbon skeletons for future needs. Many of these compounds have other uses in cells such as the structural role that hemicelluloses play in cell walls or the use of lipids to construct membranes. In some cases these compounds are allocated in dedicated storage tissues, some of which are derived form the stem. One such example is potato tubers, their developmental origin being evident from typical stem cell type arrangement. Another example is the stems of grasses, such as sugar cane and sorghum, which store large amounts sucrose and sometimes starch. In the later case, carbon accumulation often occurs during periods of slow growth to use when more favorable conditions resume. These naturally occurring mechanisms suggest that it should be possible to engineer bioenergy crops to store polysaccharides in the cell wall in vegetative tissues, particularly stem pith parenchyma, to be easily broken down into sugars for conversion to biofuels or other high-value compounds.

Our primary goal is to gain sufficient knowledge of the control of hemicellulose deposition to manipulate the accumulation of large amounts of mixed-linkage glucan (MLG) in the stem parenchyma tissue of grasses. We chose this tissue because grasses use this location to store a large variety of compounds, such as starch and sucrose as mentioned above, without detrimental affects on plant growth, and because it represents a very large storage compartment, approximately a third of the aerial part of the plant for the larger grasses such as maize and sorghum. We chose MLG because it is a polysaccharide exclusively composed of glucose, is easily extracted from the wall and enzymatically digested, and because it accumulates to large

extent in a number of grass tissues. The most pronounced example of such a tissue is the Brachypodium seed endosperm where MLG constitutes 40% of the total dry weight, giving rise to very thick cell walls almost exclusively consisting of MLG polysaccharide.

To accomplish this goal, we will require promoters that are specific for the stem parenchyma tissue and that are active at this location at the correct developmental stage. Additionally, we will need to understand how tissues such as the endosperm accumulate large quantities of MLG in the absence of cellulose accumulation.

We will present our progress on the isolation of stem pith parenchyma specific promoters. Using transcriptional profiling of developing internodes and stem pith enriched tissue samples from Brachypodium we have identified a series of candidate genes exhibiting pith parenchyma specific expression. We have cloned the upstream promoter regions of these candidate genes and are in the process of evaluating these constructs as promoter-GUS fusions by stable transformation of Brachypodium. We will present the analysis of the first four of these candidate promoters and detailed analysis of one that exhibiting high expression in the pith parenchyma cell type throughout the stem. We also present our work on using transient transformation of stem pith cells using a gene gun, which will accelerate our ability to define the promoter region required for tissue specificity.

In parallel with evaluating our promoter candidates by GUS expression characterization, we are producing a collection of transgenic plant materials of Brachypodium and maize ectopically expressing the MLG synthase gene *BdCSLF06* using our stem pith promoter candidates. These plants will server to further evaluate the promoters and provide crucial insights into the requirements for accumulation of MLG in the cell wall.

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.

Ectopic expression of *WRI1* affects fatty acid homeostasis in *Brachypodium distachyon* vegetative tissues

Yang Yang[1,2,*](young83@msu.edu), Jacob Munz[3,4,†], Cynthia Cass[3,4], Agnieszka Zienkiewicz[1,2], Que Kong[1,2], Wei Ma[2,4], Sanjaya[2, §], John Ohlrogge[2,5], John Sedbrook[3,4], **Christoph Benning[1,2,5]**

[1]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA.
[2]Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI 48824, USA.
[3] Great Lakes Bioenergy Research Center, Illinois State University, Normal, IL 61790, USA.
[4]School of Biological Sciences, Illinois State University, Normal, IL 61790, USA.
[5]Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA.
[6]MSU-Department of Energy, Plant Research Laboratory, East Lansing, MI 48824, USA.
[§]Present address: Energy and Environmental Science Institute, Department of Biology, West Virginia State University, Institute, WV 25112, USA.
[†]Present address: Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

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

Project Goals: This project aims to synthesize high-energy oils in vegetative tissues of biomass crops, including C4 grasses. Enhancing the energy content and liquid fuel yield of crop biomass can be achieved by accumulating triacylglycerols (TAG) in the crop. After extracting these compounds, lignocellulosic feedstock remains and can be used for processing and fermentation. Thus engineering oil production will result in novel biofuel crops that achieve multiple objectives. Our previous research has succeeded in producing TAG in Arabidopsis leaves and stems (Sanjaya et al., 2011; Sanjaya et al., 2013). Oil content was increased by combining transcription factors, down-regulation of starch synthesis, and addition of enzymes of TAG synthesis (Sanjaya et al., 2011). Testing these second generation approaches in the grass model plant Brachypodium is now efficient and cost-effective providing a 'proof-of-concept' strategy prior to moving into corn or other energy crops.

WRINKLED1 (WRI1) is a transcription factor, which governs fatty acid (FA) synthesis and indirectly TAG accumulation in oil storing plant tissues and ectopic expression of its cDNA has led to TAG accumulation in vegetative tissues of different dicotyledonous plants (Cernac and Benning, 2004; Cernac et al., 2006; Ma et al., 2013; Ma et al., 2015). The ectopic expression of *BdWRI1* in the grass Brachypodium (*Brachypodium distachyon*) induced the transcription of predicted genes involved in glycolysis and fatty acid (FA) biosynthesis, and TAG content was increased up to 32.5-fold in 8-week-old leaf blades. However, the ectopic expression of *BdWRI1* also caused cell death in leaves, which has not been previously observed in dicotyledonous plants such as Arabidopsis (*Arabidopsis thaliana*)(Yang et al., 2015). Lipid analysis indicated that the free FA content was 2.0-fold elevated in *BdWRI1*-expressing leaf blades of Brachypodium. The transcription of predicted genes involved in β-oxidation was induced. In addition, linoleic fatty acid treatment caused cell death in Brachypodium leaf blades, an effect

that was reversed by addition of the FA biosynthesis inhibitor cerulenin. Taken together, ectopic expression of *BdWRI1* in Brachypodium enhances FA biosynthesis and TAG accumulation in leaves as expected but also leads to increased free FA content, which has cytotoxic effects leading to cell death. Thus, while WRI1 appears to ubiquitously affect FA biosynthesis and TAG accumulation in diverse plants, its ectopic production can lead to undesired side effects depending on the context of the specific lipid metabolism of the respective plant species.

This work was funded in part by the US Department of Energy Great Lakes Bioenergy Research Center Cooperative Agreement DE-FC02-07ER64494 and MSU AgBioResearch.

References

Cernac A, Andre C, Hoffmann-Benning S, Benning C (2006) WRI1 is required for seed germination and seedling establishment. Plant Physiol 141: 745-757

Cernac A, Benning C (2004) WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. Plant J 40: 575-585

Ma W, Kong Q, Arondel V, Kilaru A, Bates PD, Thrower NA, Benning C, Ohlrogge JB (2013) Wrinkled1, a ubiquitous regulator in oil accumulating tissues from Arabidopsis embryos to oil palm mesocarp. PLoS One 8: e68887

Ma W, Kong Q, Grix M, Mantyla JJ, Yang Y, Benning C, Ohlrogge JB (2015) Deletion of a C-terminal intrinsically disordered region of WRINKLED1 affects its stability and enhances oil accumulation in Arabidopsis. Plant J 83: 864-874

Sanjaya, Durrett TP, Weise SE, Benning C (2011) Increasing the energy density of vegetative tissues by diverting carbon from starch to oil biosynthesis in transgenic Arabidopsis. Plant Biotechnol J 9: 874-883

Sanjaya, Miller R, Durrett TP, Kosma DK, Lydic TA, Muthan B, Koo AJ, Bukhman YV, Reid GE, Howe GA, Ohlrogge J, Benning C (2013) Altered lipid composition and enhanced nutritional value of Arabidopsis leaves following introduction of an algal diacylglycerol acyltransferase 2. Plant Cell 25: 677-693

Yang Y, Munz J, Cass C, Zienkiewicz A, Kong Q, Ma W, Sanjaya, Sedbrook J, Benning C (2015) Ectopic Expression of WRINKLED1 Affects Fatty Acid Homeostasis in Brachypodium distachyon Vegetative Tissues. Plant Physiol 169: 1836-1847

Rhizosphere metagenomics and metatranscriptomics of biofuel crops

Jiarong Guo[1]*(guojiaro@gmail.com), Aaron Garoutte[1], Adina Howe[2], Qiong Wang[1], Titus Brown[3], James Cole[1], **James Tiedje**[1]

[1]Michigan State University, East Lansing; [2]Iowa State University, Ames; [3]University of California, Davis, Davis

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

Project Goals: Our research supports a goal of the Great Lakes Bioenergy Research Center (GLBRC) which is to develop sustainable biofuel practices by optimizing soil, microbe, and plant interactions, since soil microbes form beneficial association with the crops' rhizospheres and also play a major role in ecosystem functions, such as N and C cycles. Our main goal is to find key eco-functional genes selected by biofuel crop rhizospheres, with a focus on nitrogen cycle genes, which is critical for sustainable biofuel production on marginal lands.

We applied deep sequencing to study rhizosphere microbiomes of biofuel crops. Through the past few years of GLBRC, our group has developed several tools (SSUsearch, khmer, and Xander) to deal with big metagenomic data that can not be analyzed by traditional bioinformatics methods. These tools enable us to explore the genetic and population components of rhizosphere communities, with special emphasis on ecological services. Here we showcase two deep sequencing enabled rhizosphere microbiome studies: 1) rhizosphere metagenomics of three biofuel crops; 2) core functional microbiome of switchgrass.

Rhizosphere metagenomics of three biofuel crops: Crop roots had strong influences on the soil microbial community. Thus large-scale plantation of biofuel crops will have significant impact on ecosystem functions regionally and beyond. We compared rhizosphere microbial communities of corn (annual) and switchgrass and *Miscanthus* (perennials). This is the first comparative study of these biofuel crops using shotgun metagenomics and one of the largest sequencing efforts to date (about 1 TB bp in total). We compared the rhizosphere metagenomes at three levels: overall community structure (SSU rRNA gene), overall function (annotation from global assembly), and N cycle genes (from Xander). All three levels showed corn had a significantly different community from *Miscanthus* and switchgrass (except for AOA). In terms of life history strategy, the corn rhizosphere was enriched with more copiotrophs while the perennials were enriched with oligotrophs, which is further supported by higher abundance of genes in "Carbohydrates" and higher fungi/bacteria ratios. In addition, corn also had a less rich and even community, so the perennials managed to maintain a more diverse community even though investing less C in the rhizosphere. Moreover, a larger dispersion of corn data in ordination plots and enriched *Penicillium* (non-beneficial fungi) also indicate corn may not be doing as well in controlling its community and selecting beneficial member. Furthermore, the nitrogen fixing community of corn was dominated by *Rhizobium* (perhaps a legacy from prior legume crops)

while the perennials had NifH sequences most related to *Coraliomargarita*, *Novosphingobium* and *Azospirillum*, indicating that the perennials can better select beneficial members. Moreover, higher numbers of genes for nitrogen fixation and lower number of genes for nitrite reduction suggest better nitrogen sustainability of the perennials. Thus our study provides comprehensive evidence showing perennial bioenergy crops have advantages over corn in higher microbial species and functional diversity and in selecting members with beneficial traits, consistent with a higher level of sustainability of perennial biofuel crops.

Core functional microbiome of switchgrass: We used a “multi-omic” approach to establish a minimum core of actively transcribed functions in rhizosphere soils of switchgrass. We used metagenomics and metatranscriptomics to identify shared transcribed functions in samples. The minimum functional core accounts for 99% or 92% of functional annotations depending on annotation database (SEED Subsystems or RefSeq respectively). We then used metaproteomic data to further confirm our minimum functional core. The metaproteome data had 460 unique SEED Subsystem annotations with an abundance of 876,429. All but 12 of the SEED Subsystem annotations were found in the minimum functional core further validating the minimum functional core. The Subsystems with the greatest relative abundance in the metatranscriptome minimum functional core are the Clustering-based subsystem (defined as a cluster of genes known to work together but for which no function has been established), Protein metabolism, Carbohydrates (partly housekeeping functions and partly functions related to sugar utilization), Miscellaneous and RNA metabolism. Obviously the majority of these functions are either “known unknowns” or functions related to housekeeping. These results identify transcripts expressed at high levels and hence targets for further investigation of their function.

Our research is supported by GLBRC, a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Alkaline Pretreatment Effectively Reduces Recalcitrance of Zip-Lignin Poplar

Aditya Bhalla,[1,2] Namita Bansal,[1,2] Leonardo da Costa Sousa,[1,2] Venkatesh Balan,[1,2] Christa Gunawan,[1,2] **Bruce E. Dale**,[1,2] and **Eric L. Hegg**[1,2]***** (erichegg@msu.edu)

[1]Michigan State University, East Lansing; [2]DOE Great Lakes Bioenergy Research Center, East Lansing, Michigan

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

Project Goals: To compete with traditional petroleum refineries, cellulosic bio-refineries must achieve high carbohydrate-to-fuel yields and facilitate lignin valorization to commodity products. The Great Lakes Bioenergy Research Center (GLBRC) adopted a two-fold strategy for improving enzymatic deconstruction of hardwoods via alkaline pretreatment. First, the GLBRC has refined the alkaline pretreatment processes employed to reduce biomass recalcitrance. These efforts have resulted in the development of two highly effective alkaline pretreatments: Extractive Ammonia (EA) and copper bipyridine-catalyzed alkaline hydrogen peroxide (Cu-AHP) pretreatment. Second, the GLBRC has genetically engineered poplar to contain readily cleavable ester bonds in the backbone of lignin (Zip-Lignin™). This modification of the cell wall was designed specifically to facilitate deconstruction of hardwoods via alkaline pretreatment and subsequent enzymatic hydrolysis.

EA pretreatment takes advantage of the properties of liquid ammonia for modifying the crystalline structure of cellulose from native cellulose I_β (CI) to cellulose III_I (CIII) as well as selectively extracting part of the lignin, leaving most of the carbohydrates intact in a single process stream. Previous studies demonstrated that the CIII allomorph is responsible for increasing enzymatic hydrolysis rates by 2-5 fold[1, 2] relative to the native CI. Furthermore, lignin is well known to inhibit both enzymes and microorganisms required to convert lignocellulosic biomass into biofuels. The combined effect of lignin extraction and CIII formation is responsible for a significant improvement in biomass conversion at low enzyme loading and high solid loading enzymatic hydrolysis compared to ammonia fiber expansion (AFEX™) pretreatment. In this work, EA pretreatment was performed using low and high severities on two Zip poplar lines. Enzymatic hydrolysis of the EA pretreated Zip poplar lines at 1% glucan loading, using 15 mg protein per g glucan enzyme loading, revealed glucan and xylan conversions up to ~80%. Relative to the wild-type line, the incorporation of Zip-Lignin into poplar improved glucan conversions by approximately 8% in this preliminary study.

Alkaline hydrogen peroxide (AHP) has been shown to be an effective pretreatment for both herbaceous feedstocks and woody biomass. The utility of this approach, however, has been limited by the prohibitively high oxidant loadings (e.g., 250-2000 mg H_2O_2 per g biomass). We recently discovered that adding small amounts of copper 2,2'-bipyridine complexes [Cu(bpy)] as

catalysts during AHP pretreatment (Cu-AHP) resulted in substantially improved delignification and enhanced sugar yields following enzymatic hydrolysis at modest oxidant loadings i.e., ~100 mg H_2O_2 per g biomass.[3] We subsequently demonstrated that copper co-localized with disrupted regions of the cell wall following pretreatment of poplar, providing indirect evidence for copper catalyzing the oxidation and removal of lignin.[4] With an aim to increase the effectiveness of the Cu-AHP process on woody biomass and further reduce chemical inputs, we utilized Zip-Lignin poplar as a substrate. Our preliminary results revealed 12% higher glucose yields (78% to 90%) relative to the wild-type line. Alternatively, the enzyme loading on Cu-AHP pretreated Zip-Lignin poplar could be decreased by 2/3 while still maintaining glucose yields at 78%. Together, these results highlight the significant potential for Zip-Lignins to reduce the recalcitrance of woody biomass.

References

1. S. P. S. Chundawat, G. Bellesia, N. Uppugundla, L. da Costa Sousa, D. Gao, A. M. Cheh, U. P. Agarwal, C. M. Bianchetti, G. N. Phillips, P. Langan, V. Balan, S. Gnanakaran and B. E. Dale, *J. Am. Chem. Soc.*, **2011**, *133*, 11163–11174.
2. D. Gao, S. P. S. Chundawat, A. Sethi, V. Balan, S. Gnanakaran and B. E. Dale, *Proc. Natl. Acad. Sci.*, **2013**, *110*, 10922–10927.
3. Z. Li, C. H. Chen, E. L. Hegg, D. B. Hodge, *Biotechnol. Biofuels*, **2013**, *6*, e119.
4. Z. Li, N. Bansal, A. Azarpira, C. H. Chen, J. Ralph, E. L. Hegg, D. B. Hodge, *Biotechnol. Biofuels*, **2015**, *8*, e123.

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

Plants for Improved Biomass Deconstruction: Native Zip-lignins across the Plant Kingdom.

Steven D. Karlen[1,2*] (skarlen@wisc.edu), Matthew L. Peck[3], Chengcheng Zhang[3], Rebecca A. Smith[1,2], Dharshana Padmakshan[1], Kate E. Helmich[1,2], Heather C. A. Free[4], Seonghee Lee[5], Bronwen G. Smith4, Fachuang Lu[1,2], John C. Sedbrook[1,6], Richard Sibout[7], John H. Grabber[8], Troy M. Runge[1,2], Kirankumar S. Mysore[5], Philip J. Harris[4], Laura E. Bartley[3], and **John Ralph**[1,2]

[1] Great Lakes Bioenergy Research Center (GLBRC); [2]University of Wisconsin, Madison, Wisconsin; [3]University of Oklahoma, Norman, Oklahoma; [4]University of Auckland, Auckland, New Zealand; [5]Samuel Roberts Noble Foundation, Ardmore, Oklahoma; [6]Illinois State University, Normal, Illinois; [7]INRA and AgroParisTech, Institut Jean-Pierre Bourgin (IJPB), Versailles, France; [8]Dairy Forage Research Center, USDA-ARS, Madison, Wisconsin

https://www.GLBRC.org

Project Goals: To determine the extent of natural monolignol ferulate conjugate utilization in cell wall lignification, in which readily cleavable ester linkages ('zips') are introduced into the lignin polymer backbone in ways that significantly improve biomass-processing energetics.

Angiosperms represent the majority of terrestrial plants and are the primary research focus for conversion of biomass to liquid fuels and co-products. Accessing the fibers and chemical energy stored in plant cell walls often requires high temperatures and/or harsh chemical treatments to cleave the lignin inter-unit bonds. However, if base-labile ester bonds, are introduced into the lignin polymer, as can be accomplished by augmenting the prototypical monomers with monolignol ferulate (ML-FA) conjugates, then lignin fragmentation occurs under mild alkaline conditions. The findings presented here provide evidence that 'zip-lignins', lignins derived in part from ML-FAs, have developed naturally via convergent evolution in diverse angiosperm lineages. The discovery of a putatively native *FERULOYL-CoA MONOLIGNOL TRANSFERASE* (FMT) enzyme in rice, encoded by *OsAT5 (OsFMT)*, and in *A. sinensis*, encoded by *AsFMT*, provides new avenues into designing plants for deconstruction, by engineering new or elevated levels of ML-FAs, as has been demonstrated in poplar and rice. This also means that breeders can select for high-zip-lignin plants that are significantly easier to process, and perhaps already inadvertently have done so. Such approaches have the potential to significantly reduce the costs and improve the energy balance of converting biomass to liquid fuels, cellulose pulps, and other value-added products.

Publication:

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

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

Global Catalog of the Transcriptional Response to Lignocellulosic Biomass-derived Inhibitors in *Escherichia coli* Identifies Promoters for Synthetic Engineering of Biofuel Microbes

Donna M. Bates[1,2]* (donna.bates@wisc.edu), Sean McIlwain,[1] Irene M. Ong[1], Jeff Grass,[1] Yaoping Zhang[1,2], Patricia J. Kiley,[1,2,3] and **Robert Landick**[1,2,4,5]

[1]University of Wisconsin-Madison; [2]Great Lakes Bioenergy Research Center, Madison, WI; [3]Department of Biomolecular Chemistry; [4]Department of Biochemistry, and [5]Department of Bacteriology

https://www.glbrc.org/

Project Goals: Knowledge of the mechanisms by which microbes respond to lignocellulosic biomass-derived inhibitors will inform approaches to engineer biocatalysts that can efficiently convert biomass to biofuels. In this study, we identified *Escherichia coli* promoters responsive to inhibitors present in ammonia-pretreated corn stover hydrolysate in order to exploit their regulatory properties for native-signal expression programming in microbes. In addition to increased knowledge of the global effects of lignocellulosic inhibitors on microbial transcription, the data obtained expands the catalog of available synthetic biology parts available for rational engineering of microbes.

A major challenge to efficient biological conversion of lignocellulosic hydrolysates to biofuels is the presence of toxic inhibitors (referred to here as lignotoxins) derived from biomass pretreatment. Knowledge of the mechanism by which microbes respond to lignotoxins (LTs) is crucial in order to engineer tolerant strains and increase biofuel yields. This study expands upon previous analysis of the global transcriptional response to LTs in *E. coli*[1] by identifying LT-specific promoters. Chromatin immunoprecipitation-sequencing (ChIPseq) assays, using antibodies directed against RNA polymerase subunits σ^{70}, σ^{S}, or σ^{N}, were performed on cultures grown in the absence or presence of LTs, from the exponential, transition, and stationary growth phases. Differentially bound ChIP sites were identified and mapped to the nearest novel or published transcription start site (TSS). Direct effects on transcription were found by comparing to the RNAseq data[1] of the associated condition. Moreover, examination of promoter occupancy by housekeeping and alternative σ factors across the genome provide insight into the effect of LTs on RNA polymerase composition and global effects on transcription. Identification of LT-specific promoters can enable engineering of microbial gene-induction systems that naturally respond to hydrolysate components and expand the microbial promoter catalog and associated regulatory elements for use in synthetic biology.

References

1. Keating, D.H., et al., *Aromatic inhibitors derived from ammonia-pretreated lignocellulose hinder bacterial ethanologenesis by activating regulatory circuits controlling inhibitor efflux and detoxification.* Frontiers in Microbiology, 2014. **5**:402, DOI: 10.3389/fmicb.2014.00402.

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

Genome-Enabled Assembly of Carbohydrate-Active Enzymes for the Deconstruction of Biomass

Evan Glasgow[1,2]* (eglasgow@wisc.edu), Emily Beebe,[1,2] Kirk Vander Meulen,[1,2] Lai Bergeman,[1,2] Johnnie Walker,[1,2] Taichi Takasuka,[1,2] Christopher Bianchetti,[1,2] Savannah Mills,[1,2] Bob Smith,[1,2] Craig Bingman,[1,2] Adam Book,[1,2] Robb Stankey,[1,2] Gina Lewin,[1,2] Evelyn Wendt-Pienkowski,[1,2] Cameron Currie,[1,2] Steven Karlen,[1,2] Hoon Kim,[1,2] John Ralph,[1,2] Leonardo Sousa,[1,3] Bruce Dale,[1,3] Sam Deutsch,[4] Kai Deng,[5] Trent Northen,[5] Seema Singh,[5] Ken Sale,[5] Blake Simmons,[5] Paul Adams,[5] Michael G. Hahn,[6] Sivakumar Pattathil,[6] and **Brian Fox[1,2]**

[1]Great Lakes Bioenergy Research Center (GLBRC), Madison, WI [2]University of Wisconsin-Madison, Madison, WI; [3]Michigan State University, Lansing, MI; [4]Joint Genome Institute, Walnut Creek, CA, [5]Joint Bioenergy Institute, Emeryville, CA, [6]Bioenergy Science Center, Oak Ridge, TN

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

Project Goals: GLBRC deconstruction research focuses on basic physical, chemical, and enzymatic strategies to overcome biomass recalcitrance. Carbohydrate-active enzymes are especially promising as powerful, specific, and green catalysts for digestion of plant polysaccharides into valuable sugars for downstream conversion into fuels. With the tools for high-throughput gene and protein synthesis in hand, GLBRC aims to 1) functionally annotate large families of putative carbohydrate-active enzymes, 2) produce high-resolution structural models of the best-performing enzymes, and 3) engineer recombinant variations with improved activity, as well as formulate synergistic combinations of enzymes to accelerate biomass hydrolysis in industrial settings. This work is part of a larger effort shared between GLBRC-supported laboratories and all DOE Bioenergy Research Centers, with the grand challenge of unlocking the renewable carbon reservoir in lignocellulosic biomass.

Overcoming the recalcitrance of lignocellulosic biomass is the primary deconstruction aim of GLBRC. The impetus for this aim is the need for increased yields from plant materials used in the production of bioenergy and renewable carbon fuel. GLBRC approaches this problem from multiple angles simultaneously: genetic engineering of plants with more accessible cell walls, development of chemical pretreatment methods to enable enzymatic digestion of lignocellulosic materials, discovery and development of enzymes with enhanced activity against plant polysaccharides, and design of microbial fermentation processes to convert biomass into valuable chemical commodities.

Within the enzymology sector of deconstruction research, the overall strategy follows a progression of phylogeny to function to structure. By sourcing genomic sequences from databases (such as CAZy) or from metagenomic analyses of diverse microbial communities, high-throughput gene synthesis of interesting carbohydrate-active targets is achieved via

collaboration with the DOE Joint Genome Institute. Cell-free translation reactions allow preparation of microgram amounts of enzyme for initial screening against polysaccharides of different monomeric compositions and chemical linkages. Enzymes with interesting reaction profiles can easily be scaled up to milligram quantity in a cellular expression system, permitting detailed biochemical characterization and crystal structure determination. This approach has been successfully applied to glycoside hydrolase family 55 (GH55) enzymes, of which the laminarinase SactELam55A (whose structure was recently solved) is a member.

A current focus is on subfamily 4 of glycoside hydrolase family 5 (GH5_4), a clade of about two hundred enzymes with broad specificity. Total functional coverage of this group is desired, starting with assignment of cellulase, mannanase, and xylanase activities to each member. In particular, multispecific enzymes with temperature and pH optima that coincide with industrial process conditions are sought, and a number of new thermophilic and acidophilic enzymes have been identified. Fusions of enzymes with carbohydrate-binding modules (CBMs) has been investigated with the well-studied cellulase CelE, and this strategy may also prove useful on newly discovered catalytic domains. Ultimately, novel combinatorial mixtures of enzymes with synergistic activities are tested on pure substrates and on pre-treated biomass samples including genetically-engineered poplar strains with altered lignocellulose structure, which were developed at GLBRC. Glycome profiling and nanostructure-initiator mass spectrometry (NIMS) are used in collaboration with DOE Bioenergy Science Center (BESC) and Joint Bioenergy Institute (JBEI) to identify specific reactions of glycoside hydrolases with polysaccharide epitopes in plant cell walls and to quantify the product distributions and kinetics of these reactions.

Thorough, rational sampling of uncharacterized enzyme space and detailed mechanistic experiments ensure a high probability of adding versatile catalysts to the existing armory of biomass deconstruction tools.

DOE Great Lakes Bioenergy Research Center, Joint BioEnergy Institute, Joint Genome Institute, and BioEnergy Science Center are supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through contracts DE-FC02-07ER64494 (GLBRC), DE-AC02-05CH11231(JBEI/JGI), and DE-AC05-00OR22725 (BESC), respectively.

Natural Genetic Variations Influence Ionic Liquid Tolerance by *Saccharomyces cerevisiae*

Trey K. Sato[1*] (tksato@glbrc.wisc.edu), Douglas Higgins[2,3], Mary Tremaine[1], Scott Bottoms[1], Quinn Dickinson[1], Jenna Fletcher[1], Margaret Agnew[1], Maria Sardi[1,4], Audrey P. Gasch[1,4], Steve Singer[2,3], Blake A. Simmons[2,5], Robert Landick[1,6] and Michael P. Thelen[2,3],

[1]DOE Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, WI; [2]Deconstruction Division, Joint BioEnergy Institute, Emeryville, CA; [3]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA; [4]Department of Genetics, University of Wisconsin, Madison, WI; [5]Biomass Science and Conversion Technology Department, Sandia National Laboratories, Livermore, CA; [6]Department of Biochemistry, University of Wisconsin, Madison, WI.

Project Goals

Despite the increased interest and efforts in cellulosic biofuels research, a number of molecular and biochemical barriers remain that prevent the efficient bioconversion of plant feedstocks into ethanol and other biofuels. One barrier for the industrial yeast biocatalyst, *Saccharomyces cerevisiae*, is the inhibition of fuel production incurred from chemical compounds in pretreated lignocellulosic feedstocks. These compounds often cause cellular stress, which in turn impacts biofuel yield and productivity. These stressors are not only degradation products generated during the pretreatment process, including acetic acid and lignin-derived phenolics, but are also compounds used in feedstock pretreatment, such as the ionic liquids (ILs) 1-ethyl-3-methylimidazolium -chloride and -acetate ($[C_2C_1im]Cl$] and $[C_2C_1im][OAc]$. At the DOE Great Lakes Bioenergy Research Center and Joint BioEnergy Institute, we have collaborated to employ genetic, phenotypic and screening approaches to determine the genetic variation that drives a range of IL tolerance in natural isolates of *S. cerevisiae*.

Abstract

Imidazolium ILs effectively solubilize lignin and cellulosic components of biomass, enabling the subsequent enzymatic hydrolysis of cellulose into fermentable sugars. Although the majority of $[C_2C_1im]$ ILs can be recovered after the pretreatment process, residual IL can remain in hydrolysates at concentrations that significantly impair yeast growth, viability, and fermentation. We phenotyped a large panel of wild and domesticated *S. cerevisiae* strains for growth tolerance in lab media containing $[C_2C_1im]Cl$ and $[C_2C_1im]OAc$. While the canonical lab strain, BY4741, grew poorly in lab media containing 125 mM $[C_2C_1im]Cl$, we identified two yeast strains with significantly greater tolerance relative to other strains. Genomic DNA fragments from one of these tolerant strains, 378604X, was then cloned into fosmids, which were subsequently transformed into the intolerant BY4741 strain and selected for improved growth in the presence of 125 mM $[C_2C_1im]Cl$. From this selection, we identified two genes that conferred greater IL tolerance when overexpressed in BY4741. Furthermore, one of these genes, *SGE1*, was required for $[C_2C_1im]Cl$ tolerance in both BY4741 and 378604X strains. Interestingly, *SGE1* is has non-synonymous polymorphisms between the BY4741 and 378604X strains, as well as between other wild and domesticated strains with a range of IL tolerance. Expressing the *SGE1* allele from 378604X into the BY4741 *sge1Δ* strain conferred IL tolerance, while expressing the $SGE1^{BY4741}$ allele in the 378604X *sge1Δ* strain imparted IL sensitivity. The Sge1

protein is predicted as a multi-spanning transmembrane protein previously characterized as a drug and cationic dye pump, suggesting that the 378604X strain retains a natural genetic variant with greater tolerance to $[C_2C_1im]Cl$. These results exemplify the importance of strain background when selecting a biocatalyst for metabolic engineering, as well as highlight the possibilities that subtle differences protein sequences can confer phenotypic improvements relevant to biofuel production. **This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494) and Joint BioEnergy Institute (DOE BER Office of Science DE-AC02-05CH11231).**

The Yeast Biodesign Library: leveraging DNA synthesis to assess and harness genes from diverse organisms

William G. Alexander*[,1,2] (walexander@wisc.edu), Nikolay Rovinskiy[1,2], Mariana Lopes[2,3], David Peris[1,2], Jin Kang[1,2], Angela Tarver[4], Miranda Harmon-Smith[4], Jan-Fang Cheng[4], **Samuel Deutsch[4]**, **Jeff S. Piotrowski[1]**, **Trey K. Sato[1]**, **Audrey P. Gasch[1,2]**, **Chris Todd Hittinger[1,2]**

[1]Department of Energy Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI; [2]Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI; [3]Universidade Federal de Minas Gerais, Belo Horizonte, Brasil; [4]Joint Genome Institute, Walnut Creek, CA

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

Project Goals: identify and characterize genes that provide novel phenotypes relevant to efficient cellulosic bioethanol production by *Saccharomyces cerevisiae* using synthesized DNA constructs.

Two major issues encountered by biofuel researchers in the optimization of *Saccharomyces cerevisiae* cellulosic ethanol strains are the tolerance to toxins in hydrolysate derived from the lignin in plant biomass and the conversion of pentoses, disaccharides, and other unfavored sugars. The latter is especially problematic as the vast majority of *S. cerevisiae* strains are completely unable to use the pentose xylose, the major component of hemicellulose, as a carbon source without genetic modification. However, many yeasts that belong to a group known as the “CUG clade” are capable of using xylose and cellobiose readily; unfortunately these yeasts also possess an alternate genetic code, complicating the evaluation of their genes in heterologous contexts. To address these difficulties, we designed 266 synthetic gene cassettes called the Yeast Biodesign Library. Each cassette is driven by a medium-strength constitutive *S. cerevisiae* promoter and contains an open reading frame that has been cleansed of problematic restriction enzyme sites and incompatible codons. To date, we have successfully used the Yeast Biodesign Library to generate artificial multi-gene pathways and to implicate novel genes in xylose fermentation.

This work was supported by the National Science Foundation under Grant No. DEB-1253634 and funded in part by the DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494). CTH is a Pew Scholar in the Biomedical Sciences, supported by the Pew Charitable Trusts. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE- AC02-05CH11231.

Leveraging natural variation in *Saccharomyces cerevisiae* to elucidate the toxicity mechanisms of lignocellulosic hydrolysate and advanced biofuels.

Maria Sardi* (sardi@wisc.edu), Michael Place, **Audrey Gasch**

DOE Great Lakes Bioenergy Center, Laboratory of Genetics, University of Wisconsin-Madison

PROJECT GOALS: Explore natural genetic and phenotypic variation in *Saccharomyces cerevisiae* to elucidate the stress response to lignocellulosic hydrolysate and advanced biofuels such as butanol and isobutanol.

ABSTRACT:

The increased demand for alternative fuels is driving the development of more efficient and economical production of biofuels. This requires the use of non-food based plant biomass to produce advanced biofuels such as butanol and isobutanol. A major challenge of implementing this new energy source is that the chemically treated plant material, known as lignocellulosic hydrolysate, contains a variety of toxic compounds that affect fermenting microbes by inhibiting growth, metabolism, and alcohol production, all of which decrease the economic efficiency of lignocellulosic biofuel production. In addition, butanol and isobutanol are toxic even at small concentrations, making end product toxicity a significant limiting factor. We are using multiple genomic strategies to identify mechanisms of toxicity and tolerance that can be then use to engineer tolerance into industrially relevant microbes. Using a collection of 165 genetically and phenotypically diverse strains of *Saccharomyces cerevisiae*, we are exploring natural genetic and phenotypic variation to understand lignocellulosic hydrolysate tolerance. First, by comparing and contrasting the transcriptional responses of tolerant and sensitive strains exposed to these stresses, we are identifying the primary toxin targets and their effects on cellular physiology. Second, we are exploring how genetic background affects engineering strategies. We are measuring phenotypic variation in the response to gene over-expression by introducing a library of 5,000 barcoded, high-copy plasmids that each expresses a different yeast gene, into four different yeast strains. By exploring background-specific effects on the fitness contribution of each gene to toxin and end-product tolerance, we hope to uncover important strategies for engineering. Finally, we are using Genome-wide Association (GWA) and bulk segregant analysis to identify connections between genotype and phenotype. By applying multiple genomic strategies and integrating the results, we expect to identify strategies for improving tolerance to the stresses found in the production of advanced biofuels.

Collaborative Proteomics- Description of the 1 Hour Yeast Proteome and its use in GLBRC

Alexander S. Hebert[1,2]* (ahebert@wisc.edu), Alicia L. Richards[3], Michael S. Westphall[2], Trey K. Sato[1], Jeff S. Piotrowski[1], Robert Landick[1], **Joshua J. Coon[1,2,3,4]**

[1]DOE Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, WI; [2] Genome Center of Wisconsin, University of Wisconsin, Madison, WI; [3] Department of Chemistry, University of Wisconsin, Madison, WI; [4]Department of Biomolecular Chemistry, University of Wisconsin, Madison, WI

Project Goals

This project was initiated to exploit contemporary proteomic methodologies to achieve two goals of the GLBRC. The first is to investigate the mechanism of toxicity of specific fermentation inhibitors (e.g. γ-valerolactone (GVL), ionic liquids, isobutanol). The second is to improve our understanding how engineered and evolved strains of *Saccharomyces cerevisiae* metabolize xylose into biofuels.

Abstract

The ability to measure differences in protein expression has become key to understanding biological phenomena. Owing to cost, speed, and accessibility, transcriptomic analysis is often used as a proteomic proxy. However, mRNA is a genetic intermediary and cannot inform on the myriad of post-translational regulation processes. For the past decade considerable effort has been invested in maturing proteomic technology to deliver information at a rate and cost commensurate to transcriptomic technologies. Recently, a new Orbitrap hybrid mass spectrometer comprised of a mass filter, a collision cell, a high-field Orbitrap analyzer, and a dual cell linear ion trap analyzer was introduced (Q-OT-qIT, Orbitrap Fusion, Thermo). This system offers a high MS^2 acquisition speed of 20 Hz, and with this fast scan rate, is capable of providing whole proteome analysis in record time. To maximize instrument performance we developed an optimized cellular lysis approach, employed trypsin digestion, and used dimethyl sulfoxide (DMSO, 5%) as an LC additive to increase abundance of acidic peptides and unify charge state. Using this novel system we report the comprehensive analysis of the yeast proteome (4,002 @ 1% FDR) following 1.3 hours of nLC-MS^2 analysis (70 minute gradient). Over the chromatographic method, the Q-OT-qIT hybrid collected an average of 13,447 MS^1 and 80,460 MS^2 scans per run to produce 43,400 peptide spectral matches and 34,255 peptides with unique amino acid sequences. These experiments delivered an extraordinary 67 proteins per minute and demonstrate that complete analysis of the yeast proteome can be routinely performed in approximately one hour.

We have applied our single shot yeast proteomic analysis to validate chemical genomic analyses studies that revealed a number of mitochondrial gene mutants had increased sensitivity to the ionic liquid, and that EMIM-Cl and BMIM-Cl alter mitochondrial membrane polarization similar to the ionophore valinomycin. High-throughput quantitative proteomics validated this prediction, as there was significant enrichment for proteins involved in mitochondrial processes among proteins with depleted abundance in the presence of EMIM-Cl. From this study we also find 36 genes that have both significantly altered protein expression and affect viability when knocked out, in

the presence of EMIM-Cl. These genes represent proteins that are both important to the *in vivo* response to EMIM-Cl treatment and have a measurable effect on susceptibility to EMIM-Cl. Additionally, we found the expression of *hsp30*, a negative regulator of Pma1p, is decreased by ~10 fold after EMIM-Cl treatment. From chemical genomics experiments we know deletion of *ptk2*, an activator of Pma1p, imparts resistance to EMIM-Cl toxicity, yet its expression is not affected by treatment. Thus over expression of *hsp30* may impart EMIM-Cl resilience similar to *ptk2* removal and may also be a pivotal gene in explaining EMIM-Cl toxicity or a new point of rational engineering.

In addition to the above, we have also analyzed proteomic profiles of engineered and evolved *S. cerevisiae* in biological triplicate under aerobic and anaerobic conditions to elucidate how genetic changes combinatorially promote *S. cerevisiae* growth and metabolism of xylose under aerobic and anaerobic conditions. Interestingly, we found that enzymes mediating the metabolism of non-fermentable carbon sources were significantly downregulated in xylose metabolizing strains. This suggested that the engineered and evolved mutations cause *S. cerevisiae* to recognize xylose as a fermentable carbon source. This analysis demonstrates the power of combining complimentary metabolomics and proteomics approaches to tease out molecular details.

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

Metabolic effects and toxicity mechanisms of lignocellulose-derived inhibitors

Tippapha Pisithkul,[1,2] David M. Stevenson,[1,2] and **Daniel Amador-Noguez**[1,2]*
(amadornoguez@wisc.edu)

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

Project Goals: The mission of the Great Lakes Bioenergy Research Center is to perform the basic research that generates technology to convert cellulosic biomass to ethanol and other advanced biofuels. In alignment with this goal, we have made use of time-resolved, quantitative metabolic flux analyses (MFA) to gain insight into the mechanisms by which lignotoxins disrupt metabolism and inhibit sugar (i.e. xylose) conversion to biofuels. Understanding the mechanisms behind these deleterious effects is of great value for devising metabolic engineering strategies to overcome them.

An outstanding challenge toward efficient production of biofuels and value-added chemicals from plant biomass is the impact that lignocellulose-derived inhibitors have on microbial fermentations. Using *E. coli* and *Z. mobilis* as model systems, we investigated the metabolic effects and toxicity mechanisms of feruloyl amide and coumaroyl amide, the predominant phenolic compounds in ammonia-pretreated biomass hydrolysates.

Using metabolomics, isotope tracers, and biochemical assays, we discovered that these two phenolic amides act as potent and fast-acting inhibitors of purine and pyrimidine biosynthetic pathways in *E. coli* (1). Feruloyl or coumaroyl amide exposure leads to (i) a rapid buildup of 5-phosphoribosyl-1-pyrophosphate (PRPP), a key precursor in nucleotide biosynthesis, (ii) a rapid decrease in the levels of pyrimidine biosynthetic intermediates, and (iii) a long-term generalized decrease in nucleotide and deoxynucleotide levels. Tracer experiments using 13C-labeled sugars and 15N-ammonia demonstrated that carbon and nitrogen fluxes into nucleotides and deoxynucleotides are inhibited by these phenolic amides. We found that these effects are mediated via direct inhibition of glutamine amidotransferases that participate in nucleotide biosynthetic pathways. In particular, feruloyl amide is a competitive inhibitor of glutamine PRPP amidotransferase (PurF), which catalyzes the first committed step in de novo purine biosynthesis.

Similarly to *E. coli*, exposure to feruloyl amide in *Z. mobilis* results in a large accumulation of the biosynthetic intermediate PRPP, suggesting that this lignotoxin also affects nucleotide biosynthetic pathways in this biofuel producer. However, while PRPP accumulation was accompanied by decreased nucleotide levels in *E. coli*, nucleotide and deoxynucleotide levels were actually elevated in *Z. mobilis* after feruloyl amide-treatment. To examine whether this increase in nucleotide levels was due to active biosynthesis or to DNA/RNA degradation, we performed isotopic tracer experiments using 13C-labeled glucose. We found that feruloyl amide did not block *Z. mobilis* nucleotide biosynthesis: 13C-carbons were incorporated into ATP, GTP, UTP, and CTP in the feruroyl amide-treated cells similarly to control cells. It is currently unclear how *Z.*

mobilis recovers from PRPP accumulation and the decrease in levels of dihydroorotate and orotate, but these unexpected observations point to the possibility that *Z. mobilis* may have additional nucleotide biosynthetic routes or mechanisms to overcome inhibition of glutamine amidotransferases by feruloyl amide. We are currently investigating these questions.

References

1. **Pisithkul T**, **Jacobson TB**, **O'Brien TJ**, **Stevenson DM**, **Amador-Noguez D**. 2015. Phenolic Amides Are Potent Inhibitors of *De Novo* Nucleotide Biosynthesis. Appl. Environ. Microbiol. **81**:5761–5772.

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

Modified Expression of HD-ZIP III and NAC Transcription Factors Impact Biomass Formation and Sugar Release Efficiency in *Populus*

Udaya C. Kalluri[1]* (kalluriudayc@ornl.gov), Raja Payyavula,[1] Sara S Jawdy,[1] Wellington Muchero,[1] Robert W. Sykes,[2,1] Mark Davis,[2,1] Tim Tschaplinski,[1] Dan Jacobson,[1] Gerald Tuskan[1] and **Paul Gilna[1]**

[1]BioEnergy Science Center and Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]BioEnergy Science Center and National Renewable Energy Laboratory, Golden, Colorado

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) designing plant cell walls for rapid deconstruction and (2) developing multi-talented microbes or converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC 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 testing large numbers of natural variants and generating specific and modified plant samples as well as developing genomics tools for detailed studies into poorly understood cell wall biosynthesis pathways.

We are undertaking a multipronged, genetics- and genomics-based, research strategy to examine regulation of cell wall and biomass properties in *Populus*. Here, we present results from studies of genes belonging to the Homeodomain-leucine zipper class III (HD-ZIP III) and NAC domain transcription factor families. HD-ZIP III and NAC domain transcription factor family members have been previously known to regulate a diverse array of plant pathways ranging from embryo, shoot, leaf and flower development to biotic and abiotic stress response and hormone signaling. We examined the roles of selected gene family members, with enhanced expression in secondary wall-enriched xylem tissue, in determining biomass properties including sugar release efficiency of *Populus*. Transgenic *Populus deltoides* plants were generated via overexpression or down-regulation under the control of either a vascular-specific 4-coumarate-CoA ligase (*4CL*) promoter or that of a ubiquitous promoter. Morpho-anatomical, metabolic, physiological, cell wall chemistry studies revealed the extent of impact on biomass properties. Such modifications in gene expression resulted in a range of phenotypes; from changes in cell wall chemistry with no apparent compromise on plant growth and performance to pleiotropic effects to favorable phenotypic combination of enhanced growth and reduced recalcitrance. We further performed SNP (single nucleotide polymorphism) network analyses to extract genotypic and phenotypic correlations significant to the targeted genes. Using *PdWND1B* (NAC) and *PdHB3* (HD-ZIP III) as examples, we discuss alternate modes of action of these transcription factor genes in impacting biomass formation and sugar release efficiency in *Populus*.

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

Pleiotropic and Epistatic Network-Based Discovery: Integrated SNP Correlation, Co-Expression and Genome-Wide Association Networks for *Populus trichocarpa*

Deborah Weighill[1,2]* (weighillda@ornl.gov), Carissa Bleker,[1,2] Gerald Tuskan,[1] Wellington Muchero,[1] Tim Tschaplinski,[1] Daniel Jacobson,[1,2] and **Paul Gilna**[1]

[1]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]University of Tennessee, Knoxville

http://bioenergycenter.org/

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) designing plant cell walls for rapid deconstruction and (2) developing multi-talented microbes or converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC researchers provide enabling technologies in biomass characterization, 'omics, modeling and data management in order to (1) understand chemical and structural changes within biomass and (2) to provide insights into biomass formation and conversion mechanisms.

Biological organisms are complex systems that are composed of pleiotropic functional networks of interacting molecules and macro-molecules. Complex phenotypes are the result of orchestrated, hierarchal, heterogeneous collections of expressed genomic variants. However, the effects of these variants are the result of historic selective pressure and current environmental and epigenetic signals, and, as such, their co-occurrence can be seen as genome-wide correlations in a number of different manners. Biomass recalcitrance (i.e., the resistance of plants to degradation or deconstruction, which ultimately enables access to a plant's sugars) is a complex polygenic phenotype of high importance to the U.S. Department of Energy's biofuels programs. We are using data derived from the re-sequenced genomes from over 1000 alternate *Populus trichocarpa* genotypes in combination with transcriptomics, metabolomics and phenomics data across this population in order to better understand the molecular interactions involved in recalcitrance. The resulting Genome Wide Association Study networks, integrated with SNP (single nucleotide polymorphism) correlation and co-expression networks, are proving to be a powerful approach to determine the pleiotropic and epistatic relationships underlying cellular functions and, as such, the molecular basis for complex phenotypes, such as 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.

GWAS Studies in *Populus* Reveal Evolution of Two Biosynthetic Enzymes into Transcriptional Regulators Modulating Phosphoenolpyruvate Input and Chorismate Output from the Shikimate Pathway

Jay Chen[1]* (chenj@ornl.gov), Anthony Bryan,[1] Meng Xie,[1] Kelsey Yee,[1] Hao-Bo Guo,[2] Olivia Thompson,[1] Miguel Rodriquez, Jr.,[1] Sara Jawdy,[1] Lee Gunter,[1] Priya Ranjan,[1] Kai Feng,[1] Stephen DiFazio,[3] Luke Evans,[3] Gancho Slavov,[3] Robert Sykes,[4] Erica Gjersing,[4] Wendy Schackwitz,[5] Joel Martin,[5] Erika Lindquist,[5] Vasanth Singan,[5] Chris Daum,[5] Gerald A. Tuskan,[1] Wellington Muchero,[1] and **Paul Gilna**[1]

[1]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]University of Tennessee, Knoxville, Tennessee; [3]West Virginia University, Morgantown, West Virginia [4]National Renewable Energy Laboratory, Golden, Colorado; [5]DOE Joint Genome Institute, Walnut Creek, California

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) designing plant cell walls for rapid deconstruction and (2) developing multi-talented microbes or converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC 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 testing large numbers of natural variants and generating specific and modified plant samples as well as developing genomics tools for detailed studies into poorly understood cell wall biosynthesis pathways.

Redirecting carbon flow to competing pathways has been proposed as one of the most efficient ways to engineer viable, low lignin cellulosic biofuels feedstocks. To achieve this goal, regulators of carbon flux at key junctions need to be identified and characterized. Using Genome-Wide Association Studies (GWAS), we identified hitherto unknown regulators of carbon flow into and out of the shikimate pathway.

At the entry point to the shikimate pathway, phosphoenolpyruvate (PEP) is the precursor molecule for 3-Deoxy-D-arabinoheptulosonate 7-phosphate (DAHP) biosynthesis. The same molecule is also the precursor for pyruvate biosynthesis in the glycolysis pathway. Similarly, chorismate, the output molecule from the shikimate pathway, is a shared precursor for both downstream phenylpropanoid and tryptophan biosynthesis. As such, pyruvate and tryptophan biosynthesis present valuable sinks to drive carbon flux away from lignin formation. However, not much is known about genetic regulation of carbon flux at both junctions. Here, we used GWAS in *Populus* to identify two transcriptional regulators of carbon flow at these entry and exit junctions. Interestingly, both of these transcriptional regulators were evolved from biosynthetic enzymes in their respective ancestral pathways.

Firstly, we observed that variation in polyglutamine (polyQ) repeat length within a *Populus* locus annotated as 2-hydroxyacid dehydrogenase led to differential activation of marker genes associated with lignin biosynthesis. Analysis of knock-out mutants in *Arabidopsis* revealed that the transcription of genes encoding pyruvate kinase and malate dehydrogenase were significantly

downregulated in the knock-out lines compared to wild type. Since these two genes lie at the entry point of PEP into pyruvate biosynthesis, we evaluated the responses of genes in the competing shikimate pathway for evidence of increased PEP shunt in that direction. There was indeed significant up-regulation of genes associated with shikimate and cell wall biosynthesis in the mutant line. In fact, a total of nine transcription factors with known regulatory activity of secondary cell wall and phenylpropanoid biosynthesis exhibited at least 2-fold upregulation in the mutant line. These results suggested that, in its functional state, this transcriptional regulator represses transcription of genes in the shikimate pathway in favor of pyruvate biosynthesis.

Secondly, we identified a novel isoform of the 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase which catalyzes the sixth step in the shikimate pathway. We observed that this isoform harbored an additional N-terminal helix-turn-helix DNA binding motif and a nuclear localization signal. Naturally-occurring loss-of-function mutations at this locus resulted in increased tryptophan biosynthesis and up to 50% reduction in lignin suggesting that in its functional state, this regulator preferentially drives chorismate towards phenylpropanoid biosynthesis. We confirmed this by evaluating *Populus* overexpression transgenic lines. Consistent with the observations in the natural variants, cell wall biosynthesis master regulators MYB46 and NST1 were significantly upregulated in the transgenic lines overexpressing this EPSP synthase isoform, suggesting that this regulator preferentially allocates chorismate towards the phenylpropanoid pathway. Interestingly, overexpression resulted in increased flavonoid biosynthesis at the expense of lignin formation.

These newly identified transcriptional regulators hold tremendous potential in engineering reduced lignin cellulosic biofuel feedstocks.

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.

Modification of GAUT12 Expression Affects Recalcitrance in the Woody Feedstock *Populus deltoides*

Ajaya K. Biswal,[1,2,3] Melani A. Atmodjo[1,2,3]* (matmodjo@ccrc.uga.edu), Xiaohan Yang,[3,4] Kim Winkeler,[3,5] Cassandra Collins,[3,5] Sushree S. Mohanty,[2,3] David Ryno,[2,3] Robert W. Sykes,[3,6] Geoffrey B. Turner,[3,6] Angela Ziebell,[3,6] Erica Gjersing,[3,6] Mark F. Davis,[3,6] Stephen R. Decker,[3,6] Debra Mohnen,[1,2,3] and **Paul Gilna**[3]

[1]Department of Biochemistry and Molecular Biology, University of Georgia, Athens; [2]Complex Carbohydrate Research Center, University of Georgia, Athens; [3]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [4]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [5]ArborGen, Inc., Ridgeville, South Carolina; [6]National Renewable Energy Laboratory, Golden, Colorado

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) designing plant cell walls for rapid deconstruction and (2) developing multi-talented microbes or converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC 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 testing large numbers of natural variants and generating specific and modified plant samples as well as developing genomics tools for detailed studies into poorly understood cell wall biosynthesis pathways.

The deconstruction of feedstocks such as *Populus* for biofuel production is challenging due to the recalcitrance of woody biomass. *Populus* biomass is rich in cellulose, xylan and lignin with smaller amounts of pectin. The high degree of cross-linking and interactions between the wall polymers is a major barrier to the conversion of cellulose and other polysaccharides into ethanol without expensive chemical and/or enzymatic pretreatment. GAUT12 (GAlactUronosylTransferase12)/ IRX8 (IrRegular Xylem8) is a putative glycosyltransferase proposed to be involved in secondary cell wall glucuronoxylan and/or pectin biosynthesis.

We manipulated the expression of one of the two *Populus deltoides* GAUT12 homologs (*PtGAUT12.1* and *PtGAUT12.2)* via overexpression and RNAi approaches and determined the effects on biomass recalcitrance and growth. A 7–48% increase (*PdGAUT12.1*-OE) and 50–67% decrease (*PdGAUT12.1*-KD) in *Populus deltoides* GAUT12.1 transcript expression compared to controls yielded a 4–11% decrease and 4–8% increase, respectively, in glucose release following enzymatic saccharification of the biomass. The *GAUT12.1*-OE lines had reduced plant height and stem radial diameter of 6–54% and 8–41%, respectively, compared to controls, a phenotype opposite to the increased growth of *PdGAUT12.1*-KD lines with a 12–52% and 12–44% increase in plant height and radial stem diameter, respectively, compared to the controls (Biswal et al., *Biotechnology for Biofuels* **8**:41, 2015). The wood of *PdGAUT12.1*-OE lines had 12–17% increased galacturonic acid (GalA) and 14–20% increased xylose content. Conversely wood from *GAUT12.1*-KD lines had 25–47% reduced GalA and 17–30% reduced xylan content. There was no effect of *GAUT12.1* overexpression or downregulation on total lignin content. Taken together, the results suggest that GAUT12 is required for the synthesis of one or more pectin and xylan-

containing structures in *Populus* and that there is an inverse relationship between the amounts of this structure in the wall and stem diameter, height and sugar release. We hypothesize that GAUT12 synthesizes a wall polymer that cross-links the wall and that increased amounts of this structure lead to a more cross-linked wall, greater recalcitrance and reduced growth. Reciprocally, we hypothesize that reduced amounts of the polymer(s) lead to loosened walls and hence decreased recalcitrance and increased growth. The results show that the directed down regulation of GAUT12 in *Populus* can lead to better woody feedstock for the biofuel industry.

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

Sebacina for Switchgrass: Application and Benefits

Prasun Ray,[1,2] Kelly D Craven[1,2]* (kdcraven@noble.org), and **Paul Gilna**[2]

[1]Plant Biology Division, The Samuel Roberts Noble Foundation, Ardmore, Oklahoma;
[2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) designing plant cell walls for rapid deconstruction and (2) developing multi-talented microbes or converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC 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 testing large numbers of natural variants and generating specific and modified plant samples as well as developing genomics tools for detailed studies into poorly understood cell wall biosynthesis pathways.

The Sebacinales belong to a taxonomically, ecologically, and physiologically diverse group of fungi in the Basidiomycota. *Sebacina vermifera* is a mycorrhizal fungus that was first isolated from the Australian orchid *Cyrtostylis reniformis*. Research from our laboratory on this fungus clearly indicates its plant growth—promoting abilities in non-orchid host plants. *S. vermifera* colonization enhanced seed germination, biomass production and drought tolerance of the native warm-season grass, switchgrass, an important bioenergy crop for cellulosic ethanol production in the United States. Towards this end, we have developed a method for large-scale dissemination of inoculum containing this growth-promoting fungus for switchgrass field trials. Further, we demonstrate the effectiveness of this carrier-based method for colonization of switchgrass in greenhouse trials for biomass enhancement of wild type and transgenic, low lignin (COMT downregulated) switchgrass lines compared to an efficient *in vitro* colonization method. *S. vermifera* colonization enhanced plant biomass regardless of delivery method, although the percentage of fungal biomass *in planta* increased with the clay-based delivery system. Further, we found that release of some clay minerals in solution was enhanced in the presence of *S. vermifera*, while others were seemingly reduced. Intriguingly, the presence of *S. vermifera* has little or no impact on cell wall composition, including lignification. This clay-based inoculum is not prone to contamination and can be stored at room temperature. These features together with the mineral composition of the clay illustrate the potential for packaging the fungus with a nutrient source en masse for large-scale delivery to the field. This research is the first report documenting the development of a bentonite clay particle-based delivery system for mass production of any symbiotic microbe and suggests that *S. vermifera* can be packaged with a mineral composite and effectively delivered to a target host plant. We are currently taking this next step, initiating field trials of transgenic and natural varieties of switchgrass for *S. vermifera*-mediated crop improvement studies.

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.

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

Xin Chen,[2,3] Manesh Shah,[2] Debbie Weighill,[2,5] Amy Mason,[1,2] Luis L. Escamilla-Trevino,[1,4] Hongli Xie,[1,2] Qin Ma,[2,3] Zeng-Yu Wang,[1,2] Richard A. Dixon,[2,4] Richard S. Nelson,[1,2] Gerald Tuskan,[2] Wellington Muchero,[2] Tim Tschaplinski,[2] Intawat Nookaew,[2] Dan Jacobson,[2] Ying Xu,[2,3] Yuhong Tang[1,2]* (ytang@noble.org), and **Paul Gilna**[2]

[1]The Samuel Roberts Noble Foundation, Ardmore, Oklahoma; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]University of Georgia, Athens; [4]University of North Texas, Denton

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) designing plant cell walls for rapid deconstruction and (2) developing multi-talented microbes or converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC 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 testing large numbers of natural variants and generating specific and modified plant samples as well as developing genomics tools for detailed studies into poorly understood cell wall biosynthesis pathways.

Switchgrass (*Panicum virgatum*) is within one of the two plant genera (*Populus* the other) targeted in BESC for biomass improvement. For switchgrass, over 140 unique genes, promoters, promoter-genes, or stacked genes, represented by over 160 constructs, were targeted for modified expression through stable transformation to diminish recalcitrance. In addition, evaluation of natural variants for enhanced biofuel traits were carried out. After preliminary analysis to determine which lines had increased sugar release and normal or enhanced plant growth, twelve transgenic and natural variant lines were identified as switchgrass TOP Lines. These TOP Lines were compared in greenhouse and field to aid in selecting the best lines for enhanced biofuel production. Transcriptome analysis by RNA-seq is one of the characterization methods used to identify underlying pathways through which modification may further improve biofuel production.

TOP Line transcriptome and wall analyses were performed with greenhouse-grown plants at reproductive stage 1 (R1). RNA-seq was conducted at Joint Genome Institute (JGI) using Illumina TruSeq technology. For each sample, a total of 40–50 million paired-end (PE) reads of 150 bp was generated. From data quality control to differential gene expression analysis, each RNA-seq dataset was processed through a pipeline compiled from a set of publically available software. Unique mapping results and gene level assembly were generated with HISAT2 and related programs against switchgrass genome assembly v1.1. To facilitate data inspection at sample and gene levels, the RNA-seq data are also being displayed in BESC Jbrowser portal along with the switchgrass reference genome. For each TOP Line, genes whose expression was different from the control were selected through comparison between each TOP Line and its control, using differential analysis software such as DESeq, NOIseq and NOIseqbio as well as an array-like method. The selected gene list from each TOP Line was then analyzed by network analysis tools developed by BESC. The switchgrass genes were converted to *Populus* orthologs which were used to extract metabolites, genes and correlations from GWAS, co-evolution and co-expression networks developed for

Populus. These networks help to identify changes contributing to the cell wall modification in the RNAi TOP Lines, providing further gene targets for cell wall modifications. In addition, RNA-seq data will also be combined with data from other wall analyses to be analyzed at system biology level to understand cell wall structures and their impacts on 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.

Comparison of Multiple Generations of TOP Line Transgenic Switchgrass with Reduced Recalcitrance for Enhanced Bioconversion to Fuels

Alexandru Dumitrache,[1] Jace Natzke,[1] Miguel Rodriguez Jr.,[1] Kelsey L. Yee,[1] Olivia A. Thompson,[1] Charleson R. Poovaiah,[1,4] Hui Shen,[1,5] Mitra Mazarei,[1,4] Holly L. Baxter,[1,4] Nancy L. Engle,[1] Angela Ziebell,[1,2] Robert W. Sykes,[1,2] Erica Gjersing,[1,2] Chunxiang Fu,[1,3] Zeng-Yu Wang,[1,3] Mark F. Davis,[1,3] Timothy J. Tschaplinski,[1] C. Neal Stewart, Jr.,[1,4] Richard A. Dixon,[1,5] Jonathan R. Mielenz,[1] Steven D. Brown[1]* (brownsd@ornl.gov), Brian H. Davison,[1] and **Paul Gilna**[1]

[1]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]National Renewable Energy Laboratory, Golden, Colorado; [3]The Samuel Roberts Noble Foundation, Ardmore, Oklahoma; [4]University of Tennessee, Knoxville; [5]University of North Texas, Denton

http://bioenergycenter.org/

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) designing plant cell walls for rapid deconstruction and (2) developing multi-talented microbes or converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC 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 testing large numbers of natural variants and generating specific and modified plant samples as well as developing genomics tools for detailed studies into poorly understood cell wall biosynthesis pathways. BESC researchers provide enabling technologies in biomass characterization, 'omics, modeling and data management in order to (1) understand chemical and structural changes within biomass and (2) to provide insights into biomass formation and conversion mechanisms.

Switchgrass is a promising bioenergy feedstock, yet its recalcitrance is a key obstacle in large-scale bioconversion applications. Advancements in plant feedstocks, through transgenic and natural variant selection, has allowed for a variety of TOP Lines to emerge with improved phenotypic qualities. In this study, a yeast-based separate hydrolysis and fermentation (SHF) bioconversion assay was used to assess TOP Line switchgrass plants. Biomass included switchgrass lines with altered expression of genes involved in pectin and lignin biosynthesis and regulation as well as C1 metabolism. Multiple years of field-grown GAUT4, miRNA, MYB4, COMT, and FPGS altered lines were all analyzed for carbohydrate composition and ethanol yield and assessed using statistics and correlation analyses of various features of these TOP Lines. Overall, results revealed a generalized year-over-year increase in glucan and xylan plant content that is more pronounced for certain transgene targets, and an overall increase in ethanol yields for the modified crops, with highs of 36% more yield for overexpression of *PvMYB4* (MYB4 line) over parental controls. Statistical analysis revealed that selection of parental lines with improved properties and transformability had also one of the largest impacts on glucan content changes and ethanol yield increases. Future advances in natural parental line screening and a better understanding of reduced recalcitrance due to various genetic target modifications will aid efforts in developing industrially relevant switchgrass lines.

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.

Field Experiments of Seven Switchgrass TOP Lines

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

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

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) designing plant cell walls for rapid deconstruction and (2) developing multi-talented microbes or converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC 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 testing large numbers of natural variants and generating specific and modified plant samples as well as developing genomics tools for detailed studies into poorly understood cell wall biosynthesis pathways.

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

Greenhouse studies have identified several potentially successful transgenic routes for reducing cell wall recalcitrance and/or improving growth in switchgrass. Downregulation of COMT (caffeic acid *O*-methyltransferase), a lignin biosynthetic gene, reduces lignin and the S/G ratio, thereby improving sugar release and ethanol yield. Overexpression of MYB4, a transcription factor that represses the expression of multiple lignin biosynthetic genes simultaneously, also reduces recalcitrance and improves sugar release efficiency and ethanol yield. Overexpression of miRNA156, a regulator of plant developmental processes, improves biomass yield and sugar release efficiency. Downregulation of CWG-1 (cell wall gene) improves biomass yield, sugar release efficiency, and ethanol yield. Downregulation of CWG-2 improves sugar release efficiency. Overexpression of ethylene response factor/shine (ERF/SHN) transcription factor improves biomass yield and sugar release in switchgrass. Downregulation of CWG-3 improves sugar release efficiency. In addition, natural accessions of switchgrass that produce high biomass with improved sugar release efficiency have been identified through association analysis.

An important validation step, especially for genetically engineered plants, is to perform multi-year field studies, which is a vital goal of BESC. It is well known that the greenhouse is not always predictive of crop performance in the field. Herein we present data from seven BESC "TOP Lines" and appropriate controls from agronomically-relevant University of Tennessee (Knoxville) field studies, in which plants were grown under USDA APHIS BRS release into the environment permits for two or three field seasons. Data include: (1) agronomic performance (morphology and end-of-season biomass), (2) lignin content and composition by high-throughput py-MBMS, (3) sugar release by high-throughput enzymatic assays, (4) ethanol yield by SHF assays, and (5) incidence of switchgrass rust, caused by the pathogen (*Puccinia emaculata*).

COMT down-regulated switchgrass grown in the field for three growing seasons (2011–2013) had consistently lower lignin levels, reduced S/G ratios, and improved sugar release across all three years. By the end of year three, both transgenic lines tested produced 36–41% more ethanol than controls and produced equivalent biomass as controls with no difference in switchgrass rust incidence. The MYB4 over-expressing plants grown over three growing seasons (2012–2014) had decreased lignin, improved sugar release, and improved ethanol yields of up to 50%, with one line also producing 63% more biomass than the control in year two. The MYB4 transgenic events were similar to the control in rust susceptibility with the exception of event L1, which did not exhibit any rust symptoms for the duration of the experiments. Some miRNA156 overexpressing plants had modest improvement in lignin content and S/G ratios, and ethanol yields in year one, with one line exhibited consistently higher biomass yields relative to the control for years one and two. Rust susceptibility varied significantly among the different lines and among years. Downregulated CWG-1 plants produced more biomass, but had unchanged recalcitrance metrics by year 2, however, one transgenic line had a 23% increase in ethanol yield in year one. Although some CWG-1 lines showed increased rust susceptibility relative to the control in year one, all lines (transgenic and control) were not different than controls by the end of year two. One line of downregulated CWG-2 switchgrass plants had 27% higher biomass relative to the control, whereas a second line produced 7% more ethanol with no change in biomass production in year one. No changes in rust susceptibility were observed between CWG-2 transgenic lines and the control in year two. Several natural variant lines showed increased sugar release and improved biomass (up to 91% increase) in year one compared with the control, with one line also exhibiting a 9% increase in ethanol production. Data are currently being collected and analyzed for switchgrass plants with overexpression of the ERF/SHN transcription factor, and downregulation of CWG-3, both of which were planted in late spring of 2015 and have just completed their first growing season.

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

Using Glycome Profiling on Plant Biomass for Functional Characterization of Cell Wall Hydrolytic Enzymes

Sivakumar Pattathil[1,3]* (siva@ccrc.uga.edu), Maryam Mirzai,[1,3] Lai Bergeman,[2,3] Johnnie Walker,[2,3] Sindhu Kandemkavil,[1,3] Raha Saremi,[1,3] Sivasankari Venkatachalam,[1,3] Ronald Clay,[1,3] Brian Fox,[2,3] Michael G Hahn,[1,3] and **Paul Gilna**[3]

[1]Complex Carbohydrate Research Center, The University of Georgia, Athens; [2]Great Lakes Bioenergy Research Center and The Department of Biochemistry University of Wisconsin–Madison; [3]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org/

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) designing plant cell walls for rapid deconstruction and (2) developing multi-talented microbes or converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC researchers provide enabling technologies in biomass characterization, 'omics, modeling and data management in order to (1) understand chemical and structural changes within biomass and (2) to provide insights into biomass formation and conversion mechanisms.

Microbes, which are responsible for the degradation of plant biomass in the biosphere, produce wide arrays of glycosylhydrolases in order to mobilize the sugars in biomass. However, the hydrolytic activity and specificity of most of these hydrolases remain unknown. Most assays for delimiting the activity of glycosylhydrolases rely upon isolated polysaccharides as substrates for assays, and thus multiple assays are required to fully characterize newly identified putative hydrolases. We describe the use of intact biomass as a substrate for novel glycosylhydrolases, (provided by the Great Lakes Bioenergy Research Center), followed by antibody-based Glycome Profiling to reveal the activities of hydrolases capable of degrading biomass components. The enzymes examined in this study are: CelEcc_CBM3a (an GH5 family enzyme engineered for multi-functionality), XynY (a potentially processive hemicellulase from the GH10 family) and XynA (a xylanase from the GH11 family). Treatment of native or AFEX™ pretreated switchgrass and corn stover biomass with CelEcc_CBM3a, XynY and XynA showed that, in addition to the removal of xylans from the biomass, xyloglucan epitopes were also depleted, suggesting multi-functionality in the activities of these three enzymes. Thus, the biomass/Glycome Profiling-based assay system shows promise as a rapid, moderate throughput method for characterization of putative enzymes for use in biomass deconstruction.

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.

Integrated ′Omics Reveals the Details of Metabolic Adaptation of *Clostridium thermocellum* ATCC-27405 Grown on Switchgrass

Suresh Poudel,[1,2] Richard J. Giannone,[1] Miguel Rodriguez Jr.,[1] Babu Raman,[3] Madhavi Z. Martin,[1] Nancy L. Engle,[1] Intawat Nookaew,[1] Steve D. Brown,[1] Timothy J. Tschaplinski,[1] David Ussery,[1] Robert L. Hettich[1,2]* (hettichrl@ornl.gov), and **Paul Gilna**[1]

[1]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]Department of Genome Science and Technology, University of Tennessee, Knoxville, Tennessee; [3]Dow AgroSciences, Indianapolis, Indiana

http://bioenergycenter.org/

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

Switchgrass is a perennial C4 grass and a dominant grass in North America. Due to favorable growth characteristics, it is a model herbaceous bioenergy crop of significant interest to the U.S. Department of Energy. As such, switchgrass is a promising feedstock, capable of augmenting and potentially replacing corn for the production of bioethanol. The plant cell wall polysaccharides (cellulose, hemicelluloses, and lignin) are a source of carbon and energy for cellulolytic microorganisms that are capable of degrading these somewhat complex biopolymers. However, the biosolubilization of these recalcitrant biopolymers requires a sophisticated microbial enzymatic system with diverse catalytic activities.

Clostridium thermocellum is a thermophillic, anaerobic Gram-positive bacterium that produces large extracellular complexes, termed cellulosomes, that are quite efficient at solubilizing and deconstructing cellulose for eventual production into biofuel materials. This cellulolytic microbe is a candidate for converting lignocellulosic biomass directly into ethanol; however, the microbial capacity to overcome the recalcitrant nature of the lignocellulosic biomass necessary for biofuel production remains challenging to fully capture at an industrial scale. *C. thermocellum* can degrade cellulose from complex lignocellulosic feedstocks, such as switchgrass and *Populus*, to form cellobiose and other small, soluble cellodextrins as the main products. Cellobiose is ultimately utilized by the organism to generate end products of ethanol, acetic acid, lactic acid, hydrogen, and carbon dioxide. The generation of lactate, formate and acetate can divert the metabolic flux away from the desired ethanol product. In order to maximize the ethanol yield, *C. thermocellum* can be genetically modified to eliminate the competing pathways. Despite studies investigating *Clostridium thermocellum* on various cellulose substrates, relatively little work has been done to systematically characterize the comprehensive range of linked proteins/metabolites across

a detailed time-dependent growth of *Clostridium thermocellum* on switchgrass. In this study, we investigated *C. thermocellum* grown on switchgrass in batch fermentation by integrating a multi-′omics approach (proteomics, transcriptomics, metabolomics) to better understand the detailed molecular machinery and regulation of cellulolytic microbial growth on this complex lignocellulose substrate.

Acetic acid and ethanol were observed as the major fermentation products for *C. thermocellum* grown on dilute acid pretreated switchgrass. Ethanol stopped accumulating after 120 hours, whereas acetic acid slowly kept accumulating in the fermentation broth. The ethanol concentration profiles show that the cultures were most metabolically active between 19 and 45 h post-inoculation, which correspond to exponential growth phase in this study. Metabolomics provided key information about the range of end products from microbial growth on this complex biomass, and revealed major hemicellulose catabolic metabolites, including xylobiose, xylose, xylitol, arabinose, and arabitol. A large accumulation (almost 9-fold) of xylitol was measured as the culture reached stationary phase. Other 5-carbon sugar alcohols, such as arabitol, ribitol, and phenolic acids of the lignin pathway, including caffeic acid, ferulic acid and p-coumaric acid were also noted to increase dramatically in abundance; some of these compounds have been found to be inhibitory to microbial growth, and thus warrant further consideration. The accumulation of long chain saturated fatty acids, and unique branched fatty acids was quite notable in *C. thermocellum* during the stationary phase, indicating cell membrane modification to tolerate increasing concentration of fermentation end-products. To better understand this process, we used proteomic and transcriptomic analyses to characterize the range of enzymatic molecular machinery that is used by this microbe during complex cellulose solubilization.

Proteomic analysis of this system (*C. thermocellum* growing on switchgrass) revealed 1551 non-redundant proteins, representing ~50% of predicted proteins, with about a third of the proteins (566) having significant changes in abundance during transition from early exponential phase to late stationary phase. Most of the metabolically related enzymes (e.g., amino acid and protein synthesis, glycolysis) decreased in abundance with growth time; in contrast, cellulosomal proteins, S-layer, and bacterial secretion systems increased in abundance, even at late stationary phase. Clustering analyses revealed significant protein groups that varied substantially with time; these were mainly populated with ABC transporters, catabolic enzymes (cellulosomal proteins), and transport related proteins. Enzymes responsible for atypical glycolysis were highly abundant and remarkably time dependent, suggesting a malate shunt pathway. In response to damage of cellular membrane integrity caused by ethanol accumulation, *C. thermocellum* diverts activity into the pentose phosphate pathway and biosynthesis of branched fatty acids. In addition, we also found that valine, leucine and isoleucine producing enzymes changed significantly, suggesting a possible diversion of PEP/pyruvate towards amino acid biosynthesis. These results indicate that the growth of *C. thermocellum* on switchgrass is a complicated process in which plant substrate is used for microbial growth, but inevitable inhibitory end-products retard microbial metabolism and thus have to be considered in the potential application of this type of celluloytic organism for effective 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.

A New Low-Cost Paradigm for Biological Conversion of Cellulosic Biomass: Evaluation of Economic Potential

Andrew Allee[1,2]* (andrew.m.allee.th@dartmouth.edu), Mark Laser,[1,2] Rita Tu,[1,2] Annie Sun,[1,2] Michael Balch,[1,2] Evert Holwerda,[1,2] Xiaoyu Liang,[1,2] Lee Lynd,[1,2] and **Paul Gilna**[2]

[1]Thayer School of Engineering, Dartmouth College, Hanover, New Hampshire; [2]BioEnergy Research Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org/

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

Detailed comparative data has recently shown that *Clostridium thermocellum* achieves carbohydrate solubilization yields for several grasses with no pretreatment other than autoclaving that are $\geq$ 35% of theoretical, and higher than industry-standard fungal cellulase by 2 to 3-fold (Paye et al., 2016; Shao et al., 2011). Although mechanical milling has excessive energy requirements when used as a pretreatment for fungal cellulase, it is known that lignocellulose-containing slurries undergo a radical liquefaction at solubilization yields < 15%, as indicated by ~ 30-fold reductions in the two variables that determine milling energy: viscosity and particle strength. Noting that the yields achievable by *C. thermocellum* in the absence of pretreatment substantially excessed those needed to reduce milling energy requirements by over an order of magnitude, we are exploring whether non-biological disruption before biological processing (pretreatment) might be replaced by biological processing prior to non-biological disruption.

Recent data supporting this concept include:

- Two-fold greater enhancement of solubilization via milling partially-fermented solids as compared to unfermented-solids (Paye et al., 2016);
- Solubilization yields by cotreatment-enhanced *C. thermocellum* fermentation comparable to those obtainable using thermochemical pretreatment; and
- Milling at intensities sufficient to achieve high solubilization yields in the absence of thermochemical pretreatment has little or no effect on *C. thermocellum* whereas it results in complete cessation of yeast fermentation.

Here we report preliminary analysis of the potential economics of combining engineered thermophilic, cellulolytic microbes with cotreatment for conversion of corn stover to ethanol using the most recent National Renewable Energy Laboratory's design as a departure point. Our results, based on assumed success in several ongoing research directions, support potential for far lower processing costs as well as feasibility at much smaller scale compared to current technology.

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.

CELF Pretreatment Enhances Conversion of Biomass to Renewable Fuels

Charles M. Cai[1,4]* (Charles.cai@ucr.edu), Thanh Yen Nguyen,[1,4] Vanessa Thomas,[1,4] Abhishek Patri,[1,4] Ninad Kothari,[1,4] Rajeev Kumar,[1,4] Bryan Donohoe,[3,4] Yunqiao Pu,[2,4] Arthur Ragauskas,[2,4] Charles E. Wyman,[1,4] and **Paul Gilna**[4]

[1]University of California Riverside, Riverside; [2]University of Tennessee, Knoxville, Tennessee; [3]National Renewable Energy Laboratories, Golden, Colorado; [4]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org/

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

The University of California Riverside recently developed a novel pretreatment called Co-solvent Enhanced Lignocellulosic Fractionation (CELF) that applies renewable tetrahydrofuran (THF) in a monophasic solution with water and dilute acid to dramatically reduce biomass recalcitrance and enhance enzymatic, biological, and catalytic conversion of lignocellulosic feedstocks into renewable fuels and fuel precursors. CELF achieves unprecedented recovery of total sugars (C5 + C6) from multiple feedstocks using 90% less enzymes than conventional dilute sulfuric acid (DSA) pretreatment while boosting ethanol yields (>90%) and titers (>58 g/L) for high solids simultaneous saccharification and fermentation (SSF) operation at low enzyme loadings (<15 mg-enzyme/g-glucan). We also demonstrate the successful first integration of CELF pretreatment with enzyme-free consolidated bioprocessing (CBP), achieving almost 100% sugar conversion from *Populus* wood or corn stover in two days. During the CELF reaction, 90% or more of the native lignin from most lignocellulosic substrates is extracted and dissolved into the liquid hydrolysate. The dissolved lignin can then be precipitated as a purified low-molecular weight lignin product (CELF lignin) upon the removal and recovery of THF from the hydrolysate by room-temperature vacuum distillation. CELF lignin is highly suitable for valorization into renewable chemicals and fuels.

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

The Secretome of *Caldicellulosiruptor bescii*: Biomass Deconstruction without Conventional Pretreatment

Jenna Young[1,4]* (jmy715@uga.edu), Daehwan Chung,[1,4] Joseph Groom,[1,4] Minsoek Cha,[1,4] Debra Mohnen,[1,4] Richard Dixon,[2,4] Yannick J. Bomble,[3,4] Michael E. Himmel,[3,4] Janet Westpheling,[1,4] and **Paul Gilna**[4]

[1]University of Georgia, Athens; [2]University of North Texas, Denton; [3]National Renewable Energy Laboratory, Golden, Colorado; [4]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org

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

Members of the bacterial genus *Caldicellulosiruptor* are the most thermophilic cellulolytic organisms so far described and have the ability to grow on lignocellulosic biomass without conventional pretreatment. A comparison of the pangenome of this genus suggests that among the genes important for cellulolytic ability are *celA* and a cluster of genes involved in pectin degradation (1). CelA is a bi-functional glycoside hydrolase that contains a Family 9 endoglucanase and a Family 48 exoglucanase joined by three Family 3 carbohydrate binding modules (CBMs), and while there are two Family 9 and three Family 48 glycoside hydrolases in *C. bescii*, CelA is the only protein that combines both activities. Deletion of CelA resulted in a significant decrease in cellulolytic activity (2). Expression of full length CelA in *C. bescii* revealed that extracellular CelA protein is glycosylated whereas intracellular CelA is not (3). The mechanism and role of protein glycosylation in bacteria is poorly understood and the ability to express CelA *in vivo* in *C. bescii* will allow the study of the mechanism of protein glycosylation in this thermophile. Moreover, heterologous expression of an additional thermophilic endoglucanase (E1) from *Acidothermus cellulolyticus*, enhanced the ability of *C. bescii* to deconstruct biomass demonstrating that while extremely effective, the ability of *C. bescii* to deconstruct plant biomass can be improved (4).

The *C. bescii* genome contains five genes predicted to be involved in pectin deconstruction/ utilization and three exist in a cluster with a predicted transcriptional regulator. Expression of this cluster is significantly up-regulated in cells growing on biomass (5). Most biomass models do not list pectin because of its low abundance in grass walls and in dicot secondary walls (6). Recent work has shown that pectin is synthesized in secondary walls (7), that some pectin biosynthetic enzymes are amplified in grasses (8), and that saccharification of plant biomass can be improved by modifying the structure of pectin (9). Deletion of the pectinase gene cluster in *C. bescii* resulted in a mutant reduced in its ability to grow on dicot and grass biomass (10). The phenotype of the *C. bescii* pectinase mutant provides direct genetic evidence that pectin is a significant barrier to

deconstruction of unpretreated plant biomass by *C. bescii* and that pectin plays an important and underappreciated role in biomass recalcitrance.

References

1. Blumer-Schuette SE, Giannone RJ, Zurawski JV, Ozdemir I, Ma Q, Yin YB, et al., "*Caldicellulosiruptor* Core and Pangenomes Reveal Determinants for Noncellulosomal Thermophilic Deconstruction of Plant Biomass," *Journal of Bacteriology* 2012;194(15):4015-28.

2. Young J, Chung D, Bomble YJ, Himmel M, Westpheling J., "Deletion of *Caldicellulosiruptor bescii* CelA Reveals its Crucial Role in the Deconstruction of Lignocellulosic Biomass," *Biotechnology for Biofuels*. 2014;7:142.

3. Chung D, Young J, Bomble YJ, Vander Wall TA, Groom J, Himmel ME, et al., "Homologous Expression of the *Caldicellulosiruptor bescii* CelA Reveals that the Extracellular Protein Is Glycosylated," PloS One 2015;10(3):e0119508. Epub 2015/03/24.

4. Chung D, Young J, Cha M, Brunecky R, Bomble YJ, Himmel ME, et al., "Expression of the Acidothermus Cellulolyticus E1 Endoglucanase in *Caldicellulosiruptor bescii* Enhances its Ability to Deconstruct Crystalline Cellulose," *Biotechnology for Biofuels* 2015;8:113. Epub 2015/08/14.

5. Kataeva I, Foston MB, Yang SJ, Pattathil S, Biswal AK, Poole FL, et al., "Carbohydrate and Lignin are Simultaneously Solubilized from Unpretreated Switchgrass by Microbial Action at High Temperature," Energy and Environmental Science 2013;6(7):2186-95.

6. Mohnen D, Bar-Peled M, Somerville C., "Cell Wall Polysaccharide Synthesis," *Biomass Recalcitrance*: Blackwell Publishing Ltd.; 2009. p. 94-187.

7. Atmodjo MA, Sakuragi Y, Zhu X, Burrell AJ, Mohanty SS, Atwood JA, III, et al., "Galacturonosyltransferase (GAUT)1 and GAUT7 are the Core of a Plant Cell Wall Pectin Biosynthetic Homogalacturonan:Galacturonosyltransferase Complex," *Proceedings of the National Academy of Sciences of the United States of America* 2011;108(50):20225-30. Epub 2011/12/03.

8. Yin Y, Chen H, Hahn MG, Mohnen D, Xu Y., "Evolution and Function of the Plant Cell Wall Synthesis-Related Glycosyltransferase Family 8," *Plant Physiology* 2010;153(4):1729-46. Epub 2010/06/05.

9. Lionetti V, Francocci F, Ferrari S, Volpi C, Bellincampi D, Galletti R, et al., "Engineering the Cell Wall by Reducing De-Methyl-Esterified Homogalacturonan Improves Saccharification of Plant Tissues for Bioconversion," *Proceedings of the National Academy of Sciences of the United States of America* 2010;107(2):616-21. Epub 2010/01/19.

10. Chung D, Pattathil S, Biswal AK, Hahn MG, Mohnen D, Westpheling J., "Deletion of a Gene Cluster Encoding Pectin Degrading Enzymes in *Caldicellulosiruptor bescii* Reveals an Important Role for Pectin in Plant Biomass Recalcitrance," *Biotechnology for Biofuels* 2014;7(1):147. Epub 2014/10/18.

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.

Multiple Levers to Solve Recalcitrance for Lignocellulosic Solubilization

Evert K. Holwerda[1,2] (evert.k.holwerda.th@dartmouth.edu), Robert S. Worthen,[1,2] Natalie M. Ruppertsberger,[1,2] Michael L. Balch,[1,2] Brian H. Davison,[2] Zeng-Yu Wang,[3,2] Richard A. Dixon,[4,2] Debra Mohnen,[5,2] C. Neal Stewart,[6,2] Daniel G Olsen,[1,2] Gerald Tuskan,[2] Mark Davis,[7,2] Michael E. Himmel,[7,2] Jeffrey V. Zurawski,[8,2] Robert M. Kelley,[8,2] Charles E. Wyman,[9,2] Lee R. Lynd,[1,2] and **Paul Gilna**[2]

[1]Dartmouth College, Hanover, New Hampshire; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]The Samuel Roberts Noble Foundation, Ardmore, Oklahoma; [4]University of North Texas, Denton, Texas; [5]University of Georgia, Athens; [6]University of Tennessee, Knoxville; [7]National Renewable Energy Laboratory, Golden, Colorado; [8]North Carolina State University, Raleigh; [9]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 multi-talented microbes or converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. 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 testing large numbers of natural variants and generating specific and modified plant samples as well as developing genomics tools for detailed studies into poorly understood cell wall biosynthesis pathways. BESC research in biomass deconstruction and conversion targets CBP by studying thermophilic anaerobes to understand novel strategies and enzyme complexes for biomass deconstruction and manipulating these microorganisms for improved conversion, yields, and biofuel titer. BESC researchers provide enabling technologies in biomass characterization, 'omics, modeling and data management in order to (1) understand chemical and structural changes within biomass and (2) to provide insights into biomass formation and conversion mechanisms.

The primary barrier to economically competitive cellulosic biofuels is the resistance of plant cell walls to deconstruction – termed recalcitrance. Overcoming this barrier may be accomplished via multiple recalcitrance "levers" including:

1. starting with nature's best with respect to feedstocks and biocatalysts;
2. biotechnology to improve plants, enzymes, and microbes; and
3. non-biological processing prior to or during solubilization and fermentation.

Here we will report progress at applying and evaluating these levers. Studies aimed at individual levers include:

- comparison of the effectiveness of various biocatalysts at mediating plant cell wall solubilization;
- targeted modification of plants to decrease recalcitrance without sacrificing growth;
- targeted modification of thermophilic anaerobes to improve ethanol yield and titer without sacrificing fermentation capacity; and
- mechanical milling during fermentation, termed co-treatment.

In addition, we have evaluated solubilization of several second year field-trial lignocellulosic feedstocks, engineered for reduced recalcitrance with three different microbial solubilization systems; *Caldicellulosiruptor bescii*, *Clostridium thermocellum* and *Saccharomyces cerevisiae* with commercial fungal enzymes. This first-of-a-kind dataset will be used to gain new insights into the interplay between plant modification and choice of biocatalyst.

Based on our results, we conclude that there are powerful, emergent strategies to overcoming plant cell wall recalcitrance, and that systematically exploring combinations of these levers is a promising approach to enabling the cost-effective production of cellulosic biofuels.

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

Consolidated Bioprocessing of Cellulose to Isobutanol in *Clostridium thermocellum*

Paul P. Lin[1,2]* (phlin@ucla.edu), Luo Mi,[1,2] Lemuel M. Soh,[1,2] Amy H. Morioka,[1,2] Kouki M. Yoshino,[1] Sawako Konishi,[1] Sharon C. Xu,[1] Beth A. Papanek,[2,3] Lauren A. Riley,[2,3] Adam M. Guss,[2,3] James C. Liao,[1,2] and **Paul Gilna**[2]

[1]Department of Chemical and Biomolecular Engineering, University of California, Los Angeles; [2]BioEnergy Science Center, Oak Ridge, National Laboratory, Oak Ridge, Tennessee; [3]Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville

http://bioenergycenter.org

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

CBP has the potential to reduce biofuel or biochemical production costs by processing cellulose hydrolysis and fermentation simultaneously without the addition of pre-manufactured cellulases. In particular, *Clostridium thermocellum* is a promising thermophilic CBP host because of its high cellulose decomposition rate. Here we report the engineering of *C. thermocellum* to produce isobutanol. Metabolic engineering for isobutanol production in *C. thermocellum* is hampered by enzyme toxicity during cloning, time-consuming pathway engineering procedures, and slow turnaround in production tests. In this work, we first cloned essential isobutanol pathway genes under different promoters to create various plasmid constructs in *Escherichia coli*. Then, these constructs were transformed and tested in *C. thermocellum.* Among these engineered strains, the best isobutanol producer was selected and the production conditions were optimized. We confirmed the expression of the overexpressed genes by their mRNA quantities. We also determined that both the native ketoisovalerate oxidoreductase (KOR) and the heterologous ketoisovalerate decarboxylase (KIVD) expressed were responsible for isobutanol production. We further found that the plasmid was integrated into the chromosome by single crossover. The resulting strain was stable without antibiotic selection pressure. This strain produced 5.4 g/L of isobutanol from cellulose in minimal medium at 50^{o}C within 75 h, corresponding to 41% of theoretical 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.

Hemicellulose Biosynthesis is Becoming Crystal Clear

Breeanna Urbanowicz[1,3]* (breeanna@uga.edu), Vladimir Lunin,[2,3] Petri Alahuhta,[2,3] Maria Peña,[1,3] Sami Tuomivaara,[1,3] Shuo Wang,[1,3] Michael Crowley,[2,3] Vivek Bharadwaj,[2,3] Jeong Yeh Yang,[1,3] Mike Himmel,[2,3] Kelley Moremen,[1,3] William York,[1,3] and **Paul Gilna**[3]

[1]University of Georgia, Athens; [2]National Renewable Energy Laboratory, Golden, Colorado; [3]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) designing plant cell walls for rapid deconstruction and (2) developing multi-talented microbes or converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC 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 testing large numbers of natural variants and generating specific and modified plant samples as well as developing genomics tools for detailed studies into poorly understood cell wall biosynthesis pathways. BESC researchers provide enabling technologies in biomass characterization, 'omics, modeling and data management in order to (1) understand chemical and structural changes within biomass and (2) to provide insights into biomass formation and conversion mechanisms.

The synthesis and assembly of polysaccharides to form the plant cell wall is a complex process. Glycosyltransferases (GT) are the ubiquitous enzymes responsible for creating the diverse and complex array of oligosaccharides and glycopolymers found in nature. Although several of the key glycosyltransferases involved in plant polysaccharide synthesis have been identified, none have been crystallized to date. The structure and mechanism of cellulose synthase (CesA), perhaps the most widely studied glycosyltransferase involved in cell wall biogenesis, is becoming clearer as a result of crystal diffraction analysis of bacterial orthologs, but the plant CesA and other glycosyltransferases involved in the biosynthesis of plant cell walls have resisted crystallographic analysis.

Recently, we have developed methods to express highly active forms of several glycosyltransferases involved in the biosynthesis of the hemicelluloses xylan and xyloglucan, which are abundant components of plant biomass. We are currently focusing on understanding the mechanisms by which these enzymes catalyze the highly substrate-specific and regio-specific transfer of sugar residues to the growing polysaccharides. The first enzyme we have studied in depth is *At*FUT1, a glycosyltransferase in the GT37 family that catalyzes the regio-specific transfer of fucosyl residues to the sidechains of xyloglucan. We have crystallized this enzyme alone, as a complex with its donor substrate (GDP-fucose), and with a xyloglucan octasaccharide acceptor substrate (XXLG). This resulted in the first crystal structures for any plant glycosyltransferase involved in cell wall polysaccharide biosynthesis and the first crystal structure of a GT37 enzyme. We have also used site-directed mutagenesis to modify specific amino acids in the *At*FUT1 sequence that are critical for its catalytic activity, biochemically characterized these mutants, and interpreted the results in the context of the crystal structure. We are currently using the combined structural and biochemical information, along with molecular dynamics simulations, to gain insight into the

key mechanisms by which *At*FUT1 and other members of the GT37 enzyme family fulfill their biochemical functions. By providing the archetypal crystal structure for GT37 enzymes, these studies also facilitate the structural modeling of other members of this family.

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.

Absolute Quantitation of Intracellular Metabolite Concentrations in *C. thermocellum* and *T. saccharolyticum*

Tyler Jacobson,[1,2] David M. Stevenson,[1,2] Daniel Olson,[2,3] Lee R. Lynd,[2,3] Daniel Amador-Noguez[1,2]* (amadornoguez@wisc.edu), and **Paul Gilna**[2]

[1]University of Wisconsin-Madison, Madison; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]Dartmouth College, Hanover, New Hampshire

http://bioenergycenter.org

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

To provide quantitative metabolomics data that supports the development of mathematical metabolic models in *C. thermocellum* and *T. saccharolyticum*, we have used LC-MS/MS and an isotope-ratio based approach to measure intracellular metabolite concentrations of central carbon metabolites (Glycolysis, TCA cycle, amino acids, nucleotides, cofactors, etc.) in *C. thermocellum* and several engineered *T. saccharolyticum* strains. The strains were grown on minimal media on cellubiose. We have now quantitated a set of ~100 metabolites. Our results show large differences in the metabolome of *C. thermocellum* vs. *T. saccharolyticum* and point to metabolic bottlenecks and inefficiencies that could be corrected via metabolic engineering. These measurements and results will facilitate the development of quantitative and dynamic models of *T. saccharolyticum* and *C. thermocellum* metabolism and ethanol fermentation pathways.

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.

Caldicellulosiruptor Pan-Genomics: Perspectives on Newly Sequenced Species, and Genus-Wide Diversity of Cellulose Binding Proteins (tāpirins)

Laura L. Lee[1,2]* (lllee@ncsu.edu), Sara E. Blumer-Schuette,[1,2] Javier A. Izquierdo,[1,2] Kasey E. Koballa,[1,2] Kevin T. Mullen,[1,2] Jonathan M. Conway,[1,2] Jeffrey V. Zurawski,[1,2] Piyum A. Khatibi,[1,2] Robert M. Kelly,[1,2] and **Paul Gilna**[2]

[1]North Carolina State University, Raleigh; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org

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

Extremely thermophilic organisms have a promising, but yet unrealized, role to play in the microbial production of lignocellulosic ethanol. Discovery of their potential has been pursued through genomics (metagenomics and pan-genomics), as well as the investigation of novel cellulose binding proteins (tāpirins). First, samples from Obsidian Pool in Yellowstone National Park were metagenomically sequenced to identify thermophilic cellulases and ethanol-forming enzymes. The 16S rRNA and Illumina DNA sequences revealed novel enzymes and organisms of interest, while PacBio sequencing has been used to obtain longer reads and potentially closed genomes. Next, we characterized the *Caldicellulosiruptor* core and pan-genomes with the GET_HOMOLOGUES software in order to determine the genetic diversity within the genus. Presently, the core genome contains 1284 genes, but the pan-genome is open, as the size is still increasing as the number of sequenced isolates grows. The three newest species' genomes (*C.* sp. str. Rt8.B8, *C.* sp. str. Wai35.B5, and *C.* sp. str. NA10) were examined for the presence of new or uncommon glycoside hydrolases (GH) and surface layer homology domain proteins. One of the most interesting finds was a multi-modular enzyme with a GH12 domain (along with a GH48, multiple CBMs, and either a GH5 or GH10); these are the first *Caldicellulosiruptor* species sequenced at this point to have a multi-modular CAZyme with a GH12 domain. Finally, novel proteins (tāpirins) were identified via transcriptomics and proteomics to be highly expressed in cellulose-bound fractions. Structural homology to other classes of proteins could not be assigned, indicating that this is truly a new class of biomolecules and establishing a new paradigm for how cellulolytic bacteria adhere to cellulose. Overall, these results bring a more comprehensive understanding of the *Caldicellulosiruptor* genus, as well as shed light on novel CAZymes and proteins in both characterized and novel species, which could have important roles in how these microbes degrade polysaccharides.

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.

Understanding the Hyperactive Multi-Component Cellulase: CelA

Roman Brunecky,[1,3] Bryon S. Donohoe,[1,3] Levi Kramer,[1,3] John M. Yarbrough,[1,3] Ashutosh Mittal,[1,3] Daehwan Chung,[1,3] Jenna Young,[2,3] Todd A. Vander Wall,[1,3] Larry E. Taylor,[1,3] Janet Westpheling,[2,3] Michael E. Himmel,[1,3] Yannick J. Bomble[1,3]* (yannick.bomble@nrel.gov), and **Paul Gilna**[3]

[1]National Renewable Energy Laboratory, Golden, Colorado; [2]University of Georgia, Athens; [3]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://bioenergycenter.org

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

Unlike fungal systems, which typically comprise a number of single catalytic domain enzymes for biomass degradation such as family 7 and family 5 glycoside hydrolases (GH), some bacterial systems utilize an alternative strategy with tethered multi-catalytic domain cellulases. The cellulase CelA from the thermophile *Caldicellulosiruptor bescii*, which is one of the most active cellulose degrading enzymes known to date, is one such example. In the saccharification of a common cellulose standard, Avicel, CelA outperforms mixtures of commercially relevant exo- and endoglucanases. The modular architecture of CelA is defined as: GH9-CBM3-CBM3-CBM3-GH48 and the enzyme is extremely thermostable and highly active at elevated temperatures.

From transmission electron microscopy studies of biomass following incubation with CelA, we have discovered morphological features that suggest CelA utilizes a biomass digestion mechanism different from the common surface ablation strategy driven by processivity and we propose that CelA and possibly other multi-functional glycoside hydrolases, act in a novel manner when compared to traditional fungal enzyme systems. We have explored the activity of CelA on a variety of pretreated substrates in order to better understand how the different bulk components of biomass, such as xylan and lignin impact CelA activity and how the effect of those bulk components may differ between traditional fungal enzymes and CelA. We have also examined the impact of cellulose crystallinity on the respective cellulolytic activity of these two cellulase systems.

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.

Tension Wood Provides Insight Into Structural Changes in Biomass Resulting from Chemical Pretreatment

Sai Venkatesh Pingali[1]* (pingalis@ornl.gov), Daisuke Sawada,[1] Udaya Kalluri,[1] Hugh M. O'Neill,[1] Volker Urban,[1] Paul Langan,[1] and **Brian H. Davison**[1]

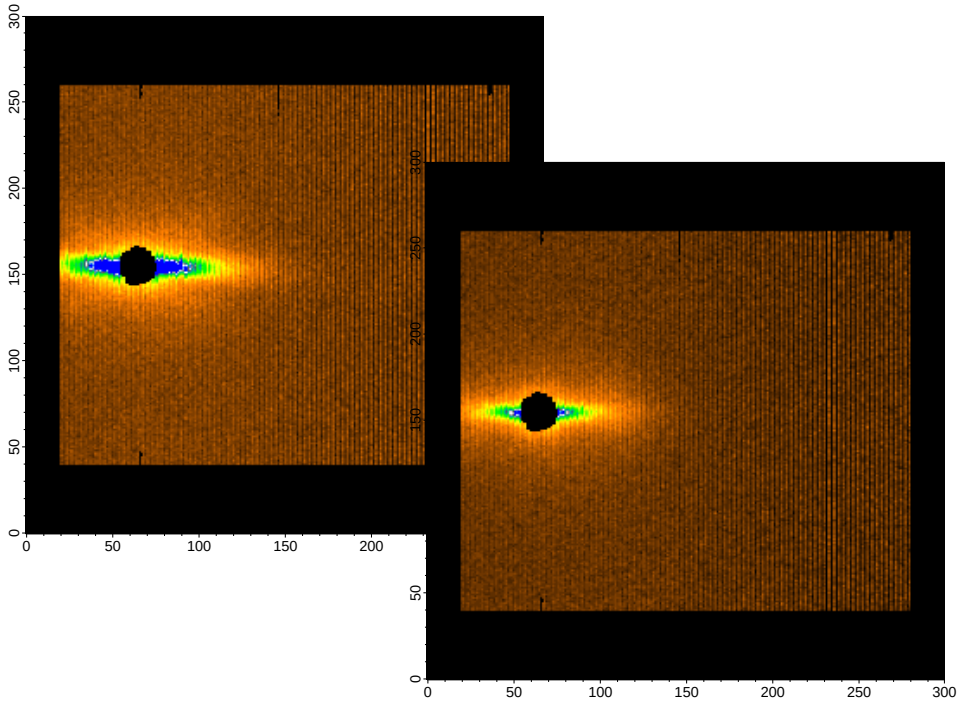
[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://cmb.ornl.gov/research/bioenergy/lignocellulose-dynamics

Project Goals: Lignocellulosic biomass comprises the vast majority of biomass on Earth and has the potential to play a major role in generation of renewable biofuels if cost-effective conversion can be achieved. Largely composed of plant cell walls, it is a complex biological composite material that is recalcitrant to the structural deconstruction and enzymatic hydrolysis into sugars that is necessary for fermentation to bioethanol. The Scientific Focus Area in Biofuels is developing "Dynamic Visualization of Lignocellulose Degradation by Integration of Neutron Scattering Imaging and Computer Simulation" for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis. This is providing fundamental information about the structure and deconstruction of plant cell walls that is needed to drive improvements in the conversion of renewable lignocellulosic biomass to biofuels.

Plant cell walls comprise the bulk of lignocellulosic biomass. Therefore, a greater understanding of the chemistry, architecture, physical and mechanical properties of cell walls is essential to improve biomass-based biofuel production. Here we propose to understand biomass properties underlying recalcitrance to enzyme-based sugar release. Towards this end, the spatio-temporal progression in chemical and anatomical changes that occur in growing plants under tension stress, called tension wood, were investigated. Hydrolysis of tension wood gives 3-fold increased glucose yield compared to normal wood.[1] Limited structural insights from previous work postulated that the absence of characteristic diffraction peaks stemmed from laterally aligned microfibrils.[2]

Therefore, we carried out structural studies of tension wood using small-angle neutron scattering performed at the Bio-SANS instrument.[3] Tension wood was produced from *Populus tremula x alba* grown subjected to tension stress and normal (or control) wood from a plant grown in identical conditions except for the application of the tension stress. Intact microtomed samples prepared from plants consisted of normal, opposite (to the tension-stressed site), and tension wood samples. The normal and tension wood samples resulted in an anisotropic scattering pattern with the latter sample showing enhanced anisotropy (see figure). Although opposite wood also showed evidence of anisotropy, it was much less pronounced than tension and normal wood samples (not in figure). Structural features observed for normal wood sample at the higher scattering angles were modeled as a long cylinder with a cross-section that represents the crystalline cellulose elementary fibril (R_g = 10 ~ 11 Å).[4] The same feature was evident in the tension wood sample, however, multiple sizes were required describe the data optimally. Size distribution analysis of the scattering curve showed evidence of association between elementary fibrils forming different sizes of aggregates. Most interestingly, when the sizes of all the aggregates were summarized, distinct multiples of the elementary fibrils were obtained such as twice and thrice the size of an average elementary fibril. This propensity of the cellulose microfibrils to associate due to the application of tension stress mimics plant nanostructural evolution as observed during most thermochemical pretreatments, especially dilute acid pretreatment.



SANS 2d image of tension (left) and normal (right) wood samples depicting the differing degree of anisotropy in the scattering profiles.

References

1. M. Foston, et al., "Chemical, ultrastructural and supramolecular analysis of tension wood in *Populus tremula x alba* as a model substrate for reduced recalcitrance." *Energy Environ. Sci.* **2011**, 4, 4962–4971.
2. Y. Nishiyama, et al., "Periodic disorder along ramie cellulose microfibrils." *Biomacromolecules.* **2003**, 4, 1013–1017.
3. The development of cell was partially funded by the Oak Ridge National Laboratory's Center for Structural Molecular Biology (CSMB), which is supported by the Office of Biological and Environmental Research, using facilities supported by the U.S. Department of Energy, managed by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725.
4. S.V. Pingali et al., "Breakdown of cell wall nanostructure of dilute acid pretreated switchgrass." *Biomacromolecules* **2010**, 11, 2329-2335.

Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. This program is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Investigating Biopolymer Structural Evolution During Different Biomass Pretreatments by In Situ SANS Studies

Volker Urban[1]* (urbanvs@ornl.gov), Sai Venkatesh Pingali,[1] Hugh M. O'Neill,[1] Barbara Evans,[1] Riddhi Shah,[1] Daisuke Sawada,[1] Seema Singh,[2,3] Paul Langan,[1] and **Brian H. Davison**[1]

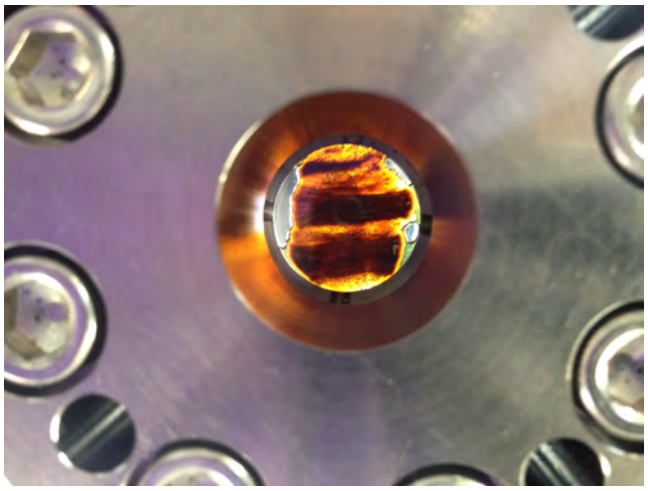
[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]Sandia National Laboratories, Livermore, California; [3]Joint BioEnergy Institute, Emeryville, California

http://cmb.ornl.gov/research/bioenergy/lignocellulose-dynamics

Project Goals: Lignocellulosic biomass comprises the vast majority of biomass on Earth and has the potential to play a major role in generation of renewable biofuels if cost-effective conversion can be achieved. Largely composed of plant cell walls, it is a complex biological composite material that is recalcitrant to the structural deconstruction and enzymatic hydrolysis into sugars that are necessary for fermentation to bioethanol. The Scientific Focus Area in Biofuels is developing "Dynamic Visualization of Lignocellulose Degradation by Integration of Neutron Scattering Imaging and Computer Simulation" for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis. This is providing fundamental information about the structure and deconstruction of plant cell walls that is needed to drive improvements in the conversion of renewable lignocellulosic biomass to biofuels.

In situ small-angle neutron scattering (SANS) was employed to examine real-time breakdown of biomass. It has previously been shown that the component biopolymers—cellulose, hemicellulose and lignin—undergo varying structural and organizational changes depending on the thermochemical treatment employed. For instance, an in situ SANS study of steam explosion pretreatment of *Populus* showed that lignin forms aggregates during heating, while hemicellulose is hydrolyzed, promoting growth of larger cellulose crystallites. The presence of lignin aggregates and higher cellulose crystallites are thought to reduce ethanol yields by inhibiting enzymatic hydrolysis of cellulose. One approach to address these issues is the preparation of transgenic materials with lower lignin proportions.[1] Recent studies have shown switchgrass lines downregulated in their lignin pathway (i.e., COMT-kd) have wild-type biomass yields but require reduced pretreatment severity and much lower cellulase dosages for equivalent product yields providing lower processing costs.[2]

Here, we report real-time SANS experiments of dilute acid pretreatment of transgenic and native switchgrass using a temperature-pressure reaction cell we developed.[3] The temperature of the cell was ramped-up from room to high temperatures (150–180°C), and after resident times between 5 to 60 min depending on the chemical treatment, the cell was cooled down to room temperature. We observed in the native switchgrass sample a structural feature that progressively moved to smaller-Q with increasing temperature and residence time at 180°C, indicating a growth in the particle size.[4] This feature which represents particles that appear at 120°C is interpreted as lignin aggregates; and this feature, consistent with our interpretation, is absent for the transgenic modified lignin material. Further, results of other pretreatment studies, ammonia and ionic liquid, of deuterated switchgrass and native poplar will also be presented. The use of deuterated switchgrass enhances the contrast thereby enabling our ability to resolve biomass structural features as observed during ionic liquid pretreatment which otherwise would not be possible. In situ SANS studies in conjunction with deuteration and contrast capabilities provide a unique vision into the nano-structural evolution of biomass leading to improvements critical for better efficiencies of the different industrial thermochemical pretreatments.



New improved temperature-pressure cell to perform in situ SANS experiments of thermochemical pretreatment of biomass.

References

1. F. Chen et al., "Lignin modification improves fermentable sugar yields for biofuel production." *Nat. Biotech.* **2007**, 25, 759–761.
2. C. X. Fu et al., "Genetic manipulation of lignin reduces recalcitrance and improved ethanol production from switchgrass." *PNAS* **2011**, 108, 3803–3808.
3. The development of the cell was partially funded by the Oak Ridge National Laboratory's Center for Structural Molecular Biology (CSMB), which is supported by the Office of Biological and Environmental Research, using facilities supported by the U.S. Department of Energy, managed by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725.
4. S. V. Pingali et al., "Breakdown of cell wall nanostructure of dilute acid pretreated switchgrass." *Biomacromolecules* **2010**, 11, 2329–2335.

Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. This program is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Developing Semi-Synthetic Composite Materials for Investigating Cellulose and Matrix Polymer Interactions During Pretreatment

Hugh O'Neill[1]* (oneillhm@ornl.gov), Riddhi Shah,[1] Daisuke Sawada,[1] Sai Venkatesh Pingali,[1] Loukas Petridis[1], Volker Urban,[1] Paul Langan, [1] Arthur J. Ragauskas,[1,2] Barbara R. Evans,[1] Jeremy C. Smith,[1,3] and **Brian H. Davison**[1]

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]University of Tennessee, Department of Chemical and Biomolecular Engineering and the Department of Forestry, Wildlife, and Fisheries, Knoxville; [3]University of Tennessee, Department of Biochemistry and Cellular and Molecular Biology, Knoxville

http://cmb.ornl.gov/research/bioenergy/lignocellulose-dynamics

Project Goals: Lignocellulosic biomass comprises the vast majority of biomass on Earth and has the potential to play a major role in generation of renewable biofuels if cost-effective conversion can be achieved. Largely composed of plant cell walls, it is a complex biological composite material that is recalcitrant to the structural deconstruction and enzymatic hydrolysis into sugars that is necessary for fermentation to bioethanol. The Scientific Focus Area in Biofuels is developing "Dynamic Visualization of Lignocellulose Degradation by Integration of Neutron Scattering Imaging and Computer Simulation" for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis. This is providing fundamental information about the structure and deconstruction of plant cell walls that is needed to drive improvements in the conversion of renewable lignocellulosic biomass to biofuels.

Plant cell walls comprise of three main components cellulose, hemicellulose, and lignin. The spatial arrangement of these components with respect to each other and changes among them during pretreatment of biomass are not clear. This is due to the complex and polymeric nature of plant cell walls. In order to understand the interactions of the plant cell wall components during pretreatment, we are developing semi-synthetic composite materials comprising of binary combinations of hemicellulose, lignin, and cellulose. In this study, hemicellulose-cellulose composites were prepared by synthesizing bacterial cellulose from *Acetobacter xylinus* in presence of glucomannan or xyloglucan dissolved in the growth media. Quantitative saccharification was carried out and the ratio of glucomannan to cellulose was 1:3 while the xyloglucan to cellulose was 1:5.

X-ray diffraction showed a significant decrease in the crystal size and crystallinity of cellulose synthesized in presence of hemicelluloses. In situ small angle neutron scattering (SANS) was used to study morphological changes during dilute acid pretreatment. The cellulose in the composite material was deuterated to provide contrast between it and the hemicellulose. SANS measurements were performed in 45% D_2O solvent, the contrast match point for protiated hemicellulose, making it possible to extract size and shape information of cellulose. The samples were heated to 170°C at 5°C/min, held at this temperature for 5 mins, before cooling to 25°C. SANS profiles were recorded at 1 min intervals. Structural changes in cellulose in presence of hemicelluloses (either xyloglucan or glucomannan) were observed during the temperature rise from 120 to 170°C. Analysis of the changes in cellulose structure and morphology supported by X-ray diffraction and compositional analysis will be presented.

Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. This program is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

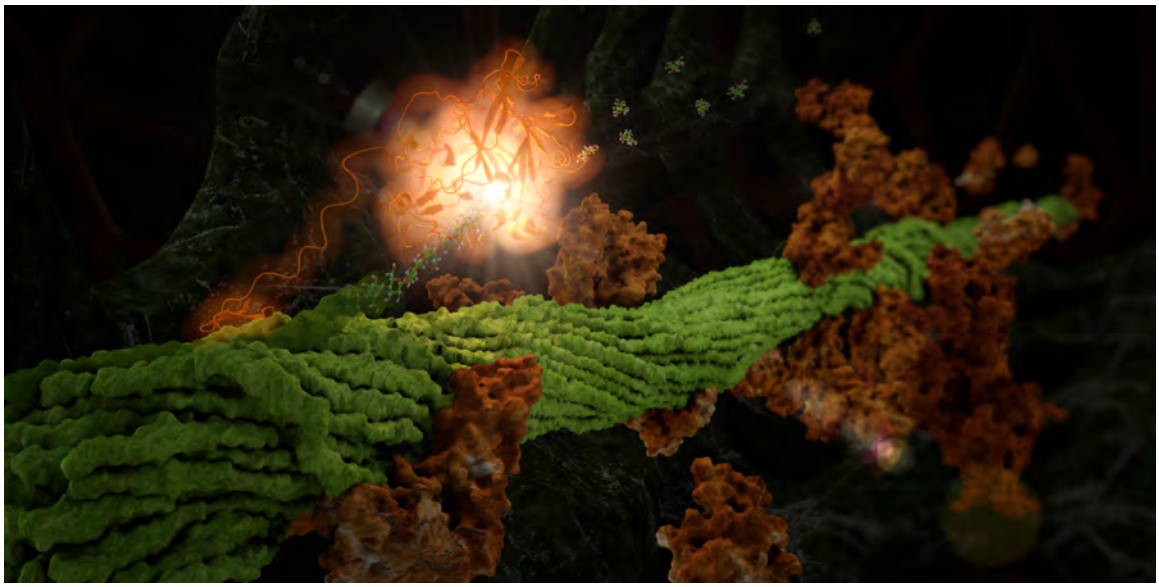
Why Is Lignin so Effective at Stopping Enzymes from Hydrolyzing Cellulose and How Does Heat-Treatment Change Lignin Characteristics?

Loukas Petridis[1]* (petridisl@ornl.gov), Derya Vural,[1] Josh Vermaas,[1] Hugh M. O'Neill,[1] Yunqiao Pu,[1] Sai Venkatesh Pingali,[1] Volker Urban,[1] Paul Langan,[1] Daisuke Sawada,[1] Riddhi Shah,[1] Arthur J. Ragauskas,[1,2] Barbara R. Evans,[1] Jeremy C. Smith,[1,3] and **Brian H. Davison**[1]

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]University of Tennessee, Department of Chemical and Biomolecular Engineering and the Department of Forestry, Wildlife, and Fisheries, Knoxville; [3]University of Tennessee, Department of Biochemistry and Cellular and Molecular Biology, Knoxville

http://cmb.ornl.gov/research/bioenergy/lignocellulose-dynamics

Project Goals: Lignocellulosic biomass comprises the vast majority of biomass on Earth and has the potential to play a major role in generation of renewable biofuels if cost-effective conversion can be achieved. Largely composed of plant cell walls, it is a complex biological composite material that is recalcitrant to the structural deconstruction and enzymatic hydrolysis into sugars that are necessary for fermentation to bioethanol. The Scientific Focus Area in Biofuels is developing "Dynamic Visualization of Lignocellulose Degradation by Integration of Neutron Scattering Imaging and Computer Simulation" for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis. This is providing fundamental information about the structure and deconstruction of plant cell walls that is needed to drive improvements in the conversion of renewable lignocellulosic biomass to biofuels.



Lignin (brown) impeding the hydrolysis of cellulose (green) by a cellulose enzyme (bright orange).

By performing atomic-detail molecular dynamics (MD) simulation of a biomass model containing cellulose, lignin, and cellulases (TrCel7A), we elucidate detailed lignin inhibition mechanisms. We find that lignin binds preferentially both to the elements of cellulose to which the cellulases also preferentially bind (the hydrophobic faces) and also to the specific residues on the cellulose-binding module of the cellulase that are critical for cellulose binding of TrCel7A (i.e., amino acids at Y466, Y492, and Y493). Lignin thus binds exactly where for industrial purposes it is least

desired, providing a simple explanation of why hydrolysis yields increase with lignin removal. To overcome lignin recalcitrance, biomass is usually subject to thermochemical pretreatment at temperatures T~160–180°C, above the lignin glass transition temperature (Tg), prior to biomass deconstruction. Pretreatment induces changes in biomass that promote its deconstruction, such as redistribution of the lignin that dramatically open up the biomass structure and improve the accessibility of the cellulose. Lignin re-distribution is enabled by temperature-dependent enhanced lignin dynamics. Here we probe the dynamics of lignin combining MD simulation and neutron scattering experiments and obtain insights in changes that occur in lignin during pretreatment.

Reference

J. V. Vermaas, L. Petridis, X. Qi, R. Schulz, B. Lindner, J. C. Smith, *Biotechnology for Biofuels* **2015**, 8, 1-16.

Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. This program is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

A Metabolic and Gene-Expression Model Reveals New Insight Into the Acetogen *Clostridium ljungdahlii*

J. Liu[1]* (jkl055@ucsd.edu), A. Ebrahim[1], C. Lloyd[1], J. Kim[1], M.M. Al-Bassam[1], **K. Zengler**[1]

[1]University of California, San Diego, 9500 Gilman Drive 0419 La Jolla, CA 92093

http://sbrg.ucsd.edu

Project Goals: We have reconstructed a metabolic and gene-expression model (ME-model) for the acetogen *Clostridium ljungdhalii*. This model details the organism's interconnectivity of metabolism, energy conservation, and macromolecular synthesis in a computable format. We are now using the model to explore the potential for biocommodity production from inexpensive sources.

The acetogen *Clostridium ljungdahlii* has emerged as a potential chassis for strain designed chemical production for not only can it grow heterotrophically on a diverse set of sugars, but it can also grow autotrophically on carbon monoxide (CO), carbon dioxide (CO_2) and hydrogen (H_2), or a mixture of all three gases (i.e. syngas). When grown autotrophically, *C. ljungdahlii* metabolizes the gases into multi-carbon organics, an ability that can be redirected and engineered to produce biocommodities from low cost substrates.

To advance towards this goal, a constraint-based modelling method was used to systematize the biochemical, genetic, and genomic knowledge of *C. ljungdahlii* into a computable mathematical framework. This metabolic and gene expression model (ME-model) accounts for 961 ORFs that are responsible for the production of transcriptional units, functional RNAs (e.g., tRNAs, rRNAs), prosthetic groups, and cofactors as well as the formation and translocation of protein complexes. This macromolecular synthesis machinery (i.e. the E-matrix) enables the metabolic network (M-model). The two networks integrated together compute the molecular constitution of *C. ljungdahlii* as a function of genetic and environmental parameters. Furthermore, the ME-model predicts relative growth conditions that are conducive for secretion of products like acetate, ethanol, and more. For example, comparison to *in vivo* data allows us to hypothesize that batch grown *C. ljungdahlii* is capable of higher secretion rates if certain shifts in proteomic expression occur. With this ME-model, we have a foundation for predicting and understanding the phenotype of *C. ljungdahlii*, which is vital for effective strain design.

Supported by DOE-DE-SC0012586, Next-Gen[3]: Sequencing, Modeling and Advanced Biofuels

Omics-Driven Analysis of *Clostridium ljundhalii* Disentangles the Complexity of Energy Conservation Pathways at the Molecular Level

Al-Bassam, M.M. (malbassam@ucsd.edu)[1*], Kim, J.[1], Liu, J.K.[1] and **Zengler, K**[1]

[1]University of California, San Diego, 9500 Gilman Drive 0419 La Jolla, CA 92093

Project goals: We aim to functionally annotate the *C. ljungdahlii* genome at ultra high resolution by integrating information from gene transcription, gene translation, transcription start sites and RNA polymerase binding sites. Furthermore, we reveal the translation efficiency in this model acetogen. We utilize the resulting knowledge to reconstruct a ME-model (*m*etabolic and gene-*e*xpression model). Finally, the optimized ME-model will guide strain design for production of valuable biocommodities and biofuels.

Acetogenesis is an integral part of the global carbon cycle and is driven by anaerobic bacteria that can fix CO_2 into acetyl-CoA and produce acetate as a by-product. *Clostridium ljungdahlii* is model homoacetogen of special interest due to its potential for production of fuels and other biocommodities from low-cost syngas (a mixture of CO_2, CO, and H_2). We developed novel omics techniques, including TSS-seq and Ribo-seq, to disentangle complex gene expression pattern at the levels of transcription and translation during growth under heterotrophic or autotrophic conditions.

We show that differentially translated genes in heterotrophic and autotrophic conditions are mainly involved in energy conservation and amino acid biosynthesis and import. Further, we show that some of these pathways are vastly regulated at the translational level. The amino acid biosynthetic pathways are differentially translated during heterotrophic growth, whereas during autotrophic growth these pathways are turned off and the oligopeptide permease is induced instead as a cost effective, but less effective source of amino acids.

Our omics-driven approach outlines the complex global regulation and gene architecture in *C. ljungdahlii* and provides valuable knowledge required for modeling, design and engineering of strains that are superior in renewable biofuels production.

Funded by DOE Grant No. DE-SC0012586 [Next-Gen3: Sequencing, Modeling and Advanced Biofuels]

Limited Forward Thermodynamic Driving Force in Glycolysis of Slow-Growing Bacteria

Junyoung O. Park[1,2]* (jopark@princeton.edu), Daniel Amador-Noguez,[1] Monica H. Wei,[3] and **Joshua D. Rabinowitz**[1,3]

[1]Lewis-Sigler Institute for Integrative Genomics; [2]Department of Chemical and Biological Engineering; and [3]Department of Chemistry, Princeton University, Princeton, New Jersey 08544

Project Goals: Rapid glycolysis during slow growth is a desirable feature for industrial biofuel production. In practice, however, glycolysis tends to slow down together with growth. Here we aim to understand glycolytic thermodynamics and regulation in fast and slow growing bacteria. Specifically, we aim to *i*) identify Gibbs free energy of glycolytic reactions using a combination of ^{2}H- and ^{13}C-labeling, and *ii*) identify flux control mechanisms by which cells may upregulate glycolysis.

C. cellulolyticum is an obligate anaerobe capable of converting cellulose into biofuels. Using LC-MS metabolomics, we identified its glycolytic pathway that involves unique cofactor (pyrophosphate and GTP) usage, akin to a recent finding in *C. thermocellum*. This results in an energy-efficient sugar catabolism that generates more usable energy in the form of high-energy phosphate bond than canonical glycolytic pathways but at the cost of forward driving force. Using ^{13}C-tracers, we found that its entire glycolysis is reversible.

The fully reversible glycolysis in *C. cellulolyticum* contrasts the canonical glycolysis as in *E. coli* where phosphofructokinase and pyruvate kinase have classically been assumed to be strongly forward-driven. We have recently quantified ΔG in *E. coli* central carbon metabolism by integrating absolute metabolite concentrations and ^{13}C- tracer data to probe forward and reverse fluxes. We observed relatively even distributions of ΔG across glycolysis that reflect sufficient driving force for almost every reaction, such that forward flux is substantially greater than reverse flux and therefore most enzyme is productively utilized. Here we present improved tools for ΔG measurement by integrating also ^{2}H-tracers. Using these tools, we find that glycolysis in *E. coli* growing slowly due to nitrogen limitation is closer to equilibrium than in nutrient-rich conditions. Because net flux through near-equilibrium reactions can change dramatically with small changes in substrate or product, this allows rapid increase in glycolytic flux upon nitrogen upshift with only slight changes in intermediate levels.

Collectively, these observations suggest that slow growing bacteria may engage in glycolysis with limited forward thermodynamic driving force, either to produce more usable energy per carbon in organisms chronically adapted to anaerobic low-sugar environments, or to facilitate seamless adaptation to changing nutrient availability.

References

1. Zhou, J., Olson, D. G., Argyros, D. A., Deng, Y., van Gulik, W. M., van Dijken, J. P., Lynd, L. R. "Atypical glycolysis in *Clostridium thermocellum*" 2013 Appl. Environ. Microbiol.

2. Park, J. O., Rubin, S. A., Xu, Y., Amador-Noguez, D., Fan, J., Shlomi, T., and Rabinowitz, J. D. "Metabolite concentrations and free energies that favor efficient enzyme usage are a conserved feature of cellular metabolism" (accepted) Nat. Chem. Biol.

3. Park, J. O., Amador-Noguez, D., Tanner, L. B., Wei, M. H., Li, H., and Rabinowitz, J. D. "Rapid modulation of glycolytic flux is accomplished by thermodynamic shifts" (to be submitted)

Funding: DOE DE-SC0012461

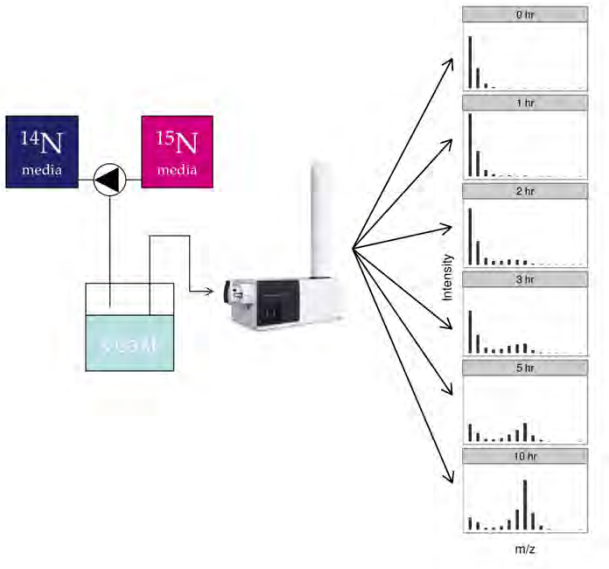
Condition-Dependent Regulation of Protein Turnover Rate in Yeast

Jonathan Goya[1*] (jgoya@princeton.edu), Sean Hackett[1], David Botstein[1], **Joshua D. Rabinowitz**[1,2]

[1]Lewis Sigler Institute for Integrative Genomics, Princeton University, Princeton NJ.
[2] Department of Chemistry, Princeton University, Princeton NJ.
http://growthrate.princeton.edu

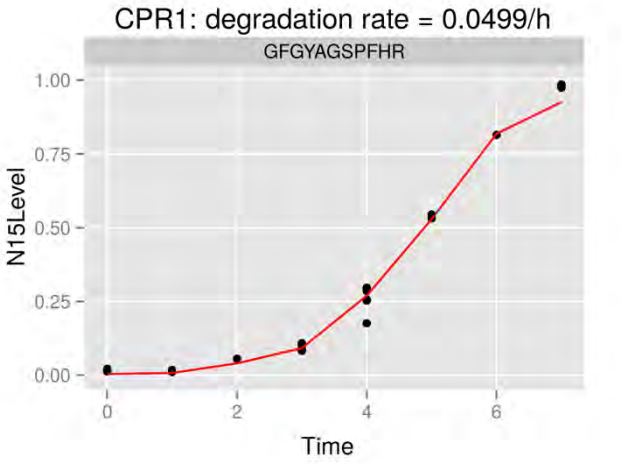
Project Goals: Despite the long history of targeted efforts to quantify protein turnover, measuring protein degradation rates in an unperturbed system at the proteome scale remains challenging. Here we aim to measure the degradation rates of a large number of proteins in nutrient-limited yeast continuous cultures (chemostats), with a focus on how degradation rates change across nutrient conditions. To this end, we switch yeast chemostat cultures into ^{15}N-labeled media and measure the dynamics of resulting protein labeling by LC-MS/MS. Through this approach, we aim to understand the extent to which protein degradation explains mismatches between protein and transcript abundance across nutrient conditions. Furthermore, we aim to uncover mechanisms controlling protein stability, which will be relevant to obtaining quantitative control of enzyme levels for optimized metabolic engineering.

Chemostat cultures provide a highly reproducible system for exploring natural variation in yeast physiology. We have previously measured transcripts[1], proteins, and metabolites[2] across 25 chemostats conditions, consisting of five different limiting nutrients each studied across five growth rates. An advantage of the chemostat setup is that biological variability across the 25 conditions is large, minimizing the impact of experimental measurement error. We were therefore surprised to observe that a median of only 19% of variation in protein abundance was explained by transcript abundance. Moreover, given that we are using relative measurements across different steady-state conditions, the discrepancy between protein and transcript levels cannot be explained by technical or biological issues impacting absolute protein and transcript abundances (e.g. MS response factors, ribosome binding strength of different mRNAs), nor differences in transcript and protein dynamics.



More careful examination of the transcript and protein data revealed stronger correlation between protein and transcript levels across growth rates for a given limiting nutrient (medians 37% – 53%), but these correlations

did not hold across limiting nutrient at specific growth rates (medians 19% - 31%). In some cases, most frequently in nitrogen limitation, we observed overt and statistically significant anti-correlation between protein and transcript abundances. This motivated us to measure protein stabilities, to see whether variability in protein degradation underlies the poor overall correlation between protein and transcript levels, and the particularly surprising cases of overt anti-correlation.



To this end, we developed an integrated experimental-computational method that uses dynamic ^{15}N-labeling to measure protein turnover in otherwise unperturbed, steady-state chemostats. We inoculate chemostats with S. cerevisiae and allow each chemostat population to stabilize. We then switch the incoming media from the natural isotope media to media that has been prepared with ^{15}N-ammonium sulfate, but is otherwise identical to the unenriched media. Switching chemostat media in this way results in an exponential decay in the amount of unlabeled nitrogen available for amino acid synthesis, so newly synthesized protein is increasingly enriched for the heavy nitrogen isotope. We sample the chemostat repeatedly over several generations after the media switch and measure the relative proportion of ^{15}N in each protein using LC-MS/MS. Using simultaneous measurements of ^{15}N enrichment in amino acids and a dynamical model of amino acid incorporation into protein, we will find the protein-specific degradation rate for each protein in each combination of nutrient condition and growth rate. We are currently in the process of completing the LC-MS/MS analysis for 9 chemostat conditions (carbon, nitrogen, and phosphorus limitation, each at 3 growth rates). We will present data on the transcript-protein correlations across the 25 chemostats conditions, as well as preliminary protein degradation results from these 9 conditions.

References

1. Matthew J Brauer, Curtis Huttenhower, Edoardo M Airoldi, Rachel Rosenstein, John C Matese, David Gresham, Viktor M Boer, Olga G Troyanskaya, and David Botstein. Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. Molecular biology of the cell, 19(1):352–67, January 2008.
2. Viktor M Boer, Christopher A Crutchfield, Patrick H Bradley, David Botstein, and Joshua D Rabinowitz. Growth-limiting intracellular metabolites in yeast growing under diverse nutrient limitations. Molecular biology of the cell, 21(1):198–211, January 2010.

Funding for this project has been provided by Department of Energy grant DE-SC0012461 and a Thought Leader Award from Agilent Technologies.

Ensemble cell-wide kinetic modeling of anaerobic organisms to support fuels and chemicals production

Satyakam Dash[1*](sud25@psu.edu), Ali Khodayari[1*](auk241@psu.edu), Evert K. Holwerda[2], Daniel G. Olson[2], M. Ahsanul Islam[3], Matthew Theisen[4], James C. Liao[4], Gregory Stephanopoulos[3], Lee R. Lynd[2], and **Costas D. Maranas**[1]

[1]The Pennsylvania State University; [2]Thayer School of Engineering at Dartmouth College; [3]Massachusetts Institute of Technology; [4]University of California, Los Angeles.

http://maranas.che.psu.edu/
https://engineering.dartmouth.edu/lyndlab/
http://bamel.scripts.mit.edu/gns/
http://www.seas.ucla.edu/~liaoj/

Project Goals: The goal of the project is to systematically construct dynamic models of *Clostridium thermocellum* and *Clostridium ljungdahlii* by making use of Ensemble Modeling (EM) paradigm through integration of multiple omic information (metabolomic & fluxomic). These models will be instrumental in exploring genetic interventions for overproduction of biofuel products.

Clostridia have garnered the interest of the bioprocess industry due to their unique metabolism. In particular, *C. thermocellum* can metabolize cellulose into biofuels such as ethanol and *C. ljungdahlii* can metabolize syn gas using the unique Wood-Ljungdahl pathway. Despite their increasing role as bio-production platforms, they remain poorly characterized with significant uncertainty in their metabolic repertoire. To this end, we report progress toward developing cell-wide dynamic metabolic model for both organisms using the Ensemble Modeling (EM) paradigm [1] built on a curated genome-scale metabolic (GSM) model foundation. The existing GSM model of *C. thermocellum* was curated to update cofactors based on experimental observations [2] and subsequently used to construct a second-generation GSM model (*i*Cth446). The model contains 446 genes and includes 598 metabolites and 637 reactions, along with gene-protein-reaction associations. The GSM model predicted higher yield of ethanol production for malate knockout (Δ*mdh* Δ*me*) mutant [3] and showed that lactate dehydrogenase knockout (Δ*ldh*) mutant did not have an effect on growth rate as observed experimentally [4]. The GSM model was next used to construct a core kinetic metabolic model of the *C. thermocellum*'s central metabolism containing 90 reactions and 76 metabolites with cellobiose as the carbon source under anaerobic growth condition. This core kinetic model accounts for all major biomass precursors. It encompasses the cellobiose degradation pathway, glycolysis/gluconeogenesis, the pentose phosphate (PP) pathway, the TCA cycle, pyruvate metabolism, anaplerotic reactions, alternative carbon metabolism, nucleotide salvage pathway, and 22 substrate level regulatory interactions extracted from BRENDA. Model parameterization was carried out using the EM approach by simultaneously imposing the mutant library data recently measured and provided by the Lynd group. This dataset includes *C. thermocellum* variants with mutations in lactate, malate, acetate, and hydrogen production pathways and combinations thereof defining 22 specific mutants with measured concentrations of various fermentation products such as acetate, lactate, formate, ethanol, hydrogen, carbon dioxide, amino acids and cellobiose (32 measured concentrations per mutant) (see Figure).

The preliminary analysis of the constructed kinetic model showed accurate predictions of cytosolic concentrations in multiple mutant strains, particularly for the pathways with measured data and substrate level regulations during parameterization. For example, the increase in metabolite concentrations (*i.e.*, 1.5 fold) of cellobiose and sugar phosphates by up-regulating amino-acid synthesis pathways under ethanol stress conditions [3] was accurately captured by the model. The up-regulation of amino acid synthesis led to an increase in ammonia uptake, which inhibits the phosphofructokinase reaction and results in fructose phosphate accumulation. Similarly, the fold changes in cytosolic concentrations for thirteen metabolites in the Δ*ldh* mutant were predicted accurately by the kinetic model [4]. Nevertheless, a leave-one-out cross-validation test to evaluate the robustness of the estimated parameters revealed that the fidelity of the model predictions remained limited to the mutants located in the proximity of mutations used for training

the model. Thus, integration of additional measured datasets under a variety of perturbation scenarios is required to address this challenge, while the effect of regulatory events must also be properly accounted for to make meaningful predictions. To this end, we are currently expanding the range of the training data by including thirteen additional fermentation datasets under different growth (i.e., media) conditions from the mutant library. At the same time, we use a data-driven approach to account for the action of all known regulators upon change in growth condition by adjusting the level of the enzymes under each condition. In addition, relative mRNA concentrations will be integrated into model to narrow down the range of enzyme levels that will vary during parameter estimation. RNA-seq data will be generated by the Liao group for the wild-type organism, as well as for mutant strains that will be constructed by modulating the activity of selected transcriptional regulators (i.e., increasing and/or decreasing promoter strength). The constructed model will also be analyzed by the Liao group using the ensemble modeling robustness analysis (EMRA) to identify the unrobust enzymes and the metabolic conditions which renders them unstable along with potential metabolic targets for iso-butanol overproduction. A similar effort is also underway for *C. ljungdahlii*. The existing GSM was already used as the scaffold for constructing the core metabolic model, composed of 77 reactions and 63 metabolites (see Figure). The network spans glycolysis, the Wood-Ljungdahl pathway and an incomplete TCA cycle with 41 substrate level regulatory interactions from BRENDA based on other Clostridia species. The Stephanopoulos group is applying ^{13}C isotope-assisted labeling experiments and metabolic flux analysis on *C. ljungdahlii* to study its central carbon metabolism and the Wood-Ljungdahl pathway (WLP), degeneracy in its TCA cycle, and the C-N assimilation and influences on the WLP. The generated data will subsequently be used to estimate the core model parameters using the same proposed framework.

The work was supported by the genomic science grant from Department of Energy, USA (grant# DE-SC0012377).

1. Khodayari, A., et al., *A kinetic model of Escherichia coli core metabolism satisfying multiple sets of mutant flux data.* Metabolic engineering, 2014. **25**: p. 50-62.
2. Zhou, J., et al., *Atypical glycolysis in Clostridium thermocellum.* Applied and environmental microbiology, 2013. **79**(9): p. 3000-8.
3. Yang, S.H., et al., *Clostridium thermocellum ATCC27405 transcriptomic, metabolomic and proteomic profiles after ethanol stress.* Bmc Genomics, 2012. **13**.
4. van der Veen, D., et al., *Characterization of Clostridium thermocellum strains with disrupted fermentation end-product pathways.* Journal of industrial microbiology & biotechnology, 2013. **40**(7): p. 725-734.
5. Nagarajan, H., et al., *Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of Clostridium ljungdahlii.* Microbial cell factories, 2013. **12**: p. 118.

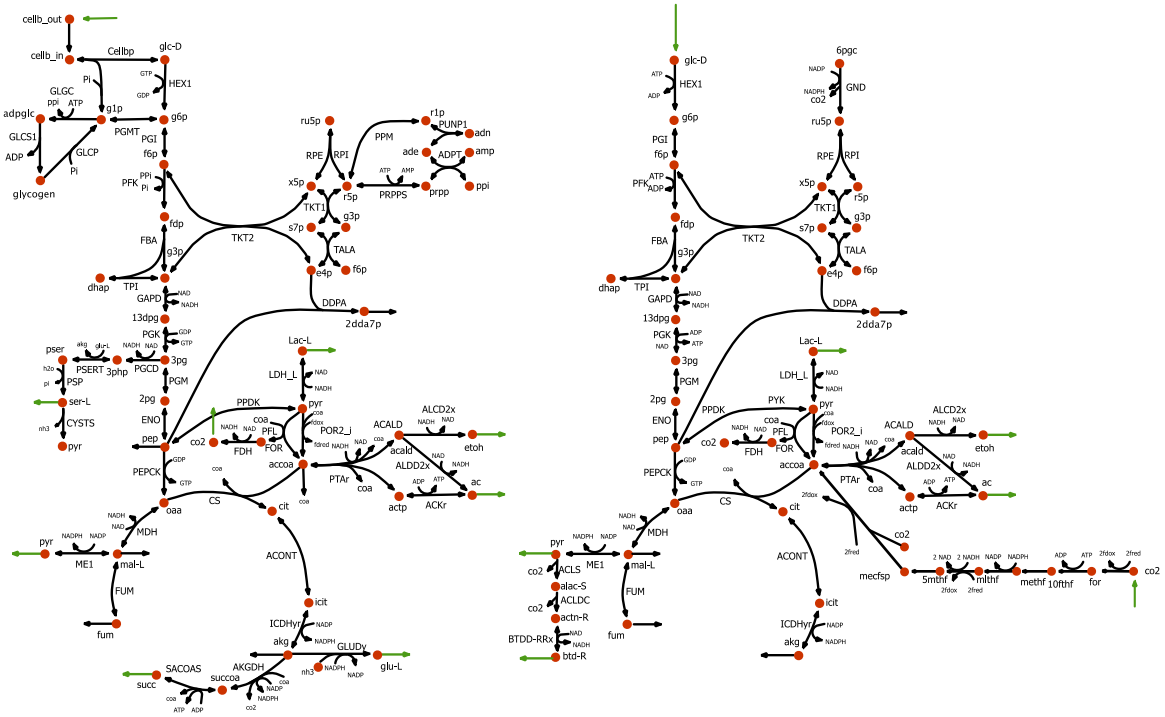


Figure 1: Core metabolic map of *Clostridium thermocellum* and *Clostridium ljungdahlii*. The green arrows represent the metabolite concentrations which are experimentally measured.

Development of a genome-scale *Escherichia coli* kinetic metabolic model satisfying flux data for multiple mutant strains

Ali Khodayari (auk241@psu.edu) and **Costas D. Maranas**

Department of Chemical Engineering, Pennsylvania State University, University Park, Pennsylvania, USA.

http://www.maranasgroup.com

Project Goals: The goal of this effort is to construct a genome-scale kinetic model of *Escherichia* coli metabolism by making use of Ensemble Modeling (EM) concepts. Model parameterization is carried out using multiple flux datasets for different substrates and growth (aerobic vs. anaerobic) conditions.

Kinetic modeling paradigm provides a promising platform to broaden our knowledge of cellular capacity and cell physiology beyond stoichiometric descriptions. However, developing kinetic models of metabolism at a genome-scale that faithfully recapitulate the effect of multiple genetic interventions is still an ongoing challenge. To this end, we introduce k-ecoli457, a genome-scale kinetic model of *Escherichia coli* metabolism that satisfies fluxomic data for a wild-type and 25 mutant strains for different substrates and growth (aerobic vs. anaerobic) conditions. The k-ecoli457 model contains 457 reactions and 337 metabolites accounting for all relevant reactions from the genome-scale *i*AF1260 model that carry flux under the experimental conditions of the flux measurements. These include reactions in glycolysis/gluconeogenesis, the Pentose Phosphate (PP) pathway, the TriCarboxylic Acid (TCA) cycle, anaplerotic reactions, amino acid synthesis/degradation, fatty acid oxidation/synthesis and a number of reactions in other parts of the metabolism, such as folate metabolism, cofactor and prosthetic group synthesis, alternative carbon, membrane lipid, cell envelope, nucleotide salvage and oxidative phosphorylation pathways. In addition, 295 regulatory interactions were extracted from BRENDA and EcoCyc and included in k-ecoli457. The model was also supplemented with a simplified version of the biomass equation including all the constituent precursors. Model predictions were tested against multiple experimentally measured datasets that were not used during model parameterization. These included (i) 898 steady-state metabolite concentrations for twenty of the mutant strains [1-4], (ii) 319 Michaelis-Menten constants (211 K_m and 108 k_{cat} values) from BRENDA and EcoCyc, and (iii) 320 experimentally reported product yields for designed strains spanning 24 different bioproducts. Comparisons revealed that 63% of the predicted metabolite concentrations as well as 60% and 64% of the estimated K_m and k_{cat} values, respectively, are within the experimentally reported ranges. These levels of agreement, in overall, exceed the previous effort [5], despite the significantly increased scope of the model and coverage of less studied pathways. The average relative error of k-ecoli457 predictions for the yield of 16 bioproducts in 140 designed strains is 0.1, while stoichiometric model based techniques such as flux balance analysis (FBA) or minimization of metabolic adjustment (MOMA) yield corresponding relative errors of 1.05 and 1.19, respectively.

This modeling effort describes a stepwise procedure for construction of genome-scale kinetic models with robust parameterization consistent with multiple sets of omics information for *E. coli* and provides guidelines for developing genome-scale kinetic models for other well-studied organisms.

The work was supported by the genomic science grant from Department of Energy, USA (grant# DE-SC0012377).

1. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A *et al*: **Multiple high-throughput analyses monitor the response of E. coli to perturbations**. *Science* 2007, **316**(5824):593-597.

2. Kabir MM, Ho PY, Shimizu K: **Effect of ldhA gene deletion on the metabolism of Escherichia coli based on gene expression, enzyme activities, intracellular metabolite concentrations, and metabolic flux distribution**. *Biochem Eng J* 2005, **26**(1):1-11.
3. Zhao J, Baba T, Mori H, Shimizu K: **Global metabolic response of Escherichia coli to gnd or zwf gene-knockout, based on 13C-labeling experiments and the measurement of enzyme activities**. *Applied microbiology and biotechnology* 2004, **64**(1):91-98.
4. Zhao J, Shimizu K: **Metabolic flux analysis of Escherichia coli K12 grown on 13C-labeled acetate and glucose using GC-MS and powerful flux calculation method**. *Journal of biotechnology* 2003, **101**(2):101-117.
5. Khodayari A, Zomorrodi AR, Liao JC, Maranas CD: **A kinetic model of Escherichia coli core metabolism satisfying multiple sets of mutant flux data**. *Metabolic engineering* 2014, **25**:50-62.

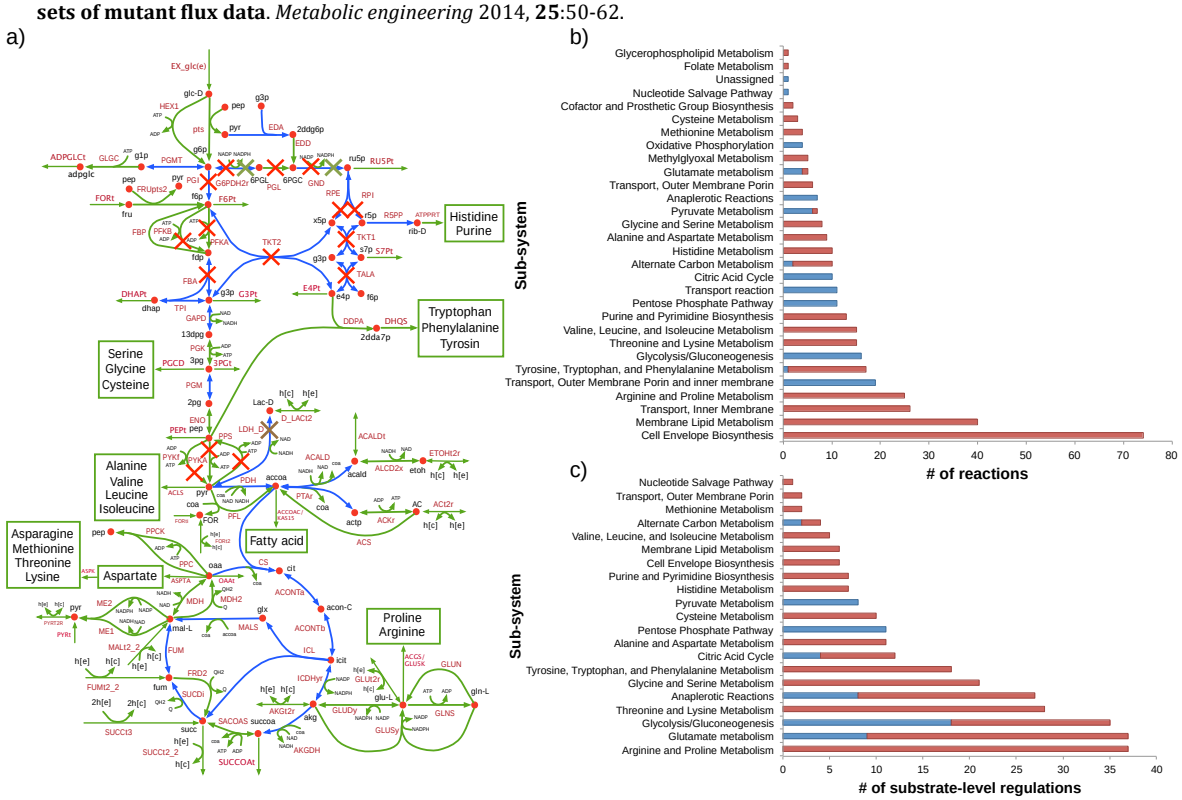


Figure 1 (a) A pictorial representation of the constructed kinetic model of *E. coli* metabolism. Red, brown and green marks denote the knockout mutants grown aerobically on glucose, anaerobically on glucose and aerobically on pyruvate, respectively, with flux data for the reactions shown in blue. (b) Sub-system classification of reactions in the constructed kinetic model. (c) Sub-system classification of the integrated regulatory interactions. Blue bars denote the content of the core model [5] while red denotes the additional reactions/regulations included in k-ecoli457.

Magnesium controls carbon flux to biomass or fermentation products.

David Christensen,[1*] (dchristensen1@luc.edu), James Orr,[2] Christopher V. Rao,[2] and **Alan Wolfe**[1]

[1]Loyola University, Chicago, IL; [2]University of Illinois at Urbana-Champaign

Project Goals: The goal of this project is to determine how protein acetylation affects metabolism in engineered microorganisms. Lysine acetylation is a common post-translational modification that eukaryotes, archaea, and bacteria employ to regulate protein activity. Multiple studies have recently shown that lysine acetylation predominantly targets metabolic enzymes – in fact, most metabolic enzymes are subject to lysine acetylation. We hypothesize that bacteria employ lysine acetylation as a global mechanism to regulate metabolism in response to their energy and redox status. Our previous work suggests that lysine acetylation may be an attractive and innovative target for metabolic engineering. We are investigating how lysine acetylation affects fuel production in engineered microorganisms. The significance of this work is that it will address a fundamental gap in our understanding of bacterial metabolism and identify new approaches for overcoming the problems associated with the production of advanced biofuels.

N^{ε}-lysine acetylation is a posttranslational modification that occurs within all three domains of life. The acetylation reaction occurs through the donation of an acetyl group from a donor molecule onto a susceptible lysine of a protein or peptide. This modification neutralizes the positive charge of the lysine side chain and increases its size. Acetylation of residues required for catalytic function can render an enzyme inactive. Additionally, neutralization of the positive charge can disrupt salt bridges necessary for protein-protein interactions. In *E. coli*, acetylation is known to be catalyzed by two mechanisms. One, the canonical enzymatic mechanism, utilizes the only known lysine acetyltransferase, YfiQ, to catalyze the donation of the acetyl group of an acCoA molecule onto a lysine. The other and more predominant mechanism employs acetyl phosphate (acP), the intermediate of the acetate fermentation (AckA-Pta) pathway, to donate its acetyl group onto proteins non-enzymatically. Therefore, conditions that promote acetate fermentation invariably lead to protein acetylation.

Previously, our lab and others have found that *E. coli* grown in carbon excess leads to high acetylation levels due to the production of acetate[1,2]. In this work, we report that carbon is directed into biomass rather than acetate when magnesium, the limiting nutrient in our experiments, is in excess. We found that *E. coli* grown in tryptone broth buffered to pH7 (TB7) supplemented with 0.4% glucose grew to an OD_{600} value of ~3. When we further supplemented the growth medium with 1 mM magnesium sulfate ($MgSO_4$), *E. coli* had an extended exponential phase and reached an OD_{600} that was 2-3 times greater. Six additional carbon sources showed the same magnesium-induced biomass increase. We also found that cells grown in media where

tryptone was replaced with either casamino acids or peptone exhibited the same effect. Even the common laboratory medium LB was found to benefit from magnesium supplementation when provided with excess carbon. More significantly, magnesium-induced biomass increase was accompanied by a significant reduction in acetylation as measured by Western blot analysis.

These results demonstrate that magnesium indirectly affects protein acetylation by determining whether carbon flux is diverted to biomass or acetate. The mechanism is likely related to ribosome abundance, because magnesium is known to increase the stability of the ribosomes[3]. We are currently testing this hypothesis. Collectively, these results provide a deeper understanding of how different media formulations influence bacterial metabolism and physiology, and demonstrate how *E. coli* regulates its metabolism accordingly.

References

1. Weinert, B.T., et al., *Acetyl-phosphate is a critical determinant of lysine acetylation in E. coli.* Mol Cell, 2013. **51**(2): p. 265-72.
2. Kuhn, M.L., et al., Structural, kinetic and proteomic characterization of acetyl phosphate-dependent bacterial protein acetylation. PLoS One, 2014. 9(4): p. e94816.
3. Pontes, M.H., A. Sevostyanova, and E.A. Groisman, When Too Much ATP Is Bad for Protein Synthesis. J Mol Biol, 2015. 427(16): p. 2586-94.

This work was supported by the U.S Dept. of Energy, Office of Science, Office of Biological & Environmental Research (PI: Rao, DE-SC0012443), grant title 'The Systems Biology of Protein Acetylation in Fuel-Producing Microorganisms' (ER213630).

Regulation of sugar consumption in *Escherichia coli* by amino acids

James S Orr[1]* (jsorr3@illinois.edu), David Christensen[2], Alan J Wolfe,[2] and **Christopher V Rao**[1]

[1]University of Illinois at Urbana-Champaign; [2]Loyola University Chicago

Project Goals: The goal of this project is to determine how protein acetylation affects metabolism in engineered microorganisms. Lysine acetylation is a common post-translational modification that eukaryotes, archaea, and bacteria employ to regulate protein activity. Multiple studies have recently shown that lysine acetylation predominantly targets metabolic enzymes – in fact, most metabolic enzymes are subject to lysine acetylation. We hypothesize that bacteria employ lysine acetylation as a global mechanism to regulate metabolism in response to their energy and redox status. Our previous work suggests that lysine acetylation may be an attractive and innovative target for metabolic engineering. We are investigating how lysine acetylation affects fuel production in engineered microorganisms. The significance of this work is that it will address a fundamental gap in our understanding of bacterial metabolism and identify new approaches for overcoming the problems associated with the production of advanced biofuels.

When we measure protein acetylation in *E. coli*, we routinely grow the cells in tryptone broth (pH 7) supplemented with glucose prior to analysis by liquid chromatography. During the course of these experiments, we observed that consumption of glucose is delayed. In particular, we found that the cells reached an OD_{600} of ~1 before they started to consume the glucose. Subsequent mass spectrometry analysis demonstrated that *E. coli* consumes multiple amino acids (serine, aspartate, and threonine) before it begins to consume glucose. Similar results were also observed with lactose, arabinose, and glycerol, where again sugar consumption is delayed by amino acids.

The mechanism is independent of the phosphotransferase system and appears to result from the switch between glycolytic and gluconeogenic growth. All of the preferred amino acids (serine, aspartate, and threonine) enter metabolism through pyruvate, leading to gluconeogenic growth. This mode of growth appears to inhibit the metabolism of glucose (and other sugars). In support of this mechanism, we found that pyruvate also inhibits the uptake of glucose.

In conclusion, we have serendipitously identified a new fact of *E. coli* physiology that may translate to other species of bacteria. The results are significant as glucose is normally thought of as the preferred carbon source for *E. coli*. Our results, however, demonstrate that easily consumed amino acids are the preferred carbon source. Furthermore, they demonstrate that metabolic regulation in *E. coli* is more complex than previously thought.

This work was supported by the U.S Dept. of Energy, Office of Science, Office of Biological & Environmental Research (PI: Rao, DE-SC0012443), grant title 'The Systems Biology of Protein Acetylation in Fuel-Producing Microorganisms' (ER213630).

Dynamic Remodeling of Protein Acetylation in Fuel-Producing *E. coli* with Different Carbon Sources

Birgit Schilling,[1] David Christensen,[2] Kori D. Dunn,[3] Dylan J. Sorensen,[1] Alexandria K. Sahu,[1] **Christopher V. Rao**,[2] **Alan J. Wolfe**,[3] and **Bradford W. Gibson[1]*** (bgibson@buckinstitute.org)

[1] Buck Institute for Research on Aging, Novato, CA
[2] Loyola University Chicago, Maywood, IL
[3] University of Illinois at Urbana-Champaign, Urbana IL

Project Goals: The goal of this project is to determine how protein acetylation affects metabolism in engineered microorganisms. Lysine acetylation is a common post-translational modification that eukaryotes, archaea, and bacteria employ to regulate protein activity. Multiple studies have recently shown that lysine acetylation predominantly targets metabolic enzymes – in fact, most metabolic enzymes are subject to lysine acetylation. We hypothesize that bacteria employ lysine acetylation as a global mechanism to regulate metabolism in response to their energy and redox status. Our previous work suggests that lysine acetylation may be an attractive and innovative target for metabolic engineering. We are investigating how lysine acetylation affects fuel production in engineered microorganisms. The significance of this work is that it will address a fundamental gap in our understanding of bacterial metabolism and identify new approaches for overcoming the problems associated with the production of advanced biofuels.

Recently, we published a study reporting that glucose-regulated lysine acetylation (Kac) was predominant in central metabolic pathways and overlapped with acetyl phosphate-regulated acetylation sites (1). We proposed that acetyl phosphate-dependent acetylation across hundreds of proteins is a response to carbon flux that can regulate central metabolism. Here, we extend our investigations to examine the effect on both protein expression levels and the alterations in acetylation status across the *E. coli* proteome when it is provided with alternative carbon sources beyond glucose.

For these experiments, *E. coli* K-12 wild-type strain BW25113 was grown in M9 media, a minimal growth media supplemented with either 0.4% or 4% glucose (Glc) or xylose (Xyl). Four independent biological replicates were performed for each of the four conditions and a total protein fraction was subjected to label-free quantitation both before and after immuno-affinity enrichment of peptides containing acetyllysine modifications. Anti-acetyllysine Western blot analysis revealed a proteome that became more acetylated with higher concentrations of glucose or xylose, respectively. To obtain a more precise understanding of glucose- and xylose-induced changes in protein acetylation, as well as in protein expression, we examined protein fractions that had undergone trypsin digestion both before and after antibody-based affinity-enrichment of acetyllysine-containing peptides by quantitative mass spectrometry. To monitor changes in protein acetylation and expression, we used a novel label-free and data-independent acquisition (DIA or SWATH) approach on a SCIEX TripleTOF 6600 LC/MS system that we have modified for this purpose (2). Overall, we confidently identified 3840 unique acetylation sites from 978 acetylated proteins across all growth conditions: 1608 Kac sites at 0.4% Xyl; 2824 Kac sites at 4% Xyl; 1489 Kac sites at 0.4% Glc, and 2949 Kac sites at 4% Glc. Both the number of identified sites and the extent of acetylation increased when the concentration of glucose or xylose was increased from 0.4% to 4%. Using our label-free quantitative approach with Skyline 3.5, an open source software project, we were able to identify hundreds of acetylation sites in over 150 proteins that showed robust and statistically significant acetylation increases when cells

were grown with the larger amount of sugar. These acetylated proteins were involved in glycolysis/gluconeogenesis, pyruvate metabolism, TCA cycle etc. In addition, performing a functional annotation analysis for these acetylated proteins showed a statistically relevant enrichment for the ontology term "generation of precursor and metabolites and energy". Interestingly, the acetylation levels and relative changes with increased sugar concentrations appear very similar between glucose and xylose. We also found, as expected, significant changes in the expression levels of proteins relevant in carbohydrate metabolism and sugar transport, as well as key metabolic proteins, including acetyl-coenzyme A synthetase (ACS), and 2-methylcitrate synthase. For example, ACS showed a ~10 fold reduction in protein levels at high concentrations (4%) of either glucose or xylose compared to those observed in the lower sugar (0.4%) supplementation experiments.

Lastly, we have begun an independent assessment of site occupancy or stoichiometry of lysine acetylation using a novel SWATH acquisition method. In this modified approach, ion intensities of both MS1 precursors and MS2 fragments are used to determine site occupancy, thus increasing the confidence and accuracy of these determinations. Such determinations are typically prone to error and over-estimation given the large dynamic range of observed acetylation stoichiometries (~20% to less than 0.01%). Our modified and optimized approach was based on a recently reported method by Baeza and colleagues (3) that depends on the accurate determination of isotopic forms that are generated both *in vivo* (normal endogenous isotopic abundances) and *in vitro*, the latter by quantitative per-acetylation with stable isotope labeled acetic anhydride at the protein level. Our revised method is based on measuring abundances of otherwise identical specific fragments ion pairs in a data-independent acquisition that encompass the same site of lysine acetylation and whose relative ion intensities represent the extent to which they originated from endogenous versus exogenous labeling. In these *E. coli* experiments, most of the lysine-acetylation stoichiometries measured were low (<5%). However, there were some 'hot spots' detected at specific sites showing higher site occupancy. These site occupancy levels reveal information in addition to the relative fold-change of acetylation levels of individual Kac sites under different sugar supplement conditions, and will provide additional mechanistic assessments of which pathways may be most functionally affected by acetylation.

References

1. B Schilling, D Christensen, R Davis, AK Sahu, LI Hu, A Walker-Peddakotla, DJ Sorensen, B Zemaitatis, BW Gibson, AJ Wolfe. Protein acetylation dynamics in response to carbon overflow in *Escherichia coli*, Molecular Microbiology, **2015**, 98(5), 847-63.

2. MJ Rardin, B Schilling, LY Cheng, BX MacLean, DJ Sorensen, AK Sahu, MJ MacCoss, O Vitek, BW Gibson. MS1 Peptide Ion Intensity Chromatograms in MS2 (SWATH) Data Independent Acquisitions. Improving Post Acquisition Analysis of Proteomic Experiments. Mol Cell Proteomics **2015**, 14(9), 2405-19

3. J Baeza, JA Dowell, MJ Smallegan, et al. Stoichiometry of site-specific lysine acetylation in an entire proteome. J Biol Chem **2014**, 289 (31), 21326-38.

Funding Statement:
This work was supported by the U.S Dept. of Energy, Office of Science, Office of Biological & Environmental Research (PI: Rao, DE-SC0012443), grant title 'The Systems Biology of Protein Acetylation in Fuel-Producing Microorganisms'. We also acknowledge the support of the NIH shared instrumentation program for the use of the SCIEX TripleTOF 6600 mass spectrometer (1S10 OD016281).

Robust Non-Oxidative Glycolysis in *Escherichia coli*

Paul P. Lin[1] (phlin@ucla.edu), Alec J. Jaeger,[1] Maria Chun,[1] Kouki M. Yoshino,[1] Sawako Konishi,[1] Sharon C. Xu,[1] and **James C. Liao[1]**

[1]Department of Chemical and Biomolecular Engineering, University of California, Los Angeles

Project Goals: This project seeks to construct strains of *Escherichia coli* to rely solely on Non-oxidative glycolysis (NOG) for glucose catabolism during aerobic growth, fundamentally rewiring the central metabolism to favor carbon conservation in acetyl-CoA biosynthesis. This work will also engineer the constructed strains for energy applications. Specifically, we aim to produce an advanced biofuel, n-butanol, with close to complete carbon conservation. Hydrogen or formate will be provided to supply reducing equivalents needed to drive the NADH dependent pathway.

Acetyl-coenzyme A (acetyl-CoA) is a two carbon metabolite and important metabolic precursor to a variety of industrially relevant compounds including biofuels. An ultimate limitation of acetyl-CoA derived biochemical production is the inherent carbon loss when forming acetyl-CoA. Most organisms use some glycolytic variation, commonly the Embden-Meyerhof Pathway (EMP), to initially degrade sugar into pyruvate. Pyruvate, a C3 metabolite, is then decarboxylated to form acetyl-CoA, losing carbon to the environment. This decarboxylation limits the carbon yield to only two molecules of acetyl-CoA from one molecule of hexose, thus inhibiting the economics of any associated bioprocess. A synthetic sugar catabolism pathway, termed non-oxidative glycolysis (NOG), was recently developed to address this problem, as it uses a combination of phosphoketolase dependent cleavage of sugar phosphates and a carbon rearrangement cycle to directly generate three C2 units per hexose in a redox neutral manner. To further expand the applications using NOG, an *Escherichia coli* strain was constructed to rely solely on NOG for sugar catabolism in this work. Therefore, the resulting strain offers significant potential to be engineered for the production of a variety of acetyl-CoA derived compounds. To implement NOG as a growth pathway, all native sugar degradation pathways, including the EMP, ED and methylglyoxal bypass, were removed, eliminating the cell's ability to grow on sugar as a sole carbon source. In addition, the glyxoylate shunt and gluconeogenesis pathways, which are necessary for the production of essential metabolites using NOG, were upregulated. Then, this engineered strain was evolved to grow in minimal glucose media supplemented with exogenous acetate. The overexpression of critical NOG protein phosphoketolase restored the cells ability to grow on sugar as a sole carbon source and further evolution improved the cell's growth phenotype in glucose and xylose minimal media. Under anaerobic conditions, it was verified that this strain produces acetate as a major fermentation product from glucose. The carbon rearrangement pathway will be further overexpressed to increase the carbon yield through NOG pathway for acetate production. To further demonstrate the potential of NOG, this strain will be additionally engineered to produce C2 derived biofuels by supplying reducing power in the form of hydrogen or formate.

Predicting Gene Targets for Optimizing Lipid Production in Acetic Acid Pathway of *Yarrowia lipolytica* by Ensemble Modeling

Po-Wei Chen,[1]* (i98512@gmail.com), Matthew Theisen,[1] and **Prof. James Liao**[1]

[1]University of California, Los Angeles

While the opportunity for optimization of Yarrowia lipolytica production of lipids from glucose is well-explored, an acetic acid metabolism based lipid production pathway is investigated by means of the Ensemble Modeling (EM) to predict potential gene targets for maximizing the lipid yield. To build a model for the acetic acid pathway, we have collected information from literatures that the major route for acetic acid metabolism is by the reaction acetyl-CoA synthetase, which converts acetic acid into acetyl-CoA powered by the conversion of ATP to AMP. We used this as the route for acetic acid metabolism in *Yarrowia* for the model. Additionally, we achieved acetyl-CoA transport into the mitochondria via the acyl-carnitine/carnitine translocase system. The glyoxylate cycle is presumed to be active in the mitochondria under acetate conditions, which allows for a net conversion of two acetyl-CoA molecules into a 4-carbon dicarboxylic acid (malate). Malate is then decarboxylated in the mitochondria by malic enzyme which has been noted to be localized there. A pyruvate carrier then moves pyruvate to the cytosol where it can be used for gluconeogenesis.

EM simulation builds large number of models (with randomly sampled but realistic parameter values). We then simulate effects for all single enzyme perturbation. The reference/steady state fluxes, which are the key information for EM simulation, are estimated using publicly available biomass and lipid yield data (this literature will be cited on the poster) and steady state is ensured using linear programming. Estimated results are within 4% of the reported value. In current stage, our EM simulation built 1000 stable models for gene targets prediction, and the perturbation analysis provides 6 knockout and overexpression gene targets for further investigation in the lab.

1. 6 knockout targets: Malate synthase, isocitrate lyase, succinate dehydrogenase, fumarase, malic enzyme NADH (mitochondria) and pyruvate carboxylase. Among them, knocking out malate synthase and isocitrate lyase can increase 5.7% of lipid yield compared to the reference.
2. 6 overexpression targets: Oxoglutarate dehydrogenase, acetyl-CoA carboxylase, isocitrate dehydrogenase, ATP transport, adenylate kinase and respiration. Among these, overexpressing oxoglutarate dehydrogenase increases 8.5% of lipid yield compared to the reference, while the second best, acetyl CoA carboxylase, increases 7%.

This model of *Yarrowia* metabolism under acetic acid feed conditions is a reasonable representation of the system and EM simulation has suggested key gene targets for maximizing lipid yield. Furthermore, since key assumptions about compartmentalization and metabolite flow have been identified, this will allow the model to be adapted to many other conditions, including high density production schemes with low biomass accumulation.

Conversion of lignin-derived aromatic compounds into lipids by engineered *Rhodococcus opacus* strains

Aki Yoneda[1,2], William R. Henson[3*] (williamrhenson@gmail.com), Nicholas K. Goldner[2], Kun Joo Park[3], Kevin J. Forsberg[2], Soo Ji Kim[3], Mitchell W. Pesesky[2], **Marcus Foston**[3], **Gautam Dantas**[1, 2, 4, 5] and **Tae Seok Moon**[3]

[1] Department of Pathology and Immunology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108, USA; [2] Center for Genome Sciences and Systems Biology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108, USA; [3] Department of Energy, Environmental & Chemical Engineering, Washington University in St. Louis, St. Louis, MO, 63130, USA; [4] Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO, 63130, USA; [5] Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108, USA

Project Goals: The goal of this project is to interrogate the metabolic networks and genetic regulation that control the utilization of and tolerance to thermochemically depolymerized lignin, focusing on phenolics, in *R. opacus*.

Lignocellulosic biomass is a renewable feedstock that can be converted into biofuels and high-value chemicals using microorganisms. However, pretreatment of lignocellulosic biomass often releases toxic compounds that inhibit microbial growth. In addition, lignin has been an untapped carbon source due to its recalcitrance and toxic monomeric units (e.g., aromatic compounds). Our work focuses on developing *Rhodococcus opacus* as a new chassis for conversion of aromatic compounds into triacylglycerols (TAGs), biodiesel precursors (1). *R. opacus* is a promising host due to its high tolerance to and utilization of aromatics as a sole carbon source. Importantly, *R. opacus* can accumulate TAGs up to ~80% of cell dry weight under nitrogen-limiting conditions. To enhance its innate aromatic-degrading capacity, we applied adaptive evolution, a growth-based strain selection method, by sequentially sub-culturing cells in diverse combinations of lignin-derived aromatic compounds as sole carbon sources. Our adapted strains demonstrated higher growth rates and higher lipid accumulation compared to the wild type strain. Whole genome sequencing, RNA-seq, and ^{13}C-fingerprinting analysis have identified possible aromatic tolerance and utilization mechanisms such as degradation pathway upregulation. We will present progress towards development of *R. opacus* as a microbial cell factory.

(1) WD Hollinshead, WR Henson, M Abernathy, TS Moon and YJ Tang. Rapid Metabolic Analysis of *Rhodococcus opacus* PD630 via parallel ^{13}C-Metabolite Fingerprinting, *Biotechnol. Bioeng.* 113, 91-100 (2016)

This work was supported by U.S. Department of Energy grant DE-SC0012705.

Lipid Production in Single Oleaginous Yeast Cells Using *In Vivo* Label-Free Imaging

Wei Shen[1], Chi-Han Chang[2]* (cchang@g.harvard.edu), Ang Li[2], Shi-You Ding[1] and **X. Sunney Xie**[2]

[1] Department of Plant Biology, Michigan State University, East Lansing, MI 48824;
[2] Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138

http://bernstein.harvard.edu/
http://dinglab.plantbiology.msu.edu/

Project Goal: To develop a system for *in vivo* label-free imaging followed by single-cell RNA-Seq for understanding the lipid production mechanisms in oleaginous yeast, and to associate genes and transcriptional factors with the lipid accumulation processes at single-cell resolution.

Hydrocarbons have great potential to be used as next generation biofuels due to their relatively high energy density and compatibility with current infrastructure of fuel usage. Lipids produced by oleaginous yeast are of particular interest, as they can accumulate at high levels in the form of triacylglycerols (TAGs). The lipid accumulation is finely controlled based on nitrogen availability, while the lipid composition may vary with the carbon sources such as glucose and xylose. We explore lipid production in oleaginous yeast under different nitrogen and sugar conditions, and focus on the critical conditions under which the yeast changes their lipid accumulation behavior dramatically. Hyperspectral Stimulated Raman Scattering (hsSRS) microscopy is used to image the lipid content *in vivo* without chemical labeling of any molecules. A microfluidic platform is also developed for high-throughput imaging and subsequent isolation of single cells for whole-transcriptome analysis. With these techniques for correlating growth conditions to lipid accumulation and transcriptomic profiles of single cells, we aim to elucidate the regulatory network of lipid accumulation and optimize the biofuel production from oleaginous yeast.

This work is funded by the U.S. Department of Energy (DOE), Office of Biological and EnvironmentalResearch, Genome Science Program, Award number DE-SC0012411.

Functional genomics of lipid accumulation in the oleaginous yeast *Rhodosporidium toruloides* using randomly bar-coded TDNA-Seq

Dominic Pinel [1,2,3], Samuel T. Coradetti [3,4], Gina Geiselman[3], Masakazu Ito[3], **A.P. Arkin** [1,2,3], **Jeffrey M. Skerker** [1,2,3*] (skerker@berkeley.edu) and **Rachel B. Brem**[3,4]

[1]Energy Biosciences Institute, Berkeley, CA; [2]Lawrence Berkeley National Laboratory, Berkeley, CA; [3]University of California, Berkeley, CA; and [4]Buck Institute for Research on Aging, Novato, CA

Project Goals: Conduct large-scale mapping of genotype to phenotype in oleaginous yeast, focused on the genes underlying lipid production, plant feedstock hydrolysate tolerance, low-oxygen metabolism, and co-utilization of sugars in plant material. The yeast chosen for this study, *Rhodosporidium toruloides,* has several advantages over more traditional model yeasts, including its native ability to metabolize the sugars in plant hydrolysates (i.e., glucose, xylose, arabinose, and cellobiose) and high *de novo* lipid productivity. Techniques for *R. toruloides* genetics and genome engineering will be developed and used to map the determinants of complex growth and lipid productivity traits in wild isolates and engineered strains. Establishment of a robust, versatile model yeast that natively accumulates high lipid levels will enable greater flexibility in developing new biofuels that produce industrial strains and will provide fundamental insights into the origins of complex traits useful for biofuels production.

The ability to rapidly functionally annotate genomes of non-model fungi is a barrier to developing and exploiting new host organisms for biotechnology and studying interesting fungal traits. We study the oleaginous basidiomycetous yeast *Rhodosporidium toruloides*, which can accumulate up to ~70% of its dry cell weight, primarily as triglycerides (TAG). These triglycerides can be extracted and converted into biodiesel or the high acetyl-CoA flux required for lipid accumulation could be harnessed for the production of fatty acid derived biofuels/chemicals. Furthermore, *R. toruloides* can also generate relatively high levels of carotenoid compounds, consume a wide array of carbon sources and tolerate common sources of inhibition found in biomass waste substrates. However, little is known about the *R. toruloides* genetics of these useful traits. Here, we have developed a functional genomics technology that is an extension of RB-TnSeq, called RB-TDNA-Seq that can be used in fungal systems. Instead of transposon mutagenesis, our method exploits the efficiency of *Agrobacterium tumefaciens* mediated transformation (ATMT) to randomly integrate a transfer DNA (TDNA) fragment into the fungal host genome. ATMT mutagenesis has been reported in hundreds of fungal species; therefore, our method should be widely applicable. We have performed proof of concept experiments demonstrating that we can quantify mutant strain fitness using a growth-based assay, and also have developed an assay based on buoyant density to identify genes required for lipid accumulation. To demonstrate our fitness-based assay, we used a ~50,000 member mutant pool to identify genes required for amino acid biosynthesis. By comparing the fitness profile in

minimal media with synthetic complete media, or media supplemented with single amino acids, we identified arginine, methionine, proline, and leucine auxotrophs. In another pilot experiment using a smaller mutant pool (~6,000 strains) we successfully segregated strains with altered lipid accumulation by differences in their buoyant density. Lipid production was induced by growth in nitrogen-limited conditions, and then cultures were segregated into high-density, medium-density, or low-density fraction by progressive centrifugation in increasing concentrations of sucrose. Over 900 mutant strains representing disruptions in over 700 genes were significantly enriched in the high-density or low-density fractions as compared to the input culture. These genes represented a wide range of cellular and biochemical functions including lipid metabolism, nutrient sensing, the cellular endomembrane system, transporters and many uncharacterized proteins. Ten genes with multiple insertions that had consistent enrichment patterns, or for which a single mutant strain was strongly enriched, were then deleted by homologous recombination using a drug resistance marker in a Ku70 deletion strain background. Preliminary experiments comparing TAG accumulation in these strains (as measured by BODIPY 495/503 intensity) to the Ku70 background confirmed that four strains had unambiguously altered lipid accumulation, three modest but significant lipid phenotypes, and three had unaltered lipid accumulation or phenotypes too subtle to measure with this protocol. Among the genes with strong lipid phenotypes were homologs to the human genes seipin and tuberin, shown to be involved in lipid metabolism in several other model systems. We have now scaled up our ATMT mutagenesis protocol and generated a randomly integrated, bar-coded mutant pool of *R. toruloides* consisting of ~400,000 members using this technology. This large mutant pool and the methods we developed will now enable us to perform genome-wide fitness and lipid accumulation screens.

This work is supported by grant DE-SC0012527 from the U.S. Department of Energy's (DOE) Genomic Science program, managed within the Office of Biological and Environmental Research (BER).

Accelerating the engineering of improved lipid accumulation in *Yarrowia lipolytica*

Kyle R. Pomraning, Erin L. Bredeweg, Tom O. Metz, Young-Mo Kim, Heino Heyman, **Scott E. Baker* (scott.baker@pnnl.gov)**

Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA

Project Goal: Our goal is to enable rapid engineering of the oleaginous yeast, *Yarrowia lipolytica,* for increased rates of biofuel production.

The oleaginous ascomycete yeast *Yarrowia lipolytica* serves as a model organism for lipid accumulation and the production of lipid based biofuels. In order to expand capabilities of *Y. lipolytica* for both biological research and industrial bioengineering applications, we have developed a set of genetic and molecular tools. In this work, strains were developed with increased homologous recombination, for targeted DNA incorporation. We have generated a library of fluorescent protein tagged strains to define organelle morphology in live cells ("The *Yarrowia* Organelle Atlas") and built a set of plasmids to allow targeted overexpression and a *Yarrowia*-optimized GFP for fluorescent tagging. We have performed genome sequencing and assembly, and RNA expression analysis and transcript discovery in the PO1g background, which differs from the CLIB122 reference strain. This work provides the *Yarrowia* community with tools for the study of cell biology and metabolism of *Y. lipolytica* to further development for biofuels and natural products. In addition, we identified and tagged and overexpressed enzymes predicted to be important for the production of triglycerides from glucose with green fluorescent protein to identify their cellular location. Analysis of localization revealed that many enzymes are localized to the endoplasmic reticulum and lipid droplets. This localization pattern is not absolute. We observed two enzymes (Diacylglycerol acyltransferase, Dga1) and (1-acyl-sn-glycerol-3-phosphate acyltransferase, Slc1) that localize to both the endoplasmic reticulum and the periphery of lipid droplets.

Publications

1. Pomraning, K.R., Kim, Y.-M., Nicora, C.D., Chu, R.K., Bredeweg, E.L., Purvine, S.O., Hu, D., Metz, T.O., Baker, S.E. 2016. Multi-omics analyses reveals regulators of the response to nitrogen limitation in *Yarrowia lipolytica.* BMC Genomics, *In press.*
2. Kerkhoven, E. Pomraning, K.R., Baker, S.E., Nielsen, J. 2016. Regulation of amino acid metabolism controls flux to lipid accumulation in *Yarrowia lipolytica.* npj Systems Biology and Applications, *In press*.
3. Pomraning, K., Baker, S.E. 2015. Draft genome sequence of the dimorphic yeast *Yarrowia lipolytica* strain W29. Genome Announc. 3(6). pii: e01211-15.
4. Pomraning, K.R., Wei, S., Karagiosis, S.A., Kim, Y.M., Dohnalkova, A.C., Arey, B.W., Bredeweg, E.L., Orr, G., Metz, T.O., Baker, S.E. 2015. Comprehensive metabolomic, lipidomic and microscopic profiling of *Yarrowia lipolytica* during lipid accumulation identifies targets for increased lipogenesis. PLoS One. 10(4):e0123188.

Funding for this research provided by the DOE BER Genome Science Program grant DE-SC0008744.

Rewiring metabolism for maximum lipid production in oleaginous yeast *Yarrowia lipolytica*

Kangjian Qiao[1] (kjqiao@mit.edu) and **Gregory Stephanopoulos**[1]

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA

Project Goals: We focused on achieving a fundamental understanding of the metabolism of the oleaginous yeast *Yarrowia lipolytica* and developing tools to characterize and engineer it towards cost-effective lipids production. More specifically, we aimed to improve its fermentation characteristics towards the development of a cost-effective process which converts renewable resources to lipids for biodiesel production. The conversion yield and volumetric productivity on various carbon sources are the key metrics for optimization.

Production of lipids by microbial fermentation of carbohydrate feedstocks outcompetes oil crops in terms of productivity, however, presently achievable carbohydrates-to-lipids process yields are not yet at a point that can support cost-effective production of lipids and biodiesel. To maximize process yields, one needs to maximize lipid content as well as capture as many of the electrons generated from the catabolism of the available substrate as possible. We show that overall lipid process yield, Y, is significantly improved via introduction of synthetic pathways that effectively recycle glycolytic NADHs into cytosolic NADPH and acetyl-CoA to be used for lipid synthesis. Strain construction is guided by a quantitative model that predicts Y from the non-lipids biomass yield, lipid content, and yield of lipid synthesis pathway, Y_L, with Y_L becoming the key driver of process yield maximization at high lipid content. In total, thirteen rationally designed strain constructs were evaluated in shake flask and bioreactor experiments to identify the best strain that achieved a lipid titer of close to 100 g/L with a productivity of 1.3 g/L/h and an overall process yield 0.27 g FAME/g glucose. These figures of merit advance the commercialization opportunities of carbohydrate-based biodiesel production.

Intracellular Metabolite Pool Size Quantification in Oleaginous Yeast *Yarrowia lipolytica* using Acetate as Carbon Source

Nian Liu[1,*] (nianl@mit.edu) and **Gregory Stephanopoulos[1]**

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA

Project Goals: This project aims to develop a cost-effective process that uses the oleaginous yeast *Yarrowia lipolytica* as a biocatalyst to convert carbon sources into fatty acids for biodiesel production. In particular, we are interested in using acetate as an inexpensive, renewable carbon source and understanding fundamentally how the metabolic pathways are coordinated to produce lipids. From this, we will be able to identify bottlenecks within the metabolic network and provide suggestions on engineering strategies that help improve the lipid yield, titer, and productivity.

The use of the oleaginous yeast *Yarrowia lipolytica* as a biocatalyst to convert acetate into fatty acids for biodiesel production is a renewable process and can potentially be cost-effective. However, due to the limited understanding of its metabolism on acetate, engineering the strain to achieve high lipid yield, titer, and productivity can be difficult. This work serves to provide a fundamental understanding of the metabolism through the quantification of intracellular metabolite pool sizes. Differences across two strains were compared in this study: the wild type strain (WT) and a previously engineered lipid overproducing strain (ACCDGA). A total of 17 intracellular metabolites in the glycolytic, tricarboxylic, and pentose phosphate pathways were quantified. Reducing cofactors (i.e. NADH, NADPH) and energy metabolites (i.e. ATP, ADP, AMP) were also analyzed. Our results indicate that the ACCDGA strain has higher reducing capacity and increased acetyl-CoA usage, both contributing to higher lipid production. Furthermore, upon transitioning from the growth phase to the lipid production phase (nitrogen depletion in media), both strains shift into a metabolic state that increases generation of the reducing cofactors favorable for lipid production. Finally, both strains are in an energy deficient state when cultured on acetate as the sole carbon source. This indicates that ATP requirements must be considered when engineering the strain.

Put on a diet: Lipid accumulation inhibits TOR signaling in *Saccharomyces cerevisiae*

Michael Gossing[1]* (gossing@chalmers.se) and **Jens Nielsen[1]**

[1]Systems and Synthetic Biology, Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

http://sysbio.se/

Project Goals: Lipids are a group of highly diverse molecules with a multitude of biological functions such as formation of biological membranes, storage of energy, cell signaling, and apoptosis. Triacylglycerides (TAG) function as energy storage and source of membrane building blocks. TAG are of particular interest since they can serve as a feedstock for production of oleochemicals and biodiesel. We are interested in studying the regulatory mechanisms behind TAG formation.

We have engineered *Saccharomyces cerevisiae* to accumulate increased levels of TAG. Accumulation was achieved by introducing a push and pull on TAG biosynthesis. A push was introduced by overexpression of acetyl-CoA carboxylase double mutant $ACC1^{S659A\ S1157A}$ (*ACC1***). The gene product of *ACC1*** is constitutively active because the protein kinase *Snf1p* cannot inactivate the mutated carboxylase by phosphorylation. A pull was introduced by overexpression of phosphatidate phosphatase *PAH1* and diacylglycerol acyltransferase *DGA1*. The resulting strain was analyzed for transcriptional changes in fermentative and respiratory growth phase compared to a reference strain.
Despite a strongly increased flux towards fatty acids and TAG in respiratory phase, we were unable to observe major transcriptional changes for genes involved in fatty acid and TAG biosynthesis. However, we observed an upregulation of β-oxidation, and a downregulation of phospholipid metabolism. Gene set analysis (GSA) revealed changes in biological processes that are not directly linked to lipid metabolism. We observed a downregulation of translational initiation, ribosome biogenesis, cellular amino acid biosynthesis, and rRNA processing, as well as an upregulation of the cell wall integrity pathway, and autophagy. These biological processes are (partially) regulated by the TOR complex. TOR is a protein complex that responds to energy and amino acid levels. It regulates many aspects that are related to cell growth. These data indicate that lipid accumulation impinges on TOR signaling.

This work is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science program, under Award Number DE-SC0008744.

Transcriptional regulation and lipid accumulation in *Yarrowia lipolytica*

Eduard J. Kerkhoven[1]* (eduardk@chalmers.se), Kyle Pomraning[2], Carrie D. Nicora[3], Thomas L. Fillmore[3], Samuel O. Purvine[2], Richard D. Smith[3], Thomas O. Metz[3], Scott E. Baker[2] and **Jens Nielsen**[1]

[1]Chalmers University of Technology, Gothenburg, Sweden; [2]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington; [3]Fundamental and Computational Science Directorate, Pacific Northwest National Laboratory, Richland, Washington

Project Goals: Our goal is to elucidate the regulation of lipid metabolism in *Y. lipolytica* to identify new targets to improve the TAG yield.

Oleaginous yeasts such as *Yarrowia lipolytica* are capable of accumulating lipids up to 70% of their biomass, predominantly in the form of triacylglycerols (TAGs), and this has fuelled interest in exploiting these fungi for the production of biodiesel. To optimise yields, we are studying the metabolic fluxes and their regulation in *Y. lipolytica* in different growth conditions on a genomic scale.

A multi-factorial chemostat experiment was designed, including two growth conditions (nitrogen and carbon limitation) and two strains (a wild-type and a DGA1 overexpressor). These cultures were sampled for various omics technologies (RNAseq, proteomics [performed under the Pan-omics program at PNNL], metabolomics, lipidomics) and integrative data analysis was performed with this comprehensive dataset and the aid of a genome-scale model of *Y. lipolytica* metabolism.

Lipid accumulation is substantially increased upon nitrogen limitation and DGA1 overexpression, however, both transcriptomics and proteomics indicate the lack of regulation of expression levels of lipid biosynthetic enzymes. The genes that were differentially expressed showed a strong correlation between transcript and protein levels, indicating that many of these are under transcriptional regulation.

The strongly correlated genes were enriched for involvement in amino acid biosynthesis. More precisely, specifically leucine biosynthesis was transcriptionally downregulated upon high lipid accumulation in the DGA1 strain during lipid accumulation. This was further supported by metabolomics data, which indicated the intermediate of leucine biosynthesis 2-isopropylmalate as overflow metabolite during nitrogen limitation in the WT strain, while absent at high lipid accumulation during nitrogen limitation in the DGA1 strain.

While the lipid biosynthetic pathway is not transcriptionally regulated at high lipid accumulation, leucine biosynthesis is specifically down-regulation, having the effect that carbon flux is redirected from amino acid biosynthesis towards lipid accumulation. This suggests new and unexpected targets for further metabolic engineering.

Designing a Microbial Community for Production of Biofuel from Lignocellulose

Meghan Thommes[1]* (mthommes@bu.edu), Andrea Lubbe[2]* (alubbe@lbl.gov), Jessica Lee[3,4,5]* (jessicalee@uidaho.edu), Arion Stettner[1], Ilija Dukovski[6], Alyssa Baugh[3], Nicholas Shevalier[3], Joshua Wirtz[3,5], Sergey Stolyar[3,4,5], Christopher Marx[3,4,5], Trent Northen[2] and **Daniel Segre**[1,6,7]

[1]Dept. of Biomedical Engineering, Boston University, Boston, Massachusetts; [2]Lawrence Berkeley National Laboratory, Berkeley, California; [3]Dept. of Biological Sciences, [4]Institute for Bioinformatics and Evolutionary Studies, and [5]Center for Modeling Complex Interactions, University of Idaho, Moscow, Idaho; [6]Bioinformatics Program, Boston University, Boston, Massachusetts; [7]Dept. of Biology, Boston University, Boston, Massachusetts

Project Goals: The goal of our research is to implement a high-throughput pipeline for the systematic, computationally-driven study and optimization of microbial interactions and their effect on lignocellulose degradation and biofuel production. We combine multiple approaches, including computer modeling of ecosystem-level microbial metabolism, mass spectrometry of metabolites, genetic engineering, and experimental evolution.

The microbial production of biofuels from lignocellulose is a complex, multi-step process. Microbial consortia are an ideal approach to consolidated bioprocessing: a community of microorganisms performs a wide variety of functions more efficiently and is more resilient to environmental perturbations than a microbial monoculture. Each organism we have chosen for this project addresses a specific challenge: lignin degradation (*Pseudomonas putida*); (hemi)cellulose degradation (*Cellulomonas fimi*); lignin degradation product demethoxylation (*Methylobacterium* spp); generation of biofuel lipid precursors (*Yarrowia lipolytica*). These organisms are genetically tractable, aerobic, and have been used in biotechnological applications.

In the past year, we have made significant experimental and computational progress both at the level of single organism characterization, and at the level of the assembly and observation of co-cultures.

In terms of individual organisms, we have analyzed the growth of *C. fimi*, *P. putida*, and *Y. lipolytica* in different environmental conditions. Experimentally, we have used mass spectrometry to characterize and measure the metabolic inputs and outputs of each of these consortium members, providing valuable information for model

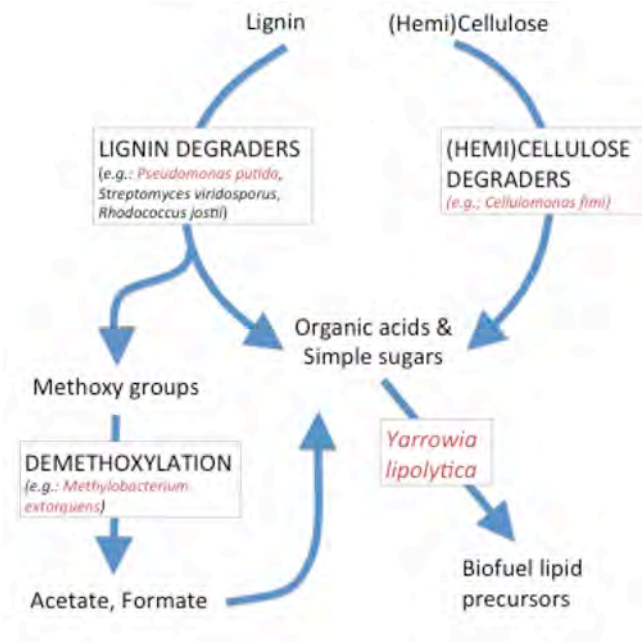


Figure 1. *Microbial community design.*

refinement, and for understanding possible metabolism-mediated interactions. On the computational side, we have implemented draft genome-scale models for all consortium members, based on KBase reconstructions and published information. Models have been imported into our platform for spatio-temporal dynamic flux balance modeling, COMETS (Computation Of Microbial Ecosystems in Time and Space). In parallel, we have started exploring the incorporation of extracellular enzymes in COMETS, in order to simulate more faithfully lignocellulose degradation processes in the community.

In addition to lignocellulose degradation, we have begun to address the challenge of removing the free formaldehyde produced by the demethoxylation of lignin monomers, which can otherwise inhibit microbial growth due to its toxicity. *M. extorquens* utilizes formaldehyde as a central metabolic intermediate, but cannot remove and detoxify the methoxy groups on lignin. We have identified soil-dwelling *Methylobacterium* species that can break down lignin degradation products, and are genetically engineering the well-studied *M. extorquens* so that it expresses these enzymes. We have also identified potential pathways to introduce the demethoxylation and cleavage of complex aromatics, with the goal to introduce these pathways into *M. extorquens*.

At the level of consortia – a key goal of this project - we have started studying how the lignocellulose degradation by *C. fimi* and *P. putida* can ensure a flow of saccharides to *Y. lipolytica* and of lignin degradation byproducts usable by *M. extorquens*. Using mass spectrometry, we have identified potential cross-feeding metabolites for synergistic growth of *C. fimi* with *Y. lipolytica* and *P. putida*, including metabolites released by *Y. lipolytica* and *P. putida* that are consumable by *C. fimi*. In order to develop an optimized defined medium for the biofuel consortium, we are developing minimal defined media for each individual consortium member that sustains growth and induces the production of cross-feeding metabolites.

Given the availability of draft genome-scale models, we have in parallel started performing COMETS simulations of metabolism in co-cultures of consortium members, yielding predictions of the metabolic activity of the whole community. Simulations will help us refine the optimal environmental conditions towards enabling the formation of a stable biofuel precursor-producing ecosystem.

This work was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy, Award No. DE-SC0012627, and the Office of Science Early Career Research Program, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

Spatial Connectomics to Identify Agents Relevant to Lignocellulose Deconstruction in Fungi

Jiwei Zhang[1], Melania Figueroa[1], Katarina Sweeney[1], Kenneth E. Hammel[2], Christopher J. Hunt[2], Ellen A. Panisko[3], Gerald N. Presley[1], and **Jonathan S. Schilling**[1]

[1]University of Minnesota, Saint Paul, [2]U.S. Forest Products Laboratory (FPL); [3]Pacific Northwest National Laboratory

URL: http://schillinglab.cfans.umn.edu

Project Goals: Our goal is to discover which genes are differentially up-regulated across the mycelia of brown rot wood-degrading fungi *in planta*, particularly at the leading edge of wood decomposition. These unique fungi accomplish what we have difficulty achieving – energy from plant biomass. To do this, brown rot fungi apparently couple an oxidative pretreatment step with enzymatic saccharification in discrete space, via partitioning of reactions. We have previously shown evidence that these are governed by differential expression, but with genes putative, transport out of hyphae unclear, and the secretome poorly studied, historically, a comprehensive approach is needed that can also limit ('winnow') data sets from powerful global analytical tools to focus on the genes and pathways that matter. To address this, we are comparing global expression profiles among mycelial regions and matching what we see to what they do, in terms of changing wood physiochemistry. To map a 'connectome' in a wood-fungal interaction, specifically, we must also match gene expression patterns with the extracellular secretome and with physiochemical wood modifications.

Objective 1 – **Zone localization:** Use the wood wafer design to resolve a depolymerization zone in *P. placenta* near the mycelial leading edge, and optimize RNA extraction for thin-sectioning.

Objective 2 - Fungal connectomics: Co-localize gene expression with the secretome and with relevant physiochemical modifications made to the wood, e.g. hemicellulose loss, porosity changes.

Objective 3 - Clade comparisons: Compare key zones among brown rot clades, in context of white rot same-clade ancestors, to target universal 'brown rot' genes and candidates for bioprocessing.

Abstract:
Certain filamentous fungi are uniquely able to deconstruct lignocellulose, and their poorly understood mechanisms have potential biofuels applications. A key hindrance to harnessing these fungal mechanisms has been their spatial complexity. Our past work has shown that differentiated networks of hyphae that penetrate wood are not metabolically uniform, with critical reactions occurring near the hyphal front. Standard omics analyses of these fungi from artificial media or from colonized wood ground en masse fail to distinguish expression of key gene products occurring in localized regions along growing hyphae.

Our focus for this research is specifically on brown rot fungi, a more recently evolved decay fungal group (relative to white rot) that circumvents the lignin barrier to extract sugars from lignocellulose. The genetic basis for how this capacity evolved away from white rot multiple times remains unknown, despite the modern options to align the compare brown rot and white rot fungal genomes. Our new collaboration aims to focus omics techniques to map and integrate expression over networks of wood-degrading fungal hyphae *in planta*. A similar approach,

‘connectomics,’ has been used to map the human nervous system, and its application here is timely. First, wood-degrading fungal genomes are an emerging resource, particularly brown rot functional types. Second, we recently optimized a thin-section wood set-up that can finely resolve reaction zones along an advancing mycelium. Within these zones, we can employ deep omics approaches without the typical noise of whole-sample homogenization. By co-localizing gene expression, secretions, and wood modifications, we can prioritize the genes most useful for application.

Our goal is to discover which genes are differentially up-regulated across the mycelia of wood-degrading fungi *in planta*, particularly at the leading edge of wood decomposition. To do this, we need to compare global expression profiles among mycelial regions. To map a ‘connectome’ in a wood-fungal interaction, specifically, we must also match gene expression patterns with the extracellular secretome and with physiochemical wood modifications. Given this potential for substrate-fungus feedback, we will cross-check genes using separate clade representatives for brown rot fungi alongside their white rot ancestors, harnessing the JGI MycoCosm portal and several key resources and expertise at the USDA Forest Products Laboratory and the Pacific Northwest National Laboratory.

References:

Kaffenberger, J.T. and Schilling, J.S. (online first) Comparing lignocellulose physiochemistry after decomposition by brown rot fungi with distinct evolutionary origins. Environmental Microbiology doi:10.1111/1462-2920.12615.

Schilling, J.S., Kaffenberger, J.T., Liew, F.J., Song, Z. (2015) Signature wood modifications that reveal decomposer community history. PLoS ONE 10(3): e0120679. doi:10.1371/journal.pone.0120679.

Schilling, J.S., Duncan, S.M., Presley, G.N., Filley, T.R., Jurgens, J.A, and Blanchette, R.A. (2013) Colocalizing incipient reactions in wood degraded by the brown rot fungus *Postia placenta*. International Biodeterioration and Biodegradation 83: 56-62.

Schilling, J.S., Ai, J., Blanchette, R.A., Duncan, S.M., Filley, T.R., and Tschirner, U.W. (2012) Lignocellulose modifications by brown rot fungi and their effects, as pretreatments, on cellulolysis. Bioresource Technology 116: 147-154.

Funding statement

This work was supported by the U.S. Department of Energy Office of Science (Early Career Grant DE-SC0004012 from the Office of Biological and Ecological Research (BER) to J.S.S.; and BER Grant DE-SC0012742). Confocal microscopy is funded by user Facility grant #48607 at the Environmental Molecular Sciences Laboratory of Pacific Northwest National Laboratory (to J.S.S.).

The Effect of Authentic Sinusoidal Light Variation on Growth and Productivity of Microalgae

Mark Hildebrand (mhildebrand@ucsd.edu),[1] Juergen Polle[2], and Michael Huesemann[3]

[1]Scripps Institution of Oceanography, University of California, San Diego; [2]CUNY Brooklyn College; [3]Pacific Northwest National Laboratory

Project Goals: The overarching goal of this project is to understand and manipulate fundamental molecular mechanisms involved in maximizing growth rate and lipid accumulation in diverse classes of microalgae under authentic diurnal conditions to enhance production capabilities for biofuels. The aim is to fill two critical knowledge gaps in microalgal cultivation by addressing fundamental cellular issues that govern growth and fuel molecule accumulation in conjunction with realistic climate-simulated cultivation: 1) The insufficient understanding of the effect of metabolic topology on cellular carbon partitioning and its regulation with regards to productivity, 2) the understudied effect of the diurnal and cell cycles on cellular metabolism, resource allocation, and resource mobilization under simulated production conditions.

Our team has focused on a comparative analysis of metabolic pathway topology in two classes of microalgae, diatoms and chlorophytes, as well as addressed the disconnect between laboratory cultivation and outdoor cultivation under sinusoidally-varying light intensity conditions. Microalgal performance under laboratory conditions commonly does not correlate with outdoor productivity because outdoor sinusoidal light intensity variation and diurnal temperature changes are not mimicked. Microalgae must accumulate defined amounts of energy, energy-storage compounds, and biochemical constituents to replicate. *Typically, cells are not continuously "growing" throughout the day, but they alternate periods of enlargement (G1 cell cycle phase) and division (S and G2+M phases).* Microalgal cultures commonly synchronize their division processes, dividing once or twice per day at specific times. We monitored gene transcript changes during synchronized cell cycle progression in the diatom *Thalassiosira pseudonana*, and identified cell cycle stage-specific dependence on nutrients. Energetic, nutrient, and metabolic processes vary throughout the cell cycle, and understanding the relationship between these parameters and times of day should have a significant impact on productivity, providing, for instance, insight into optimal times for addition of nutrients and harvesting.

We evaluated the response of the diatom *Cyclotella cryptica* to sinusoidal variation of daylight intensity and temperature on a 12:12 l:d cycle. Cell cycle synchronization occurred, with cells undergoing the division process in S and G2+M phases for a substantial portion of the day. There was no strict correlation between OD750, typically used as a proxy for biomass, and cell number. Small changes in OD750 occurred at night, suggesting minimal respiratory losses. Timing of chloroplast division and changes in cell volume related to the growth and division process. Variation in triacylglycerol (TAG) levels suggested accumulation prior to a maximal period of membrane synthesis, then additional accumulation of a pool prior to nighttime. Because productivity is affected by photosynthetic performance, we examined the photoadaptive response by tracking changes in photosynthetic and photoprotective pigment levels. Continuing investigation should lead to a fuller understanding of the effect of diurnal conditions and synchronized cell cycle progression on growth and productivity.

SYNECHOCOCCUS ELONGATUS UTEX 2973: METABOLIC MODELS AND FLUXOMIC ANALYSIS

Thomas J. Mueller[1*] (tjm5293@psu.edu), Saratram Gopalakrishnan[1*] (sxg375@psu.edu), Mary H. Abernathy[2*] (mary.h.abernathy@wustl.edu), Fangfang Ma[3], Michelle Liberton[4], Whitney Hollinshead[2], Lian He[2], Doug K. Allen[3], **Yinjie J. Tang**[2], **Costas D. Maranas**[1], and **Himadri B. Pakrasi**[2,4]

[1]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA; [2]Department of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, MO; [3]USDA-ARS/Donald Danforth Plant Science Center, St. Louis, MO; and [4]Department of Biology, Washington University, St. Louis, MO

http://pages.wustl.edu/photo.synth.bio

Project Goals: The overall objective of this project is to use an integrated systems biology approach to develop *Synechococcus* 2973, a fast growing cyanobacterial strain, as a platform organism for photobiological production of advanced biofuels and other useful chemicals. The project aims to improve the understanding of metabolic processes in this microbe through *in silico* analysis followed by experimental validations to enable efficient strain design. Overall the goals are three-fold: (a) to reconstruct a genome-scale metabolic model of *Synechococcus* 2973, (b) to develop a carbon-mapping model of this strain, and (c) to better understand its phenotypical properties by isotopically nonstationary metabolic flux analysis (INST-MFA).

Photosynthetic microbes offer advantages over commonly used heterotrophic microbes such as *E. coli* and *Saccharomyces cerevisiae* due to their ability to harvest light energy and fix carbon dioxide. Researchers have demonstrated that photosynthetic microbes like cyanobacteria have the ability to convert light energy and CO_2 into value-added chemicals. However, there are still key challenges in the use of cyanobacteria as microbial cell factories: low biomass yield and production rates. Recently, *Synechococcus elongatus* UTEX 2973 has been identified in the Pakrasi lab as a unicellular cyanobacterial strain with an extraordinarily fast growth rate under high light and CO_2 conditions.

Modeling and subsequent comparison of the metabolism of *Synechococcus* 2973 and its close (but slower-growing) relative *Synechococcus elongatus* PCC 7942 can provide further insights into these strains and their differing growth rates, along with enabling future design of metabolic interventions. Genome-scale metabolic (GSM) models were developed for both *Synechococcus* 7942 (*i*Syf686) and *Synechococcus* 2973 (*i*Syu627) using a semi-automated method that combined information from UniProt, NCBI Protein Clusters, and the RAST annotation pipeline with the previously developed *Synechocystis* GSM model *i*Syn731. The *i*Syf686 model correctly predicted 232 of 267 non-essential genes and 234 of 338 essential genes present within the model. Experimental measurements for CO_2 consumption were used to constrain the

corresponding fluxes. Differing rates of CO_2 fixation resulted in a growth rate for *Synechococcus* 2973 that was 1.7 times higher than that of *Synechococcus* 7942, which is in the same order of magnitude of the experimental observations (i.e., 3.7 times). In addition, carbon-mapping models can be useful in tracking the flow of carbon within metabolic models. However, the availability of a carbon-mapping model remains the major bottleneck in the genome-scale ^{13}C metabolic flux analysis of cyanobacteria. To this end, we have constructed the genome-scale metabolic mapping (GSMM) model *i*mSyf608 describing reaction atom mapping for all carbon-balanced reactions contained within *Synechococcus* 7942. This model was constructed by appending to the already published *E. coli* mapping model, *i*mEco726, and reaction mapping information generated using our previously developed CLCA algorithm. We found that the 99.8% genomic sequence similarity between *Synechococcus* 7942 and the fast-growing *Synechococcus* 2973 gives rise to a 100% identity of carbon-balanced reactions contained within their respective genome-scale models, resulting in the applicability of *i*mSyf608 for ^{13}C-MFA of *Synechococcus* 2973.

In parallel to these modeling efforts, to better understand the phenotypical properties of *Synechococcus* 2973, pulse-chase experiments and isotopically nonstationary metabolic flux analysis (INST-MFA) were performed. INST-MFA allows for the quantification of autotrophic carbon fluxes through transient ^{13}C-labeling patterns of free metabolites. Using a “deep-frozen” metabolite quenching method followed by ion-pairing LC-MS/MS and ESI-TOF-MS, we efficiently recovered fast-turnover metabolites (including sugar phosphates and organic acids) for the analysis of isotopomer dynamics as well as relative pool size determination. Using such data and the computational platform, INCA, we deciphered the fluxomic topology for photoautotrophic growth of *Synechococcus* 2973. Our research yielded three new insights. First, *Synechococcus* 2973 has no measureable fluxes through the oxidative pentose phosphate pathway. This was further confirmed by the normal growth of a *zwf* inactivation mutant. Second, metabolic channeling from 3PGA towards the TCA cycle is evident, indicating an internal heterogeneous distribution of central metabolic enzymes. Third, compared to *E. coli*, cyanobacterial metabolite pool sizes are much higher for sugar phosphates, while the TCA cycle metabolites and acetyl-coA pools are about 5-10 fold lower. This indicates that *Synechococcus* 2973 is an ideal chassis for producing chemicals derived from glycolysis and the Calvin cycle.

In future work, we will use the genome-scale and carbon-mapping models to derive efficient strain designs for *Synechococcus* 2973, in order to study its growth phenotypes for various carbon sources and thereby pinpoint the key reasons for its fast growth from the metabolic point of view. Furthermore, these models will also be utilized to carry out additional INST-MFA analyses. Ultimately, such studies will help develop this cyanobacterium as a novel chassis for the production of fuels and value-added chemicals.

These studies have been supported by funding from the Office of Biological and Environmental Research in the DOE Office of Science to HBP, YJT and CDM.

SYNECHOCOCCUS ELONGATUS UTEX 2973, A NEW CYANOBACTERIAL STRAIN THAT EXHIBITS RAPID AUTOTROPHIC GROWTH

Justin Ungerer[1]* (JustinUngerer@wustl.edu), Kristen E. Wendt[1], Rajib Saha[1], **Yinjie J. Tang**[2], **Costas D. Maranas**[3], and **Himadri B. Pakrasi**[1,2]

[1]Department of Biology, Washington University, St. Louis, MO; [2]Department of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, MO; and [3]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA

http://pages.wustl.edu/photo.synth.bio

Project Goal: The overall objective of this project is to use an integrated systems biology approach to develop *Synechococcus* 2973, a fast growing cyanobacterial strain, as a platform organism for photobiological production of advanced biofuels and other useful chemicals. An aim is to develop a genetic tool kit that will enable facile metabolic engineering of this strain. Another aim is to generate a genome wide library of transposon mutants that will be made available to researchers for future studies. We intend to use this library to identify essential genes for specific metabolic characteristics such as sugar utilization, high light tolerance, or rapid growth. Finally, we will measure photosynthetic parameters to identify factors that are critical to rapid growth.

Photosynthetic microbes are of considerable interest for applications in carbon sequestration, photosynthetic production of fuels and other valuable chemicals. The advantage of using cyanobacteria as biofactories is that they can grow on CO_2 and sunlight alone, which reduces greenhouse gas emissions and moves society away from dependence on petroleum-based products. Unfortunately, commonly used cyanobacterial strains exhibit growth rates that are much slower than conventional heterotrophic biofactories such as *E. coli* and yeast. This leads to inherently lower productivity from the cyanobacteria. We have recently identified a cyanobacterial strain, *Synechococcus elongatus* UTEX 2973, that has significant industrial potential. *Synechococcus* 2973 exhibits autotrophic biomass productivity that is comparable to that observed in heterotrophs such as yeast. Under high light and high CO_2 conditions, this strain exhibits a doubling time of 1.9 hours [1]. Genome sequencing revealed that *Synechococcus* 2973 is a close relative of the slower growing and widely used strain, *Synechococcus elongatus* PCC 7942. The two strains differ by only 55 SNPs, a 188 kb inversion, and a 7.5 kb deletion.

Synechococcus 2973 has proven resistant to natural transformation. Recent research efforts have aimed to develop a facile genetic system for this strain. We have been successful in introducing a replicating plasmid containing a fructose uptake system and engineering an *nblA* deletion mutant using conjugation [1, 2, 3]. The strain with the fructose uptake system actively transports fructose and exhibits increased growth rates under mixotrophic conditions. Strains with the *nblA* knockout show the non-bleaching phenotype that is characteristic of this deletion. Additionally, we have also set out to develop a CRISPR/Cas9 genome editing system for *Synechococcus* 2973. As a proof of concept for this system, we have generated a markerless Δ*nblA* strain, exhibiting the expected non-bleaching phenotype.

We have developed a high efficiency transposon mutagenesis system and demonstrated that it inserts randomly across the genome. We are currently using this transposon along with next

generation sequencing technology to generate a library of mutants of *Synechococcus* 2973 that will be made publicly available. Several of the most interesting mutants will be selected from the library for MFA analysis by the Tang lab. Data collected from these mutants will be used to further refine the metabolic models developed by the Maranas lab. The refined models will be exploited to generate hypotheses regarding efficient strain designs as well as fluxomic states under given conditions. These findings will then be tested in the Pakrasi and Tang labs. Using NGS we will screen the library for all mutants that have a changed fitness value under selected conditions. This will allow us to assign essentiality to any gene needed for the process we are querying. Conditions that are being initially studied include light tolerance, CO_2 utilization, and mixotrophic growth.

We are also characterizing photosynthetic parameters for *Synechococcus* 2973, and comparing them with those for the slower growing *Synechococcus* 7942. Such studies will identify differences in photosynthesis that differentiate the fast growing strain from the slower growing strain and will point to adaptations that can be reconstructed in other organisms to increase their photosynthetic output.

References:

1. Yu, J., et al., *Synechococcus elongatus UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO_2.* Scientific Reports, 2015. **5**: 8132.
2. Ungerer, J., B. Pratte, and T. Thiel, *Regulation of Fructose Transport and Its Effect on Fructose Toxicity in Anabaena spp.* Journal of Bacteriology, 2008. **190**: 8115-8125
3. Collier, J.L. and A.R. Grossman, *A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria.* The EMBO Journal, 1994. **13**: 1039-1047.

These studies have been supported by funding from the Office of Biological and Environmental Research in the DOE Office of Science to HBP, YJT and CDM.

The Algal Ferredoxin Interactome

Venkataramanan Subramanian[1]* (venkat.subramanian@nrel.gov), Regina Vasconcellos Antonio[1,2], Alexandra Dubini[3], David Gonzalez-Ballester[3], Patrice Hamel[4], **Maria Ghirardi[1]**

[1]National Renewable Energy Laboratory, Golden, CO; [2]Universidade Federal de Santa Catarina, Araranguá, Brazil; [3]Universidad de Córdoba, Córdoba, Spain; [4]The Ohio State University, Columbus; OH.

Project Goals: Our goal is to unravel the specific roles of the different ferredoxins in the green alga, *Chlamydomonas reinhardtii* and their contribution, either singly or in combination, towards overcoming different physiological and environmental stresses.

Abstract: Ferredoxins (FDX) are small, iron-sulfur cluster containing proteins with strong negative redox potentials (-350 to -450 mV) that transfer electrons to reductive steps in various metabolic pathways. There are 13 FDX isoforms predicted in the *Chlamydomonas* genome that are differentially expressed in response to varying environmental conditions such as availability of copper, oxygen, iron and ammonium. Previous research in our lab has established a global FDX interaction network along with demonstrating that they may have redundant functions (Peden et al. 2013, Boehm et al. 2015). That work revealed that (1) FDX2 may have an overlapping role with FDX1 in donating electrons for H_2 production and $NADP^+$ reduction and, possibly, in mediating state transitions and/or cyclic electron transport; (2) FDX3, together with FDX1 and FDX2 is involved in nitrogen assimilation; (3) FDX4 interacts with glycolytic enzymes and enzymes involved in protection against ROS; and (4) FDX5 may be required for hydrogenase maturation, together with FDX2 and FDX4, and has been shown to be involved in fatty acid synthesis in the dark (Yang et al. 2015). It has also been shown in the literature that overexpression of specific FDXs (FDX1 and FDX5) results in an increase in starch and oil content as well as increased tolerance to heat-, salt-, and oxidative- stresses in this alga.

Towards understanding the specific role of the FDXs, we have analyzed the response of the *FDX2* gene knock-out (Δ*fdx2*) towards H_2 production, starch content, oxidative stress, and triacylglycerol (TAG) accumulation, under sulfur deprived (-S) conditions. Based on our observation, there are no structural phenotypic differences between the wild-type and the Δ*fdx2* strain. Both the wild-type and the Δ*fdx2* strain produce ~ 18 mmol H_2 over a period of 72 h under -S conditions. They have similar starch accumulation profiles, which increases up to 24 h after sulfur deprivation, followed by a gradual reduction in starch levels, which is accompanied by a consistent increase in H_2 production. Although starch accumulation was found to be similar in both the strains, chlorophyll synthesis was found to be slightly higher in the wild-type strain (27 μg/ml) in comparison to the Δ*fdx2* (22 μg/ml) during this time period. Metabolite analysis revealed that both strains accumulated almost similar levels of ethanol and formate, measured at 72 h post sulfur-deprivation. The levels of acetate decreased in the culture at 24 h post sulfur-deprivation and then remained the same till 72 h. This was consistent with consumption of

acetate during the oxic phase of the culture. On the other hand, our plate-based assay to determine tolerance to oxidative stress revealed that Δ*fdx2* strain was more sensitive to H_2O_2 under –S conditions. This effect was more pronounced under high light intensities, as expected. Based on microscopy, there was no apparent change in the neutral lipid content of Δ*fdx2* strain. Overall, deletion of *FDX2* gene seems to affect the strain's ability to combat reactive oxygen (ROS) species in this organism, especially under high light conditions when ROS levels are high. On the other hand, preliminary analysis reveals that Δ*fdx2* shows lower H_2 photoproduction activity (~ 30%) during a 24 h dark, anaerobic induction period, as well as slightly lower (~ 10%) dark fermentative H_2 production.

References and publications generated by this project:

1. **Boehm M**, Alahuhta M, Mulder DW, **Peden EA**, Long H, Brunecky R, Lunin VV, King PW, **Ghirardi ML**, **Dubini A**. (2015). Crystal structure and biochemical characterization of Chlamydomonas FDX2 reveal two residues that, when mutated, partially confer FDX2 the redox potential and catalytic properties of FDX1. Photosynth. Res. DOI 10.1007/s11120-015-0198-6.

2. **Peden EA**, **Boehm M**, Mulder DW, Davis R, Old WM, King PW, **Ghirardi ML**, **Dubini A**. (2013). Identification of global ferredoxin interaction networks in *Chlamydomonas reinhardtii*. J. Biol. Chem. 288(49): 35192-209.

3. Yang W, Wittkopp TM, Li X, Warakanont J, **Dubini A**, Catalanotti C, Kim RG, Nowack EC, Mackinder LC, Aksoy M, Page MD, D'Adamo S, Saroussi S, Heinnickel M, Johnson X, Richaud P, Alric J, **Boehm M**, Jonikas MC, Benning C, Merchant SS, Posewitz MC, Grossman AR. (2015). Critical role of *Chlamydomonas reinhardtii* ferredoxin-5 in maintaining membrane structure and dark metabolism. Proc Natl Acad Sci U S A. 112(48):14978-83.

This research is being supported by DOE's Office of Science, Biological and Environmental Research Office under a Science-Focused Area (SFA) project to NREL.

Design Principles Controlling Hydrogen Metabolism in Phototrophic Organisms

Task 1

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

[1]National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO 80401
[2]Genebiologics, LLC, Boulder, CO 80303
*Presenting author

Project goals: To obtain a systems-level understanding of the biological barriers that control hydrogen metabolism and prevent sustained H_2 photoproduction in the green alga *Chlamydomonas reinhardtii*.

Photobiological H_2 production from water is a clean, non-polluting and renewable technology. Although the potential light conversion efficiency to H_2 by biological organisms is theoretically high (about 10%), the system is currently limited by biochemical and engineering constraints. The specific objectives of this research are covered by two Tasks: (1) development, testing, validation and utilization of novel high-throughput assays to identify photosynthetic organisms with altered H_2-producing activities, thus leading to the discovery of novel strategies to circumvent known biochemical limitations (this task is now completed); and (2) deconvolution of the network of metabolic pathways centered on six ferredoxin homologs found in Chlamydomonas, aimed at understanding reductant flux in photobiological hydrogen production, and identifying targets for future metabolic pathway engineering strategies to reduce flux to non-productive pathways.

In completing Task 1, we developed a novel high-throughput assay that allows the screening of up to 500K algal colonies per minute for H_2-production characteristics. This assay is now in use by eight different labs; it is notable for its ease of use as a bench-top screen for H_2-producing organisms whose H_2 production systems are essentially O_2-intolerant. The assay uses the H_2-sensing system of *Rhodobacter capsulatus* and couples the sensor to activation of a GFP signal (Wecker et al., 2011). We validated this assay with well-characterized mutants that are either low or high H_2-producers (Wecker and Ghirardi, 2014). We then used the assay on a *C. reinhardtii* insertional mutagenesis library and isolated four strains of *C. reinhardtii* capable of high-light H_2 production. These mutants showed up to 180-fold greater levels of H_2 production at elevated light levels in comparison to the wild type strain. The greatest high-light H_2-production increase is putatively the result of a mutation in a pherophorin gene. We are complementing this mutation to determine if the mutant with a replacement gene regains its original phenotype.

We have also created a model system for developing photosynthetic H_2 production. We inserted a heterologous hydrogenase from *Clostridium acetobutylicum* into our *R. capsulatus* sensor strain. The imported enzyme produces H_2, and the H_2 produced in turn causes the organism to fluoresce. Moreover, H_2 production can be repeatedly induced by dark cycles followed by illumination. This

system therefore represents a novel means of testing, developing and selecting for *in vivo* hydrogenase activity, for photosynthetic H_2 production, and for metabolic support of H_2 production with real time output of hydrogen production reported on a cellular level. A manuscript is near completion.

Publications:

1. Wecker, M.S. and M.L. Ghirardi, *High-throughput biosensor discriminates between different algal H_2-photoproducing strains*. Biotechnology and Bioengineering, 2014. 111(7): p. 1332-1340.
2.
3. Wecker, M.S. and M.L. Ghirardi, *Expression and Photosynthetic Activity of a Clostridial [FeFe] Hydrogenase in the Bacterium Rhodobacter capsulatus* (In Preparation)

This project was supported by the Office of Science (BER) under FWP #ERWER38.

A day in the life of Chlamydomonas

Daniela Strenkert[1*] (danielas@chem.ucla.edu), Sean D. Gallaher[1], Stefan Schmollinger[1], Eric Soubeyrand[2], Gilles J. Basset[2] and **Sabeeha S. Merchant**[1]

[1]Institute for Genomics and Proteomics, University of California, Los Angeles, CA, USA
[2]Current address: Horticultural Sciences Department, University of Florida, Gainesville, FL, USA

Project goals: The green alga *Chlamydomonas reinhardtii* is a reference organism for many questions in biology, especially photosynthesis. To better understand this important model organism, we generated a rich functional genomics dataset, combining transcriptomic data with physiological measurements over the course of a Chlamydomonas cell cycle. Cultures were grown in a flat panel bioreactor system for reproducible synchronization. RNA-seq data was collected in triplicate at regular intervals over the course of a 12-h light/12-h dark cycle. In order to obtain a global view of expression, the transcriptomic data was sorted using a K-means clustering strategy. We observed that 81% of all transcribed genes were differentially expressed at one or more time points, and these were grouped into 16 major expression patterns describing the sequence of events during a day in the alga's life. Among the most interesting responses were the orchestration of cell division and fermentation during the night, and the balance between photoprotection and photosynthesis during the day.

During the cell cycle, DNA replication and cell division are initiated at the beginning of the dark period. The demand to duplicate the core histones during S phase is met by the coordinated expression of multiple genes encoding each of the histones (2). In metazoans, replication-dependent histone mRNAs differ from canonical mRNAs in that they lack a poly-A tail. In contrast, plants and many unicellular eukaryotes express only poly-Adenylated histone transcripts (3). The existence of non-poly-Adenylated histone mRNAs was assumed to be a unique to metazoans, until their recent discovery in the green algae Chlamydomonas (4) and Volvox (5). In our RNA-seq data, we observed coordinated expression of multiple genes encoding each of the core histones H2A, H2B, H3 and H4, as well as the linker histone H1. Interestingly, we also observed constitutive expression from two of each of the histone genes, suggesting that they provide an emergency supply of histones. The helicase of the mini-chromosome maintenance (MCM) complex involved in DNA replication, precedes the expression of core histone genes, ensuring that newly synthesized nucleosomes can be loaded on replicated DNA. The expression of the MCM complex increased more than 20-fold just prior to the onset of the dark phase.

After replication, the cells remain in G_0 for the remainder of the night. In Chlamydomonas, starch represents the major carbon storage molecule. We observed a maximum of starch accumulation at the end of the light period, directly prior to cell division. We also observed a dramatic reduction in the quantity of starch per cell occurring shortly after the onset of the night phase, which is consistent with redistribution of starch to daughter cells during mitosis. We measured the oxygen consumption of the cells during the cell cycle and found that the cells do not respire throughout the night. Instead, we observed that genes involved in anaerobic metabolism, such as the gene encoding an iron hydrogenase (*HYDA1*) were up-regulated.

The dark-light transition at the onset of a 12-hour photoperiod was concomitant with a rapid increase in stress related light harvesting genes involved in photo-protection. This increase was accompanied by reduced photosystem II capacity. Surprisingly, dusk-dawn experiments, in which the onset of light was gradually increased over a period of 2 hours, demonstrated that photosystem II capacity was still reduced during the dark light transition.
The plastid and nucleus encoded subunits of the chloroplast electron transfer chain were coordinately expressed, with peak expression occurring in the middle of the photoperiod. The increase in subunits of the electron transfer chain was also accompanied by an increased plastoquinone pool and more chlorophyll per cell, finally leading to increased oxygen evolution at the end of the day. This oxygenic photosynthesis was inevitably accompanied by the production of reactive oxygen species (ROS) that are generated within the electron transfer chain in chloroplasts. ROS production via photosystem I and photosystem II is known to be elevated, especially when carbon dioxide is limited, as it was in this experiment (1). Genes encoding ROS scavenging enzymes, such as ascorbate peroxidase 1 (*APX1*) and and Mn and Fe superoxide dismuates (*MSD1* and *FSD1*) were collectively up-regulated throughout the light period. In addition, there was increased expression of genes involved in the biosynthesis of a specific group of non-enzymatic anti-oxidants throughout the light period, namely tocopherols (Vitamin E). The major tocopherol, α-tocopherol was strongly induced after onset of light, but only the accumulation of γ-tocopherol increased proportionately with the increase in biomass.

References

1. Gururani, M. A., Venkatesh, J., and Tran, L. P. (2015) Regulation of Photosynthesis during Abiotic Stress-Induced Photoinhibition. *Mol Plant* 8(9):1304-20
2. Marzluff, W. F., Wagner, E. J., and Duronio, R. J. (2008) Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nature Reviews Genetics* 9, 843-854
3. Marzluff, W. F., and Duronio, R. J. (2002) Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. *Current Opinion in Cell Biology* 14, 692-699
4. Fabry, S., Muller, K., Lindauer, A., Park, P. B., Cornelius, T., and Schmitt, R. (1995) The Organization Structure and Regulatory Elements of Chlamydomonas Histone Genes Reveal Features Linking Plant and Animal Genes. *Current Genetics* 28, 333-345
5. Muller, K., Lindauer, A., Bruderlein, M., and Schmitt, R. (1990) Organization and Transcription of Volvox Histone-Encoding Genes - Similarities between Algal and Animal Genes. *Gene* 93, 167-175

Iron storage in ferritin versus a lysosome-related compartment in the green alga *Chlamydomonas reinhardtii*

Anne Glaesener,[1] Stefan Schmollinger,[1*](schmolli@chem.ucla.edu); Marcus Miethke,[1] Allegra Aron,[2] Jefferson Chan,[2] April Kusowski,[3] Peter K. Weber,[4] Timothy L. Stemmler,[3] Christopher J. Chang,[2] Jennifer Pett-Ridge,[4] and **Sabeeha S. Merchant**[1]

[1]University of California, Los Angeles; [2]University of California, Berkeley; [3]Wayne State University, Detroit, MI; [4]Lawrence Livermore National Laboratory, Livermore, CA

URL: http://www.chem.ucla.edu/dept/Faculty/merchant/#research

Project goals: Iron is an element of crucial importance, iron-deficiency limits primary productivity on a global scale, in croplands as well as in the oceans. The photosynthetic apparatus is thereby a key link, because of its central role in light-driven carbon assimilation, its abundance and iron-dependence. The single-celled, eukaryotic green alga *Chlamydomonas reinhardtii* is an excellent model to study iron deficiency, with all the advantages of a microbial system and the components of the iron assimilation machinery identified. In eukaryotes, the predominant site for iron storage is either ferritin or a vacuole. While ferritin is more and more viewed as a dynamic reservoir or iron buffer, the vacuole recently gained more attention as an iron storage site. The goal of this project is to provide a cellular view of iron metabolism, by understanding the pathways and dynamics of iron distribution between different iron storage compartments, especially during metabolic transitions.

Abstract: Chlamydomonas requires a broad spectrum of metal cofactors to sustain its photosynthetic and respiratory capacities, and iron (Fe) is one of the major transition metals involved in these processes [1]. The demand for Fe is reflected by its intracellular quota, which in Chlamydomonas strongly depends on the lifestyle (phototrophic, heterotrophic or photoheterotrophic) and the available external Fe concentrations. Four stages of iron nutrition can be distinguished in Chlamydomonas: iron-limited (below 0.5µM in the growth media), iron-deficient (between 1-3 µM), iron-replete (5-30 µM) and iron excess (above 50 µM). While iron-limited cells are chlorotic and growth-inhibited, iron-deficient cells have no obvious phenotype, but already activated the full transcriptional response towards iron deficiency. Iron-replete and iron-excess cells have not activated this response, but can be distinguished by their intracellular iron content. Especially interesting regarding iron storing, is that the more abundant of the two ferritin proteins that are encoded in the Chlamydomonas genome, the chloroplast-targeted Fer1, shows an unusual accumulation pattern upon different external iron concentrations [2]. Fer1 preferentially accumulates in iron deficient and iron-limited cells, while it is less abundant in iron-replete and -excess conditions, when iron is abundantly present, suggesting a different role than Fe storage.

To better understand Fe homeostasis in this reference organism during acclimation to changing Fe bioavailability, we used a suite of complementary techniques including (HPLC)-ICP-MS, confocal microscopy with Fe(II)-specific fluorescent sensors, NanoSIMS isotopic imaging and XAS to visualize individual compartments of iron storage, while providing different concentrations of extracellular Fe, ranging from 0.1 µM Fe (iron-limited) to 200 µM Fe (iron-excess). We found that intracellular iron concentrations observed in these conditions vary from $3x10^7$ atoms/cell at the lowest concentrations up to $6x10^8$ atoms/cell in excess iron conditions. We isolated the soluble Fer1 fraction grown under different Fe concentrations by size exclusion chromatography and quantified both protein and Fe contents. Fe content / ferritin varies dramatically between the different conditions, and is especially increased in iron-excess conditions, where ferritin abundance is reduced but cellular Fe content is at its peak. In iron-limited and iron-deficient cells the major iron storage compartment is indeed ferritin, where most of the intracellular iron is found. To understand how accumulated Fe in excess conditions is stored at the individual cell level, we stained cells with a Fe(II)-specific fluorescent dye, and in the high iron treatment, found it appears to be sequestered in distinct subcellular compartments. Complementary NanoSIMS studies revealed a partial co-localization of Fe, Ca and P, suggesting that at higher external iron concentrations an additional iron storage compartment was identified, linked to a lysosome-related structure known as acidocalcisome [3]. Mutants lacking the acidocalcisome compartment were not capable of accumulating additional Fe in iron-excess conditions and didn't display the additional Fe sequestration sites. Using stable-isotope labelling (^{57}Fe), we were able to follow the kinetics of Fe enrichment in the different storage compartments during a shift from iron-limited conditions (^{56}Fe) to iron-excess conditions (^{57}Fe). During the initial hours of the shift, the ^{57}Fe label was strongly enriched in the Fer1 pool before it accumulated in the lysosome-related compartment and subsequently was depleted from the Fer1 pool. Thus, these studies allowed us to dissect the intracellular Fe distribution pathways both kinetically and spatially in different Fe regimes.

References

1. Blaby-Haas CE, Merchant SS (2013) Iron sparing and recycling in a compartmentalized cell. Curr Opin Microbiol 16: 677-685.
2. Long JC, Sommer F, Allen MD, Lu SF, Merchant SS (2008) FER1 and FER2 encoding two ferritin complexes in Chlamydomonas reinhardtii chloroplasts are regulated by iron. Genetics 179: 137-147.
3. Hong-Hermesdorf A, Miethke M, Gallaher SD, Kropat J, Dodani SC, et al. (2014) Subcellular metal imaging identifies dynamic sites of Cu accumulation in Chlamydomonas. Nat Chem Biol 10: 1034-1042.

This research was supported in part by the LLNL Biofuels Scientific Focus Area, funded by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under FWP SCW1039. Work at LLNL was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344.

Iron storage in ferritin versus a lysosome-related compartment in the green alga *Chlamydomonas reinhardtii*

Anne Glaesener,[1] Stefan Schmollinger,[1*](schmolli@chem.ucla.edu); Marcus Miethke,[1] Allegra Aron,[2] Jefferson Chan,[2] April Kusowski,[3] Peter K. Weber,[4] Timothy L. Stemmler,[3] Christopher J. Chang,[2] Jennifer Pett-Ridge,[4] and **Sabeeha S. Merchant**[1]

[1]University of California, Los Angeles; [2]University of California, Berkeley; [3]Wayne State University, Detroit, MI; [4]Lawrence Livermore National Laboratory, Livermore, CA

URL: http://www.chem.ucla.edu/dept/Faculty/merchant/#research

Project goals: Iron is an element of crucial importance, iron-deficiency limits primary productivity on a global scale, in croplands as well as in the oceans. The photosynthetic apparatus is thereby a key link, because of its central role in light-driven carbon assimilation, its abundance and iron-dependence. The single-celled, eukaryotic green alga *Chlamydomonas reinhardtii* is an excellent model to study iron deficiency, with all the advantages of a microbial system and the components of the iron assimilation machinery identified. In eukaryotes, the predominant site for iron storage is either ferritin or a vacuole. While ferritin is more and more viewed as a dynamic reservoir or iron buffer, the vacuole recently gained more attention as an iron storage site. The goal of this project is to provide a cellular view of iron metabolism, by understanding the pathways and dynamics of iron distribution between different iron storage compartments, especially during metabolic transitions.

Abstract: Chlamydomonas requires a broad spectrum of metal cofactors to sustain its photosynthetic and respiratory capacities, and iron (Fe) is one of the major transition metals involved in these processes [1]. The demand for Fe is reflected by its intracellular quota, which in Chlamydomonas strongly depends on the lifestyle (phototrophic, heterotrophic or photoheterotrophic) and the available external Fe concentrations. Four stages of iron nutrition can be distinguished in Chlamydomonas: iron-limited (below 0.5µM in the growth media), iron-deficient (between 1-3 µM), iron-replete (5-30 µM) and iron excess (above 50 µM). While iron-limited cells are chlorotic and growth-inhibited, iron-deficient cells have no obvious phenotype, but already activated the full transcriptional response towards iron deficiency. Iron-replete and iron-excess cells have not activated this response, but can be distinguished by their intracellular iron content. Especially interesting regarding iron storing, is that the more abundant of the two ferritin proteins that are encoded in the Chlamydomonas genome, the chloroplast-targeted Fer1, shows an unusual accumulation pattern upon different external iron concentrations [2]. Fer1 preferentially accumulates in iron deficient and iron-limited cells, while it is less abundant in iron-replete and -excess conditions, when iron is abundantly present, suggesting a different role than Fe storage.

To better understand Fe homeostasis in this reference organism during acclimation to changing Fe bioavailability, we used a suite of complementary techniques including (HPLC)-ICP-MS, confocal microscopy with Fe(II)-specific fluorescent sensors, NanoSIMS isotopic imaging and XAS to visualize individual compartments of iron storage, while providing different concentrations of extracellular Fe, ranging from 0.1 µM Fe (iron-limited) to 200 µM Fe (iron-excess). We found that intracellular iron concentrations observed in these conditions vary from $3x10^7$ atoms/cell at the lowest concentrations up to $6x10^8$ atoms/cell in excess iron conditions. We isolated the soluble Fer1 fraction grown under different Fe concentrations by size exclusion chromatography and quantified both protein and Fe contents. Fe content / ferritin varies dramatically between the different conditions, and is especially increased in iron-excess conditions, where ferritin abundance is reduced but cellular Fe content is at its peak. In iron-limited and iron-deficient cells the major iron storage compartment is indeed ferritin, where most of the intracellular iron is found. To understand how accumulated Fe in excess conditions is stored at the individual cell level, we stained cells with a Fe(II)-specific fluorescent dye, and in the high iron treatment, found it appears to be sequestered in distinct subcellular compartments. Complementary NanoSIMS studies revealed a partial co-localization of Fe, Ca and P, suggesting that at higher external iron concentrations an additional iron storage compartment was identified, linked to a lysosome-related structure known as acidocalcisome [3]. Mutants lacking the acidocalcisome compartment were not capable of accumulating additional Fe in iron-excess conditions and didn't display the additional Fe sequestration sites. Using stable-isotope labelling (^{57}Fe), we were able to follow the kinetics of Fe enrichment in the different storage compartments during a shift from iron-limited conditions (^{56}Fe) to iron-excess conditions (^{57}Fe). During the initial hours of the shift, the ^{57}Fe label was strongly enriched in the Fer1 pool before it accumulated in the lysosome-related compartment and subsequently was depleted from the Fer1 pool. Thus, these studies allowed us to dissect the intracellular Fe distribution pathways both kinetically and spatially in different Fe regimes.

References

1. Blaby-Haas CE, Merchant SS (2013) Iron sparing and recycling in a compartmentalized cell. Curr Opin Microbiol 16: 677-685.
2. Long JC, Sommer F, Allen MD, Lu SF, Merchant SS (2008) FER1 and FER2 encoding two ferritin complexes in Chlamydomonas reinhardtii chloroplasts are regulated by iron. Genetics 179: 137-147.
3. Hong-Hermesdorf A, Miethke M, Gallaher SD, Kropat J, Dodani SC, et al. (2014) Subcellular metal imaging identifies dynamic sites of Cu accumulation in Chlamydomonas. Nat Chem Biol 10: 1034-1042.

This research was supported in part by the LLNL Biofuels Scientific Focus Area, funded by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under FWP SCW1039. Work at LLNL was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344.

Detecting Cysteine Modifications in methanogen *Methanosarcina mazei* Gö1

Rachel Ogorzalek Loo[1*] (rloo@mednet.ucla.edu), Phuong H.N. Nguyen[1,2], Hong Hanh Nguyen[1], **Joseph Loo**[1], and **Robert Gunsalus**[1]

[1]University of California-Los Angeles, Los Angeles, CA, USA; [2]University of Science, VNU-HCMUS, Vietnam

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

Project Goals: To elucidate the biological pathways of microbes relevant to microbial biofuel production and to global carbon cycling. These studies employ proteomics and mass spectrometry to characterize protein post-translational modifications.

Archaea in genus *Methanosarcina* are distributed broadly from marine to fresh water environments. They produce methane from a wide range of substrates including acetate, methylamines, and methanol and account for a large percentage of global methane emission. In methanogenesis, several important steps rely on thiol intermediates; *e.g.*, methyl transfer from tetrahydrosarcinopterin (H4SPT) to coenzyme M (mercaptoethanesulfonate), methane release by oxidation of coenzyme M and coenzyme B to form a heterodimer, and recycling of coenzymes M and B after reduction by heterodisulfide reductase. The importance of thiols to methanogenesis encouraged us to explore cysteine modifications in *Methanosarcina mazei.*

Tryptic peptides were generated with and without reduction/alkylation from cell lysates of *Methanosarcina* cultivated on methanol and on other carbon substrates. Peptides were analyzed by LC-MS/MS to identify proteins and to inventory post-translational modifications. Among the most abundant modifications observed was cysteinylation (Cys+119), identified on over 40 of proteins. Protein cysteinylation was observed not only from cultures maintaining reducing conditions with Na_2S/cysteine addition, but also from those supplementing with Na_2S only. Other modifications detected included Cys+30 (trisulfide in multi-cysteine peptides), Cys+140, Cys+151, and Cys+152. Modified cysteines appeared in active sites of some metabolic enzymes. The significance of these modifications is being explored.

This work was supported by the Department of Energy Office of Science (BER) through DE-FC-02-02ER63421 to the UCLA-DOE Institute.

Detecting Cysteine Modifications in methanogen *Methanosarcina mazei* Gö1

Rachel Ogorzalek Loo[1*] (rloo@mednet.ucla.edu), Phuong H.N. Nguyen[1,2], Hong Hanh Nguyen[1], **Todd Yeates**[1], Joseph Loo[1], and Robert Gunsalus[1]

[1]University of California-Los Angeles, Los Angeles, CA, USA; [2]University of Science, VNU-HCMUS, Vietnam

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

Project Goals: To elucidate the biological pathways of microbes relevant to microbial biofuel production and to global carbon cycling. These studies employ proteomics and mass spectrometry to characterize protein post-translational modifications.

Archaea in genus *Methanosarcina* are distributed broadly from marine to fresh water environments. They produce methane from a wide range of substrates including acetate, methylamines, and methanol and account for a large percentage of global methane emission. In methanogenesis, several important steps rely on thiol intermediates; *e.g.*, methyl transfer from tetrahydrosarcinopterin (H4SPT) to coenzyme M (mercaptoethanesulfonate), methane release by oxidation of coenzyme M and coenzyme B to form a heterodimer, and recycling of coenzymes M and B after reduction by heterodisulfide reductase. The importance of thiols to methanogenesis encouraged us to explore cysteine modifications in *Methanosarcina mazei.*

Tryptic peptides were generated with and without reduction/alkylation from cell lysates of *Methanosarcina* cultivated on methanol and on other carbon substrates. Peptides were analyzed by LC-MS/MS to identify proteins and to inventory post-translational modifications. Among the most abundant modifications observed was cysteinylation (Cys+119), identified on over 40 of proteins. Protein cysteinylation was observed not only from cultures maintaining reducing conditions with Na_2S/cysteine addition, but also from those supplementing with Na_2S only. Other modifications detected included Cys+30 (trisulfide in multi-cysteine peptides), Cys+140, Cys+151, and Cys+152. Modified cysteines appeared in active sites of some metabolic enzymes. The significance of these modifications is being explored.

This work was supported by the Department of Energy Office of Science (BER) through DE-FC02-02ER63421 to the UCLA-DOE Institute.

Phototroph-heterotroph symbiosis for biofuels applications

Stephanie Hays[3*] (sgracehays@gmail.com), Tingting Li,[1*] (tli46@jhu.edu) Emily F. Freed[2], Geng Yu,[1] Rahul Tevatia,[1] Chien-Ting Li,[1] Pamela A. Silver,[3,4] George A. Oyler,[1] Michael T. Guarnieri,[2] **Michael J. Betenbaugh[1]**

[1]Johns Hopkins University, Baltimore, MD; [2]National Bioenergy Center, National Renewable Energy Laboratory, Golden, CO; [3]Harvard Medical School, Boston, Massachusetts; [4]Harvard University, Boston, Massachusetts

Project Goals:

The goal of this study is to establish symbiotic pairs of photoautotrophic and heterotrophic microbes in order to explore their interactions for growth and for biosynthesis of products, including metabolites and biofuels, from carbon dioxide and sunlight.

In natural ecosystems, microbes live in heterogeneous populations in which individual members contribute different functions to enhance the overall performance of the group. In our study, a cyanobacterium, *Synechococcus elongatus* or the microalgae *Chlorella* capture light and carbon dioxide via photosynthesis. Sucrose secretion has been engineered into the cyanobacteria. The sucrose is then used as a direct feedstock for either bacterial growth or growth and lipid accumulation in multiple yeast species. As previously observed [1], sucrose secretion by *S. elongatus* is strongly dependent on light, osmotic pressure, and pH, which is consistent with the sucrose/proton symport activity of heterologously expressed sucrose permease. Thus, successful pairing of *S. elongatus* and heterotrophs requires robust heterotroph growth on sucrose in an appropriate medium. In this work, we first evaluated monocultures of different phototrophic and heterotrophic species by examining different environmental parameters to guarantee a well-adapted coculture system. Of the bacteria tested, we focus on *Escherichia coli* growth. Of the four yeast species of *Cryptococcus curvatus*, *Rhodotorula glutinis*, *Yarrowia lipolytica,* and *Saccharomyces cerevisiae* investigated, the former three showed great potential to adapt the co-culture system, while *S. cerevisiae* had a high sucrose level requirement to support its growth.

Next, co-culture of phototrophs (sucrose-secreting *S. elongatus* or microalgae) and heterotrophs (bacteria or yeast) was performed, and the effect of each partner on the expansion of the symbiont was evaluated by monitoring cell numbers, final biomass, lipid production, and other culture parameters. Many of the heterotrophs can grow and efficiently utilize sucrose produced by the partner *S. elongatus* in coculture, even when sucrose levels were low in the initial growth stage. Some co-culture systems were also successfully maintained in a semi-continuous culture system. Interestingly, in some cases, the growth of the cyanobacteria *S. elongatus* cells was more robust in co-culture compared with monoculture.

These artificially established symbioses strongly support that sucrose secreted by cyanobacteria can sustain eukaryotic cell expansion. Similarly, microalgae produce nutrients that can facilitate growth of heterotrophs. Our results help shed light on the establishment of a bioenergy platform, which would combine the metabolic capacity of photoautotrophs to capture inorganic carbon and then channel the resulting organic carbon directly to heterotroph partners for producing biofuel precursors.

REFERENCES

1. Ducat, D.C., et al., *Rerouting carbon flux to enhance photosynthetic productivity.* Applied and environmental microbiology, 2012. **78**(8): p. 2660-2668.

Supported by DE-SC0012658.

Systems Biology of Autotrophic-Heterotrophic Symbiosis for Biofuel Production

Cristal Zuñiga[1*] (crzuniga@eng.ucsd.edu), Chien-Ting Li,[2] Jennifer Levering,[1] George Oyler,[2] Brian O. McConnell[4], Christopher P. Long,[4] Michael Guarnieri,[3] Maciek R. Antoniewicz,[4] Karsten Zengler,[1] and **Michael Betenbaugh**[2]

[1]University of California, San Diego; [2]Johns Hopkins University, Baltimore, Maryland; [3]National Renewable Energy Laboratory, Golden, Colorado; [4]University of Delaware, Newark

Project goals: The goal of this work is to build and combine autotrophic and heterotrophic organisms as a novel sustainable symbiotic platform for growth and generation of lipids for biofuels. Genome-scale reconstructions of autotrophic-heterotrophic co-cultures will be deployed to decipher and predict metabolic interactions and will lay the foundation for optimizing system performance for growth and biofuel precursor production.

Organisms do not exist in isolation in the environment but form stable communities. For example, lichens form a well-established and beneficial relationship between two types of organisms; a phototroph, which uses light as energy source and a heterotroph, which utilizes organic material. In order to get insights into these natural communities, we are designing and constructing synthetic stable co-cultures experimentally and computationally. One of these co-cultures consists of the phototroph *Chlorella vulgaris* and the heterotroph *Saccharomyces cerevisiae*. Genome-scale metabolic models for *C. vulgaris* and *S. cerevisiae* are generated to delineate computationally possible metabolic interactions between these organisms in co-culture. A model for *S. cerevisiae* already exists, while the model for *C. vulgaris* has been recently reconstructed. Initial work involves the reconstruction, validation, and application of a genome-scale metabolic model for *C. vulgaris* UTEX 395, *i*CZ842. The reconstruction consists of six compartments: the cytoplasm, mitochondrion, chloroplast, thylakoid, glyoxysome, and the extracellular space. It contains 842 out of 7,100 annotated genes (around 12%), delineating 1,763 metabolites and 2,280 reactions. *C. vulgaris* can grow under different trophic conditions (e.g. photoauto-, hetero-, and mixotrophic). Each of these growth conditions is represented mathematically through different biomass objective functions (BOFs). Every equation contains the stoichiometric coefficients for the most important metabolites of the biomass. Here, the lipid, protein, carbohydrates and ribose (RNA) contents were determined experimentally. The reconstruction represents the most comprehensive model for any eukaryotic photosynthetic organism to date based on genome size and number of genes included in the reconstruction. The highly curated model was validated against experimental data.

The model accurately predicts growth rates under photoauto-, hetero-, and mixotrophy. Flux distributions under different trophic conditions show that not only central carbon metabolism but also amino acid, nucleotide, and pigment biosynthetic pathways are impacted when the microalgae is under nitrogen starvation. Furthermore, prediction of growth rates under various medium compositions using *i*CZ842 suggested an increased growth rate with the addition of tryptophan and methionine, which was experimentally verified. This effort lays the foundation for generation of synthetic co-culture models going forward. Initial work and issues to address for building and integrating complementary heterotrophic models will also be described.

Supported by DE-SC0012658.

Advanced Computational and Modeling Analyses Provide Novel Insights into Interactions in Model Complex Microbial Consortia

Hyun-Seob Song[1]*(hyunseob.song@pnnl.gov), Ronald C. Taylor[1], Pamela B. Weisenhorn[2], Christopher S. Henry[2], Joon-Yong Lee[1], Jeremy D. Zucker[1], Hans C. Bernstein[1], Yukari Maezato[1], Abigail E. Tucker[1], Margaret F. Romine[1], Jessica Cole[1], Stephen R. Lindemann[1], Lee Ann McCue[1], **Janet K. Jansson**[1]

[1]Pacific Northwest National Laboratory, Richland, WA; [2]Argonne National Laboratory, Argonne, IL

http://www.pnnl.gov/biology/programs/fsfa/

Project Goals: The PNNL FSFA goal is to identify the fundamental mechanisms by which microbial interactions and spatial organization impact rates and pathways of carbon and energy flow in microbial communities. The strategy involves the study of highly interactive and tractable model autotroph-heterotroph consortia whose member genome sequences have been defined. Our project leverages unique capabilities including multi-omics measurements, advanced functional imaging, taxonomic profiling and metabolic and regulatory network modeling to elucidate underlying reaction mechanisms within complex microbial communities. Our research plan supports DOE goals to achieve a predictive understanding of microbially-mediated carbon and energy transformation.

A mechanistic understanding of the complex interplay between member species in microbial communities is essential to enable predictions of community dynamics and response to environmental perturbations, and for eventual design and control of community-level functions. It is known that the populations of microorganisms in communities are balanced through positive, neutral, or negative interactions. However, we currently have insufficient fundamental knowledge of the types of metabolic and cellular interactions that control member populations and that regulate their functions. Microbial association networks of complex communities have primarily been inferred from species co-occurrence data using similarity- or regression-based approaches. However, network inference is an underdetermined inference problem. In addition, existing approaches cannot differentiate between direct and indirect interactions; thus microbial association inferences are often inconsistent across different methods. By contrast, community metabolic network analyses can predict microbial interactions based on predicted metabolite exchanges between species. Currently, the great majority of metabolic network modeling is limited to monocultures or very simple communities (mostly, binary cultures), primarily due to the difficulty in reconstructing reliable species-level and community metabolic networks from metagenome sequences.

Our hypothesis is that ***by combining interaction network inference and metabolic network analysis modeling approaches we can improve predictions of microbial community interactions***. First, we developed a mixed integer linear programming-based regression model that infers microbial interactions by fitting population growth models (such as generalized Lotka-Volterra [gLV] and evolutionary game theory-based models) to temporal profiles of species abundance. Unlike conventional analyses that yield one specific interaction network (despite uncertainties on direct versus indirect interactions), our modeling approach systematically

identifies *all* possible scenarios of microbial interactions that equally represent a given dataset. We demonstrate this new modeling approach using comprehensive simulated data as well as our own experimental data. For example, we used this approach to predict metabolic interactions in an experimentally tractable model autotroph-heterotroph community composed of 20 members. This approach enabled us to model and predict multiple scenarios of microbial interactions in the consortium. Specifically, we predicted a dramatic shift in species interactions during succession under an ammonium-amendment condition. These predictions serve as the basis for development of new testable hypotheses.

In addition, we are developing genome-scale metabolic network models for complex microbial consortia, pushing well beyond the boundaries established for typical mono- or binary cultures. To address this aim, we have integrated our genomic and transcriptomic data into the DOE Knowledgebase (KBase), and utilized the KBase Platform to construct single genome and community metabolic models and to predict interactions between the species comprising our model community. We are collaborating with the KBase team to supplement the standard KBase environment with additional methods, including integration of core and whole-genome models and gene expression profile-guided gap filling. These modules are currently being evaluated for deployment through KBase for use by the larger scientific community.

The next steps will be to use this modeling approach to reveal the mechanistic underpinnings of microbial interactions and division of labor that are responsible for transforming carbon and other nutrients in complex communities. Specific interactions will also serve as key inputs to advanced modeling platforms for integrating multi-omics data to predict temporal and spatial organization of complex communities.

This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL).

Spatial Interactions of Autotrophs and Heterotrophs Elucidated Using Advanced Quantitative Imaging Techniques

Ryan S. Renslow[1]* (ryan.renslow@pnnl.gov), Stephen R. Lindemann[1], Jessica K. Cole[1], Zihua Zhu[1], Christopher R. Anderton[1], Janine R. Hutchison[1], Young-Mo Kim[1], William B. Chrisler[1], Heather E. Engelmann[1], Alice C. Dohnalkova[1], Dehong Hu[1], Thomas O. Metz[1], H. Steven Wiley[1], Haluk Beyenal[2], Phuc T. Ha[2], Jim K. Fredrickson[1], and **Janet K. Jansson**[1]

[1]Pacific Northwest National Laboratory, Richland, WA; [2]Washington State University, Pullman, WA

http://www.pnl.gov/biology/programs/fsfa

Project Goals: The PNNL FSFA goal is to identify the fundamental mechanisms by which microbial interactions and spatial organization impact rates and pathways of carbon and energy flow in microbial communities. The strategy involves the study of highly interactive and tractable model autotroph-heterotroph consortia whose member genome sequences have been defined. Our project leverages unique capabilities including multi-omics measurements, advanced functional imaging, taxonomic profiling and metabolic and regulatory network modeling to elucidate underlying reaction mechanisms within complex microbial communities. Our research plan supports DOE goals to achieve a predictive understanding of microbially-mediated carbon and energy transformation.

Elucidating the molecular mechanisms by which individual microorganisms in microbial communities exchange metabolites is an important challenge for increasing our fundamental understanding of the roles of specific microbial interactions in key biogeochemical cycles. An additional scientific challenge is to understand the importance of spatial distance between community members because natural microbial communities are spatially heterogeneous and the means by which microbes interact across various spatial gradients is not well understood. Here we addressed the hypothesis that autotrophs serve as the sole nitrogen source for heterotrophs when only nitrate is available (see Lindemann et al. poster). A second hypothesis is that certain heterotrophs would spatially associate with the autotroph, due to the need for both carbon and nitrogen sharing. Thus, providing a soluble nitrogen source should disrupt the spatial linkages by removing obligatory interactions. To test these hypotheses we employed a suite of advanced imaging capabilities together with novel quantitative image analysis techniques. For the experiments, we leveraged a tractable model consortium in which a single photoautotroph (a cyanobacterium) provides the fixed carbon for a suite of naturally, co-isolated heterotrophs. We used this model system to decipher the spatial context of interactions occurring between members. In particular, the unicyanobacterial consortium, UCC-O, was a valuable resource in this respect because most of the members have been isolated and genome sequences are available, facilitating the use of genome-enabled approaches for deriving metabolic models.

Dual-channel confocal laser scanning microscopy (CLSM) was used to image the community level spatial organization and dynamics of autotrophs and heterotrophs. The morphology of cells in actively growing consortial biofilms was investigated by analyzing the CLSM micrographs using custom developed image analysis software that could quantify total biomass and the biomass distribution between the autotroph and heterotrophs[1]. Nanoscale secondary ion mass spectrometry (NanoSIMS), a powerful, high spatial resolution imaging approach, was then used

to visualize the metabolic activities of single cells. We created and applied a semi-automated image-processing pipeline for quantitative analysis of the NanoSIMS data[2]. This pipeline includes both elemental and morphological segmentation, thus producing a final segmented image that allows for discrimination between autotrophic and heterotrophic biomass. It also allows the detection of individual cyanobacterial filaments and heterotrophic cells, the quantification of isotopic incorporation of individual heterotrophic cells, and calculation of relevant population statistics. Fluorescence in situ hybridization (FISH) imaging with species-specific probes was used to spatially localize heterotroph species that were hypothesized to be dependent on metabolic linkages with the cyanobacterium based on genome predictions.

The combined use of these imaging tools was applied to analyze the community structure over time and to understand the uptake of ^{15}N provided as either nitrate or ammonium through the consortium. We found that the degree of ^{15}N incorporation by individual cells was highly variable when labeled with $^{15}NH_4+$, but much more even when biofilms were labeled with $^{15}NO_3-$. In the $^{15}NH_4+$-amended biofilms, the heterotrophic distribution of ^{15}N incorporation was highly skewed, with a large population showing moderate ^{15}N incorporation and a small number of cells displaying very high ^{15}N uptake. We also observed that one of the heterotrophs (HL-53) was always proximal to the cyanobacterial filaments, suggesting a new hypothesis that these cells form an epibiotic or parasitic relationship. We are currently performing additional experiments to test this. Our results show the feasibility of using imaging technologies for the quantitation and visualization of the spatial organization of individual cells within a community, their uptake of specific elements, and their orientation to other member species. These techniques will be employed in future experiments to gain further understanding of the spatial dependencies of microbial community interactions.

References

1. Jessica K. Cole, Janine R. Hutchison, Ryan S. Renslow, Young-Mo Kim, William B. Chrisler, Heather E. Engelmann, Alice C. Dohnalkova, Dehong Hu, Thomas O. Metz, Jim K. Fredrickson and Stephen R. Lindemann. *Phototrophic biofilm assembly in microbial-mat-derived unicyanobacterial consortia: model systems for the study of autotroph-heterotroph interactions*. Frontiers in Microbiology. 2014. 5.
2. Renslow RS, Lindemann SR, Cole JK, Zihua Z, and Anderton CR. *Quantifying element incorporation in multispecies biofilms using nanoscale secondary ion mass spectrometry (NanoSIMS) image analysis*. Biointerphases. 2016. In Print.
3. Renslow RS, Lewandowski Z, and Beyenal H. *Biofilm image reconstruction for assessing structural parameters.* Biotechnology and bioengineering. 2011. 108 (6), 1383-1394.
4. Resat H, Renslow RS, and Beyenal H. *Reconstruction of biofilm images: combining local and global structural parameters.* Biofouling. 2014. 30 (9), 1141-1154.
5. Ha PT, Renslow RS, Atci E, Reardon PN, Lindemann SR, Fredrickson JK, Call DR, and Beyenal H. *Regulation of electron transfer processes affects phototrophic mat structure and activity. Frontiers in microbiology.* 2015. 6.
6. Moran JJ, Doll CG, Bernstein HC, Renslow RS, Cory AB, Hutchison JR, Lindemann SR, and Fredrickson JK. *Spatially tracking 13C-labelled substrate (bicarbonate) accumulation in microbial communities using laser ablation isotope ratio mass spectrometry*. Environmental microbiology reports. 2014. 6 (6) 186-791.

This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL). A portion of this work (proteomics) was performed in the William R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by OBER and located at PNNL. RSR was partially supported by a Linus Pauling Distinguished Postdoctoral Fellowship at PNNL. This work was made possible in part by the OpenMIMS software whose development is funded by the NIH/NIBIB National Resource for Imaging Mass Spectrometry, NIH/NIBIB 5P41 EB001974-10.

Spatially-Resolved Carbon and Energy Transformations in a Vertically Laminated, Phototrophic Mat

James Moran* (james.moran@pnnl.gov), Hans Bernstein, Mary Lipton, Amy Boaro, Colin Brislawn, Krystin Riha, Karl Dana, Angela Melville, Eric Huang, Jennifer Mobberly, William Nelson, James Fredrickson, and **Janet K. Jansson**

Pacific Northwest National Laboratory, Richland, WA

http://www.pnnl.gov/biology/programs/fsfa/

Project Goals: The PNNL FSFA goal is to identify the fundamental mechanisms by which microbial interactions and spatial organization impact rates and pathways of carbon and energy flow in microbial communities. The strategy involves the study of highly interactive and tractable model autotroph-heterotroph consortia whose member genome sequences have been defined. Our project leverages unique capabilities including multi-omics measurements, advanced functional imaging, taxonomic profiling and metabolic and regulatory network modeling to elucidate underlying reaction mechanisms within complex microbial communities. Our research plan supports DOE goals to achieve a predictive understanding of microbially-mediated carbon and energy transformation.

The localization of carbon (C) fixation in vertically laminated phototrophic mat systems is largely controlled by photon accessibility and generally relegated to only an upper portion of the overall community. Light attenuation through the mat can produce 'dark' regions (negligible photosynthetically active radiation) below the photic zone, thereby precluding photoautotrophy from these laminae. Lack of *in situ* photoautotrophy below the photic zone means that the resident heterotrophs are reliant on the photosynthetic activity of upper mat layers and C transport processes to deliver fixed C deeper into the mat. **This work describes our team's progress toward understanding the spatial controls on C and energy transformations in complex microbial mat communities.** Our studies are guided by two central hypotheses:

1. *Diel dynamics drive the localization and dispersal of fixed C within a photosynthetically driven system*
2. *Species richness in spatially laminated, photoautotrophically supported communities is decoupled from light energy input*

To address these hypotheses, we used stable isotope labeled (^{13}C) bicarbonate to track C fixation and accumulation into biomass within a phototrophic mat system under two sets of conditions: 1) field-based, stationary incubations of intact mat, using *in situ* medium and light (diel) conditions, and 2) laboratory flow-through incubations containing artificial medium and constant light intensity.

In the field-based incubations, we assessed C accumulation and migration over a complete natural diel cycle. Spatially resolved analysis (Figure 1) via laser ablation isotope ratio mass spectrometry (LA-IRMS) provided localized quantification of C uptake (day) and partial loss (night). We confirmed our hypothesis that net photoautotrophy (as recorded by accumulation of ^{13}C) was highly localized near the upper portions of the mat, above the maximum photic depth. C penetration below the photic zone progressed consistently during daylight, indicating that C slowly migrated while C was being actively fixed, versus a rapid spatial migration of C after a diel shift (i.e. light-to-dark). Proteomic analysis of isotopically labeled peptides was used to assess the degree to which photosynthetic products were transferred to associated heterotrophic community members. In an expanded set of field-based incubations, we used isotopically labeled (^{13}C) glucose and acetate as substrates to differentiate heterotrophic from autotrophic consumption. Proteins associated with autotrophs (i.e. cyanobacteria) dominated the labeled fraction for all substrates (bicarbonate, glucose, and acetate) but the bicarbonate-

derived ^{13}C labeled peptides from heterotrophs were used to identify trophic C exchange within the community. By clustering the labeled peptides based on their changing abundance over the diel cycle, we identified a disproportionately large increase in proteins incorporating acetate-derived C at night and bicarbonate and glucose -derived C during the day.

Similar C accumulation profiles were observed when the mat was incubated in a flow-through, laboratory-based system with (^{13}C) bicarbonate; namely a distinctive focusing of bicarbonate-derived C near the mat's surface. Shallower penetration of freshly fixed C into the mat profile was consistent with the reduced incident irradiance that was used in the laboratory versus the field. When the mat samples were exposed to constant light conditions, there was persistent migration of fixed C to deeper layers of the mat over the course of the experiment. In addition to measuring autotrophy rates, we used oxygen microsensors to determine *in situ* photosynthetic rates (both net and gross) and used 16S rRNA gene sequencing to determine the microbial diversity in cryosectioned mat laminae. We found an inverse relationship between community diversity and photosynthetic rates (Figure 2) with regions of highest O_2 production (near the mat surface) also having the lowest diversity based on number of observed OTUs.

Our studies of the spatial relationships between C and energy acquisition within photosynthetically driven microbial communities is helping to reveal basic principles about complex photoautotroph:heterotroph interactions. Together these data reveal new candidate principles to be validated with further experimentation:

1. Diel light cycling drives the timing of photosynthetically fixed C migration through the community and its exchange between phototrophs and heterotrophs.
2. Light energy input in spatially laminated, photosynthetically driven systems can constrain species richness.

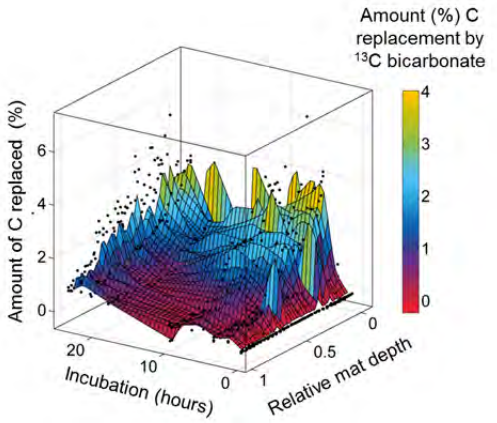


Figure 1: Spatially resolved isotope analysis to identify regions of label accumulation. We used LA-IRMS to pinpoint regions showing the highest accumulation of substrate-derived C (percent of native biomass replaced by the added substrate). Bicarbonate-derived C is predominately focused near the mat's surface (relative depth = 0) and displays a diel influx and efflux of photosynthetically derived C. Penetration of C below the photic depth is most prominent during the day suggesting a constant, depth-ward migration of C versus a stepwise transition linked to shifts in diel light regimes.

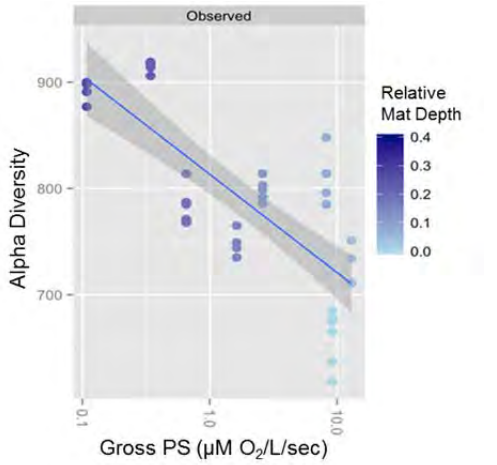


Figure 2: Depth-resolved microbial diversity (based on number of distinct OTUs) showed an inverse relationship with photosynthetic rates (both gross and net rates with gross shown above).

*This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL). A portion of this work (*proteomics*) was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by OBER and located at PNNL.*

Autotroph-Heterotroph Communities Increase Interactions and Metabolite Exchange as They Undergo Ecological Succession

William C. Nelson* (William.Nelson@pnnl.gov), Stephen R. Lindemann, Jessica K. Cole, Eric Huang, Jason Xu, Yukari Maezato, Jennifer Mobberley, Beau Morton, Ronald A. Taylor, Stephen Wiley, Mary S. Lipton, Thomas Metz, Margaret F. Romine and **Janet K. Jansson**

Pacific Northwest National Laboratory, Richland, WA

http://www.pnl.gov/biology/programs/fsfa

Project Goals: The PNNL FSFA goal is to identify the fundamental mechanisms by which microbial interactions and spatial organization impact rates and pathways of carbon and energy flow in microbial communities. The strategy involves the study of highly interactive and tractable model autotroph-heterotroph consortia whose member genome sequences have been defined. Our project leverages unique capabilities including multi-omics measurements, advanced functional imaging, taxonomic profiling and metabolic and regulatory network modeling to elucidate underlying reaction mechanisms within complex microbial communities. Our research plan supports DOE goals to achieve a predictive understanding of microbially-mediated carbon and energy transformation.

A major barrier to investigating microbial community function and dynamics is the high organismal diversity found in most environments. To address this challenge, we have derived stable consortia of reduced complexity from benthic microbial mats, by sequential enrichment on defined media under constant light[1]. Two of the resulting consortia, referred to as unicyanobacterial consortia (UCCs), each include a single cyanobacterium and up to 18 heterotrophs. Complete or near-complete genome sequences from these consortia were generated using a combination of metagenome and isolate genome sequencing[2]. Comparison of the metagenomic sequences to the assembled genome sequences uncovered microdiversity within the community, both in the form of distinct species conflated within OTUs and genomic microheterogeneity within species.

These consortia are a unique and powerful model system for the study of microbial interactions because of: 1) their tractable richness, 2) stable community structure and reproducible ecological succession, 3) a defined growth medium enabling detailed analyses of metabolic interdependencies, 4) axenic isolates of most members, 5) available genetic systems for most of the isolates, 6) the presence of microdiversity and 6) complete genomic information. We have used these consortia to examine microbial community dynamics and mechanisms associated with division of labor among the community members.

To test the hypothesis that ***microbial interactions increase over community ecological succession***, we applied a multi-omics approach to 28-day succession cycles of two different UCC consortia, each with a different cyanobacterium, and nearly identical heterotrophic membership. Reproducible changes in the relative abundances of populations over time were observed, with distinct succession patterns between the consortia, suggesting that the two consortia display different photoautotroph-driven carbon and energy fluxes. Analysis of metatranscriptomics data for nitrogen metabolism genes suggested that nitrogen acquisition shifted from an initial phase dominated by incorporation of inorganic nitrogen (nitrate in the

growth medium) to a recycling phase in which organic resources were exchanged. Metaproteomics data supported this observation, showing a decrease in expression of the assimilatory nitrite reduction protein NirA and a concomitant increase in the expression of the ammonium transporter Amt, the urea transporter UreT and urease UreC. Conversely, sulfur metabolism genes had stable expression across succession, suggesting that sulfur metabolites were either retained or remineralized, leading to a consistent requirement for acquisition of inorganic sulfur.

We also quantified extracellular amino acid levels during the succession experiments. This metabolite analysis required the development of a novel suite of methods for the isolation of polar metabolites from hypersaline media. The workflow, called MetFish, is based on chemoselective enrichment in which dansyl or related probes are used to derivatize metabolites *in situ*, followed by their isolation and analysis. In this study, dansyl chloride was used to isolate amine-containing metabolites. Using MetFish, we found that the concentrations of extracellular amino acids peaked early in succession, declining thereafter. These data suggest that the efficiency of recycling of nitrogen-containing metabolites increases as the community matures.

Together our observations suggest that early in community formation, microorganisms must act more independently and acquire nutrients from inorganic sources. As succession progresses, specific interactions are established, likely in concert with the development of spatial structure, and metabolite exchange dominates nutrient and energy fluxes. Moving forward, we plan to continue to investigate mechanisms of interaction, explore the relationship between spatial arrangement and cooperative metabolism, and test the hypothesis that microdiversity stabilizes community function to environmental change through the presence of subpopulations with similar functional profiles but differing optimal growth conditions.

References

1. Cole, J.K, Hutchison, J.R., Renslow, R.S., Kim, Y., Chrisler, W.B., Engelmann, H.E., Dohnalkova, A.C., Hu, D., Metz, T.O., Fredrickson, J.F., and Lindemann, S.R. *Front in Microbiol.* 2014. 5:109.
2. Nelson, W.C., Y. Maezato, Y.W. Wu, M.F. Romine, and S.R. Lindemann. *Appl. Environ. Microbiol*. 2015. 82(1):255:267.

This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genome Sciences Program (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory. The work conducted by the DOE Joint Genome Institute was supported by the Office of Science of DOE under Contract No. DE-AC02-05CH11231 and Community Sequencing Project 701.

Deciphering Microbial Community Interactions Using Model Complex Microbial Consortia

Mary S. Lipton* (Mary.Lipton@pnnl.gov), Alex S. Beliaev, Hans C. Bernstein, Jim K. Fredrickson, Stephen R. Lindemann, Matthew J. Marshall, Lee Ann McCue, Jason E. McDermott, Thomas O. Metz, James J. Moran, William C. Nelson, Ryan S. Renslow, Margaret F. Romine, Hyun-Seob Song, Steven Wiley, Aaron T. Wright, and **Janet K. Jansson**

Pacific Northwest National Laboratory, Richland, WA

http://www.pnnl.gov/biology/programs/fsfa/

Project Goals: The PNNL FSFA goal is to identify the fundamental mechanisms by which microbial interactions and spatial organization impact rates and pathways of carbon and energy flow in microbial communities. The strategy involves the study of highly interactive and tractable model autotroph-heterotroph consortia whose member genome sequences have been defined. Our project leverages unique capabilities including multi-omics measurements, advanced functional imaging, taxonomic profiling and metabolic and regulatory network modeling to elucidate underlying reaction mechanisms within complex microbial communities. Our research plan supports DOE goals to achieve a predictive understanding of microbially-mediated carbon and energy transformation.

Microbial autotroph-heterotroph interactions influence biogeochemical cycles on a global scale, but the molecular mechanisms underlying microbial community interactions and functional processes they perform are largely unknown. The diversity and complexity of natural systems and their intractability to *in situ* manipulation make it challenging to elucidate the principles governing these interactions in natural systems. We have addressed this challenge by development of tractable model microbial autotroph-heterotroph consortia as experimental platforms. We apply a genome-enabled approach together with integrated experimental and modeling to paint a cohesive picture of fundamental mechanisms of microbial community interactions, metabolic interdependencies and spatial constraints on functional processes and community properties.

The majority of our experiments have focused on the study of interactions in consortia consisting of one cyanobacterium and up to 18 associated heterotrophs. These unicyanobacterial consortia (UCC) have proven to be a unique and valuable resource for the team and are becoming recognized as the "***E. coli of complex communities***" because of the following attributes: 1) the UCC consist of a primary producer and its heterotrophic epibionts that has been stably maintained in simple defined medium, and thus must employ a complex interaction network to support all its members, 2) complete or near-complete genome sequence from nearly every member has been derived enabling the prediction of member metabolic interdependencies, testing the effect of member microdiversity, and application of transcriptomic and proteomic methodologies to construct regulatory networks, 3) axenic cultures have been developed and sequenced, filling in gaps in member sequence and allowing us to manipulate community membership, and 4) genetic systems have developed for isolates allowing us to manipulate community function. We have also developed new methods for study of communities: 1) assays to measure metabolites in high salt conditions, 2) software for quantitative analysis of cellular relationships from images, 3) chemical probes to identify metabolite- protein interactions and image substrate uptake, and 4) protein SIP to determine the fate of metabolites in communities.

Several multidisciplinary studies addressing specific hypotheses within our model communities, using novel applied imaging, metabolomic, and other spatiotemporal profiling capabilities lead to new fundamental knowledge about the realm of interspecies interactions guiding complex community functions. The main findings are briefly summarized in this FSFA overview poster as follows:

Poster by Nelson *et al*. To test the hypothesis that microbial interactions increase over community

Extensive Interdependencies and Unexpected Roles for B Vitamins in Microbial Communities

Aaron T. Wright[1*] (Aaron.Wright@pnnl.gov), Margaret F. Romine, Premchendar Nandhikonda[1], Yukari Maezato[1], Lindsey N. Anderson[1], Stephen R. Lindemann[1], Mary S. Lipton[1], Young-Mo Kim[1], Thomas O. Metz[1], Chengdong Xu[1], Dmitry Rodionov[2], Irina Rodionova[2], and **Janet K. Jansson**[1]

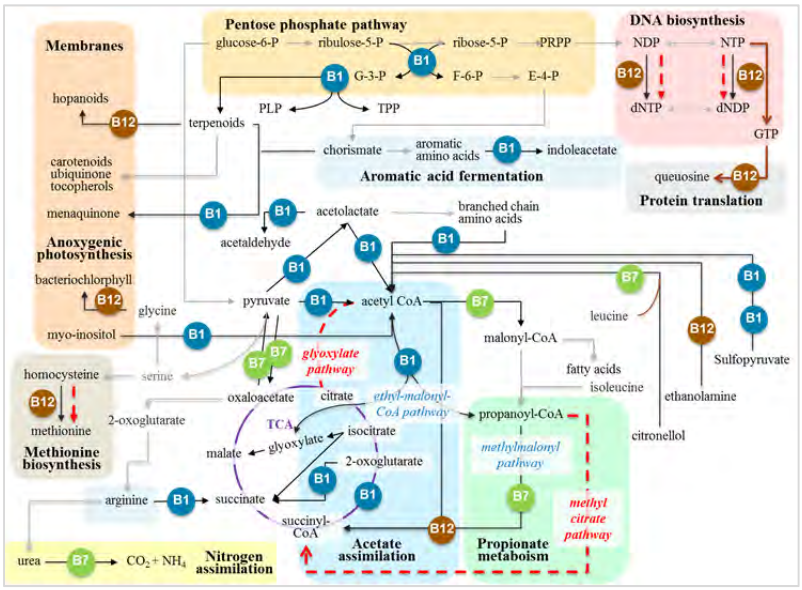
[1]Pacific Northwest National Laboratory, Richland, WA; [2]Sanford-Burnham Medical Research Institute, La Jolla, CA

http://www.pnnl.gov/biology/programs/fsfa/

Project Goals: The PNNL FSFA goal is to identify the fundamental mechanisms by which microbial interactions and spatial organization impact rates and pathways of carbon and energy flow in microbial communities. The strategy involves the study of highly interactive and tractable model autotroph-heterotroph consortia whose member genome sequences have been defined. Our project leverages unique capabilities including multi-omics measurements, advanced functional imaging, taxonomic profiling and metabolic and regulatory network modeling to elucidate underlying reaction mechanisms within complex microbial communities. Our research plan supports DOE goals to achieve a predictive understanding of microbially-mediated carbon and energy transformation.

Individual members of complex microbial communities can interact by exchanging metabolites and signaling compounds. However, the interdependency of species interactions in communities has only begun to be explored. To address this challenge we have focused our studies of microbial interactions on model, tractable autotroph-heterotroph consortia comprised of one cyanobacterium and associated heterotrophs (UCCs, Unicyanobacterial Consortia). These model consortia are valuable resources because they are stable and tractable, and have undergone extensive genome reconstruction and characterization. Our hypothesis was that ***specific members of the community would be dependent on other members for essential metabolites that they could not synthesize themselves***. This could be predicted from the genome information by determining whether each member possessed the ability to synthesize specific required nutrients, or not. Our genome reconstruction work suggested that extensive B vitamin interdependencies could exist among community members (see Figure below), and that these dependencies could govern community-coordinated carbon and energy cycling. Additionally, the requirement for vitamin exchange to support growth and possible regulatory roles in our model UCC suggests a new hypothesis to test. Namely, ***that the phototrophic producers could modulate vitamin availability to control member abundances and coordinate overall community metabolism***.

To test these hypotheses, we performed a combination of chemical probe profiling, regulon analysis, metabolomics, imaging, and microbial genetics experiments. We have previously developed and employed chemical profiling probes as B vitamin mimics to identify the function and specificity of a wide range of experimentally unidentified and/or predicted membrane-embedded B vitamin transporters, and also to directly characterize intracellular enzyme-cofactor associations in living microbial systems.[1] Here, we developed an activity based probe (ABP) to mimic vitamin B_{12}, which is produced by only four members of a UCC model community that contains >18 species. The probe was tested for its ability to capture B_{12}-binding proteins



expressed by *Halomonas* sp. HL-48, an isolate from the consortium that is capable of both synthesizing and salvaging B_{12}. Proteomic analysis revealed that the probe captured a total of 45 proteins. Components of all three expected B_{12}-dependent enzymes were among those detected. A review of the remaining proteins, suggested that B_{12} acts as an allosteric regulator of enzyme activity, a finding that has not previously been observed in any organism. Three of the captured proteins are involved in porphyrin biosynthesis; one at the branch point between heme and B_{12} biosynthesis, another associated with cobalt insertion into B_{12}, and the last likely involved in salvage of the B_{12} precursor, cobinamide. Allosteric effects at these positions could result in redirection of metabolism between biosynthesis of heme versus B_{12} as well as between B_{12} biosynthesis versus salvage. An additional 17 proteins could be linked to methionine recycling or C1 metabolism via folate. Folate is required for biosynthesis of proteins and nucleotides and thus required in high amounts during cell division. We also identified a new B_{12}-mediated regulator, designated PhrR, and performed regulon analysis to identify likely DNA promoters controlled by PhrR. Significantly, the genes under control of PhrR encode functions associated with light damage repair, chemotaxis, and biosynthesis of folate and ubiquinone. Metabolomic analysis of wild type and a PhrR knockout mutant validated a role for PhrR and B_{12} in controlling folate levels.

We then translated our B_{12} probe analyses and regulon analysis to determine that the five phototrophic species in the model community: *Rhodobacteriaceae*, *Algoriphagus*, and *Halomonas* species, utilize B_{12} as a light sensing ligand for transcriptional control of genes required for biosynthesis of light harvesting apparatus. These genes include those encoding production of bacteriochlorophyll, carotenoids, folate, and queousine as well as enzymes involved in DNA damage repair. We are now evaluating a new hypothesis that ***vitamin interdependency is important for establishing community stability***. We are currently investigating this hypothesis by testing mechanisms of vitamin exchange and evaluating the impact of vitamin supplementation on community population dynamics and metabolism.

References

1. Anderson, L.N., P.K. Koech, A.E. Plymale, E.V. Landorf, A. Konopka, F.R. Collart, M.S. Lipton, M.F. Romine, and A.T. Wright. *ACS Chem Biol*. **2016**. Epub ahead of print.

This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory. Proteomic analyses were performed in the William R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by OBER and located at PNNL.

ecological succession, we used a multi-omics approach, including a novel metabolomics methodology, to examine community composition and function over a 28-day growth of the UCCs. We found evidence of increases in efficiency of nitrogen species recycling as the community matures.

Poster by Wright *et al*. Coordination of metabolism can occur via vitamin exchange leading to a microbial division of labor. Because vitamins are synthetically expensive, vitamin sharing should conserve energy for other tasks. Importantly, we demonstrated that B_{12} utilization is widespread throughout the entire community, yet production is limited to only a few species, thus the exchange process is a necessary interaction for community success.

Poster by Lindemann *et al*. It has been observed that N affects carbon uptake and partitioning in a phototrophic dominated system, though the mechanisms driving these phenomena are not understood. We hypothesized that the cyanobacteria in the consortia served as the major primary assimilators and providers of community N, and found N availability to microbial photoautotrophs governs community C partitioning.

Poster by Renslow *et al*. Communities are typically spatially heterogeneous and the means by which microbes interact across various environmental gradients is not well understood. Here we developed a novel imaging method applied to the UCC's to decipher the spatial context of interactions occurring between members. We found the carbon based trophic relationships are dynamic and show a sharp dependence on N availability.

Poster by Moran *et al*. While photosynthetic carbon (C) fixation in vertically laminated phototrophic mat systems is generally relegated to only an upper portion of the overall community, microbial interactions must govern the availability of C to the rest of the community. We have tracked stable isotope-labeled (^{13}C) bicarbonate, acetate and glucose through the biomass of a phototrophic mat system. We identified that the community regions with highest productivity did not correlate with broadest biodiversity but that, rather, high autotrophy correlated with low biodiversity.

Poster by Song *et al*. Two complementary modeling efforts are focused on inferring microbial interactions in our model communities. We have integrated our genome and transcriptomic data into the DOE's Systems Biology Knowledgebase (KBase) platform to construct single genome and community metabolic network models as an *in silico* tool for predicting interspecies metabolite exchanges. We found coordinated and highly correlated transcriptional responses between species that could be used to infer specific interactions, particularly those that regulated synthesis and exchange of resources, such as amino acids, sugars and organic acids and organic C sources.

Poster by Bernstein *et al*. Specific mechanisms that mediate metabolic coupling and acclimation to microbial partnerships centered on C-N exchanges are not well defined. Transcriptome measurements coupled with metabolic modeling of a constructed binary phototroph-heterotroph consortium found coordinated and highly correlated transcriptional responses between species that could be used to infer specific interactions, particularly those that regulated synthesis and exchange of resources, such as amino acids, sugars and organic acids and organic C sources.

This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL). A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by OBER and located at PNNL.

Nitrogen Dynamics Control Carbon Partitioning in Model Complex Microbial Consortia

Stephen R. Lindemann* (stephen.lindemann@pnnl.gov), Jessica K. Cole, Christopher R. Anderton, Jennifer M. Mobberley, Amy A. Boaro, Alexandra Cory, Beau R. Morton, James J. Moran, Mary S. Lipton, William C. Nelson, Margaret F. Romine, and **Janet K. Jansson**

Pacific Northwest National Laboratory, Richland, WA

http://www.pnnl.gov/biology/programs/fsfa/

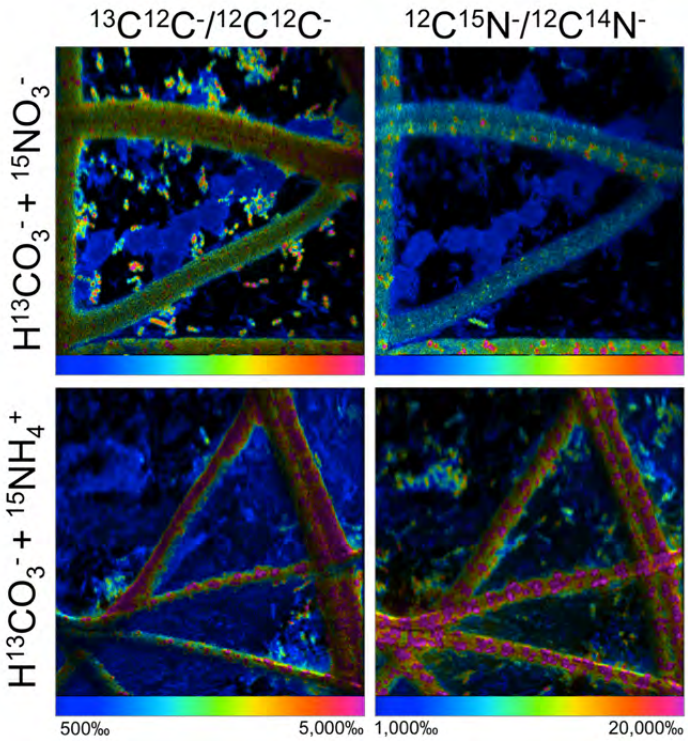
Project Goals: The PNNL FSFA goal is to identify the fundamental mechanisms by which microbial interactions and spatial organization impact rates and pathways of carbon and energy flow in microbial communities. The strategy involves the study of highly interactive and tractable model autotroph-heterotroph consortia whose member genome sequences have been defined. Our project leverages unique capabilities including multi-omics measurements, advanced functional imaging, taxonomic profiling and metabolic and regulatory network modeling to elucidate underlying reaction mechanisms within complex microbial communities. Our research plan supports DOE goals to achieve a predictive understanding of microbially-mediated carbon and energy transformation.

Interactions between microbial autotrophs and heterotrophs exert globally significant impacts on the cycling of carbon and energy, as well as other nutrients (e.g. nitrogen, phosphorus, sulfur). In order to predict the consequences of changing environmental conditions on autotroph-heterotroph interactions, it is first necessary to understand the fundamental principles driving their community dynamics. Here we investigated the principles governing interactions between phototrophs and heterotrophs in a model unicyanobacterial consortium (UCC), which contains a single cyanobacterium and co-isolated heterotrophs[1,2]. Species-resolved metagenome reconstruction[3] of two consortia revealed that most member species within the consortia were unable to directly assimilate nitrate as a nitrogen source, despite the fact that nitrate was the sole nitrogen source. Our aim was to elucidate the key metabolic interactions within the consortium that enabled relatively few member organisms (those that can consume nitrate) to supply the nitrogen requirements of the other members.

Because of the high energy cost of nitrate reduction to ammonium, ***we hypothesized that the cyanobacteria in the consortia served as the major primary assimilators and providers of community nitrogen***. To test this hypothesis, we investigated member macronutrient acquisition over a 28-day succession cycle of the consortium UCC-O, which contains the cyanobacterium *Phormidium* sp. OSCR and 19 associated heterotrophs, using an integrated transcriptomic and proteomic approach. These analyses revealed that cyanobacterial nitrate assimilation was initially high, but was supplanted late in succession by apparent recycling of reduced nitrogen species (e.g. ammonium, urea). However, cyanobacterial expression of proteins involved in ammonium incorporation into amino acids (i.e., glutamine synthetase, GlnA) was constant. Notably, heterotrophic species responded divergently in their expression of glutamine synthetase and ammonium transporters, revealing species-specific strategies to acquire reduced N within the community.

Subsequently, we tested the hypothesis ***that reliance upon the cyanobacterium for access to bioavailable N resulted in heterotrophic N limitation*** by amending UCC-O with reduced N (NH_4^+) at inoculation. This amendment drastically altered the community dynamics compared to NO_3^--only cultures, but in ways that were not readily explainable from the heterotrophs' predicted ability to reduce NO_3^-. Although there was no change in total carbon uptake rates between the two conditions, spatially-resolved analysis of ^{13}C and ^{15}N flow through the biofilm via nanoSIMS suggested that the cyanobacterium shared carbon with its heterotrophic partners much more rapidly when NO_3^- was the sole

nitrogen source than when amended with NH_4^+ (Fig. 1). Proteomic investigation revealed significant differences in proteins involved in cyanobacterial glycolysis between the two conditions, suggesting that kinetic limitation in NO_3^- uptake resulted in release of fixed C in the photosynthate to maintain redox balance. The protein data also revealed that lactate dehydrogenase and pyruvate-formate lyase were more highly expressed when NH_4^+ was absent, presumably as routes for shunting fixed C away from the TCA cycle via export of formate and/or lactate. Interestingly, these pathways for organic acid production are typically anaerobic processes. Together these observations suggest two new candidate principles governing autotroph-heterotroph interactions in communities: (1) N availability to microbial photoautotrophs governs community C partitioning, and (2) Spatial organization around hypoxic pockets in phototrophic biofilms enables anaerobic autotroph-heterotroph interactions that balance community redox and increase total productivity. Future research will test the generality of these principles using a broader array of consortia.



References

1. Lindemann, S. R. *et al.* The epsomitic phototrophic microbial mat of Hot Lake, Washington: community structural responses to seasonal cycling. *Microb. Physiol. Metab.* **4,** 323 (2013).
2. Cole, J. K. *et al.* Phototrophic biofilm assembly in microbial-mat-derived unicyanobacterial consortia: model systems for the study of autotroph-heterotroph interactions. *Syst. Microbiol.* **5,** 109 (2014).
3. Nelson, W. C., Maezato, Y., Wu, Y.-W., Romine, M. F. & Lindemann, S. R. Identification and resolution of microdiversity through metagenomic sequencing of parallel consortia. *Appl. Environ. Microbiol.* AEM.02274–15 (2015). doi:10.1128/AEM.02274-15.

This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (OBER), as part of BER's Genomic Science Program (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL). A portion of this work (transcriptomics, proteomics, and nanoSIMS) was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by OBER and located at PNNL. The work conducted by the U.S. Department of Energy Joint Genome Institute (JGI) was supported by the Office of Science of DOE under Contract No. DE-AC02-05CH11231 and Community Sequencing Project 701.

Cross-Talk Through Microbial Interaction Coordinates Microbial Community Properties

Hans C. Bernstein[1]*(hans.bernstein@pnnl.gov), Hyun-Seob Song[1], Ryan McClure[1], William Chrisler[1], Ronald C. Taylor[1], Pamela B. Weisenhorn[2], Christopher S. Henry[2],Alexander Beliaev[1], Margaret Romine[1], Thomas Metz[1], Young-Mo Kim[1], Jason McDermott[1], James Fredrickson[1], **Janet Jansson[1]**

[1]Pacific Northwest National Laboratory, Richland, WA; [2]Argonne National Laboratory, Argonne, IL

http://www.pnnl.gov/biology/programs/fsfa/

Project Goals: The PNNL FSFA goal is to identify the fundamental mechanisms by which microbial interactions and spatial organization impact rates and pathways of carbon and energy flow in microbial communities. The strategy involves the study of highly interactive and tractable model autotroph-heterotroph consortia whose member genome sequences have been defined. Our project leverages unique capabilities including multi-omics measurements, advanced functional imaging, taxonomic profiling and metabolic and regulatory network modeling to elucidate underlying reaction mechanisms within complex microbial communities. Our research plan supports DOE goals to achieve a predictive understanding of microbially-mediated carbon and energy transformation.

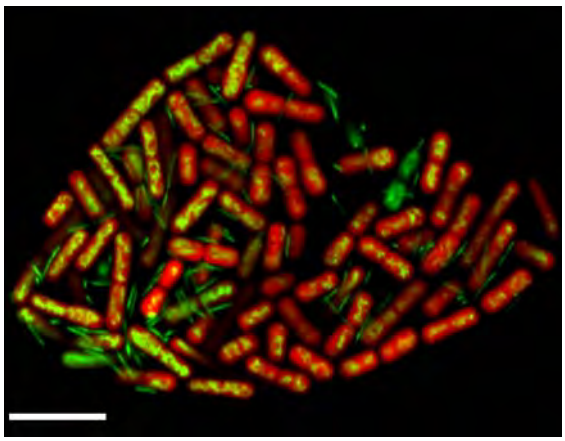


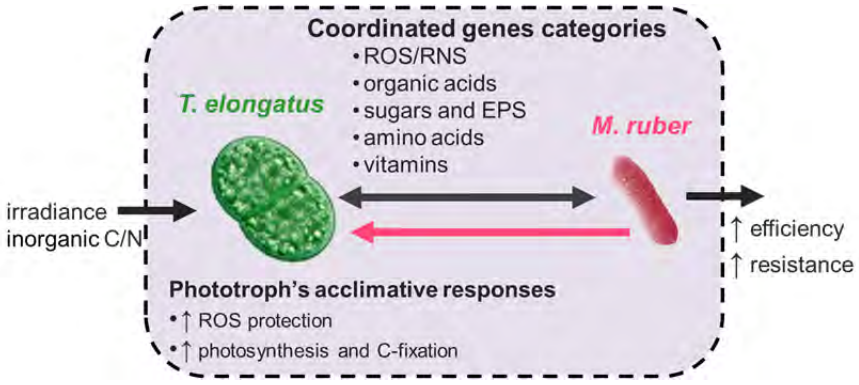
Figure 1. The *T. elongates* and *M. ruber* consortium was constructed to specifically test how microbes coordinate their metabolisms when exchanges of reduced carbon and nitrogen are essential for co-existence. We found that the cyanobacterium responded to its heterotrophic partner by altering transcriptional events and that the energy efficiency for biomass production is greater in the binary consortium as compared to the *T. elongatus* axenic control. The scale bar marks a 30 µm distance.

The means by which individual microorganisms affect each other's gene expression and coordinate metabolism in communities are largely unknown. To explore this, we interrogated specific interactions in a photoautotroph- heterotroph consortium to infer mechanisms that mediate metabolic coupling and acclimation to microbial partnerships centered on exchanges of reduced carbon and nitrogen. We constructed a binary consortium from a cyanobacterium, *Thermosynechococcus elongatus* BP-1 and an obligate aerobic heterotroph, *Meiothermus ruber* Strain A, that relies upon cyanobacterial organic carbon, O_2 and reduced nitrogen. We sought to identify the details of the interactions that supported growth of *M. ruber* and asked whether there were associated costs or benefits to the primary-producer, *T. elongatus*. We used species-resolved transcriptomic analyses in combination with growth and photosynthesis kinetics to infer species interactions and the environmental context under which they occur. In addition, we evaluated the transcriptional and physiological responses to partnerships across tightly controlled, nutrient-replete steady-states maintained via discrete energy inputs (incident irradiance) and

environmental stress imposed by elevated dissolved O_2 tensions. The guiding hypothesis for our work, to-date, is that that exchanged resources (i.e., reduced-C, -N, vitamins and O_2) act as the molecular singles that coordinated gene expression between the cyanobacterial producer and its heterotrophic partner.

We observed coordinated transcriptional responses from both species and used these to infer specific interactions resulting from the synthesis and exchange of resources, including amino acids, sugars and organic acids and other organic carbon sources. For example, the cyanobacterium responded to heterotrophic partnership by altering expression of genes involved with photosynthesis, carbon uptake, vitamin synthesis and scavenging of reactive oxygen species; the latter likely providing community-wide protection from oxidative stress. Heterotrophic partnership increased the efficiency of biomass production and resistance to stress induced by high levels of dissolved O_2 as compared to axenic cyanobacterial controls. Ongoing efforts to understand how these members might regulate each other's function are focused on the development of community-scale gene and metabolite association networks which can lead to the identification of interspecies, metabolic and regulatory coordination. From these networks, we are also building predictive genome-scale metabolic reaction networks to model specific exchanges of metabolites that may coordinate community-level behavior. By using the KBase Platform, we plan to predict specific interactions based on a novel network building procedure that integrates coordinated meta-transcriptional data directly into a interspecies network, the results of which are compared with conventional integration of data derived for individual species models (*see companion poster; Hyun-Seob Song et al. 2016*).

The resulting, genome-level discoveries and community metabolic models presented here provide a benchmark, systems-level foundation to infer specific interactions occurring between a unicellular cyanobacterium and its heterotrophic partner driven by direct metabolite exchange. Our findings, to date, support our hypothesis that microorganisms respond to partnership at the transcriptional level and are capable of acclimating to each other through indirect, interspecies regulation metabolic processes. We propose that some of the interactions, identified here, represent generalizable principles for autotroph-heterotroph community properties.



This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the PNNL. Author Hans Bernstein is grateful for support provided by the Linus Pauling Distinguished Postdoctoral Fellowship, a Laboratory Directed Research and Development program, at the Pacific Northwest National Laboratory (PNNL).

Enhanced Cross-Feeding in a Bacterial Coculture Undermines a Mutualistic Relationship

Breah LaSarre, Alexandra L. McCully, and **James B. McKinlay*** (jmckinla@indiana.edu),

Indiana University, Bloomington

Project Goals: The goals of this project are to (i) develop a stable hydrogen gas-producing coculture between *Rhodopseudomonas palustris* and fermentative microbes, such as *Escherichia coli*, (ii) use genetic, biochemical, evolutionary, and systems biology approaches to characterize and manipulate microbial interactions and H_2 production.

Synthetic microbial communities, or cocultures, preserve core aspects of microbial interactions found in natural environments while allowing for an expanded range of experimental approaches and control. Cocultures are thus valuable for addressing questions in microbial ecology and evolution, such as how microbes transform carbon as a community. Cocultures can also combine complementary traits from diverse microbes to convert renewable resources into fuels and other useful chemicals. However, establishing cocultures that support stable coexistence and yield reproducible results has proven challenging in many cases.

Our work focuses on an anaerobic coculture between fermentative *Escherichia coli* and an engineered strain of phototrophic *Rhodopseudomonas palustris* (Nx) that together convert carbohydrates into H_2 gas, a potential biofuel. The two bacteria form a syntrophic relationship wherein *E. coli* ferments glucose and excretes essential carbon (organic acids) for *R. palustris* Nx while *R. palustris* Nx fixes N_2 and excretes essential nitrogen (NH_4^+) for *E. coli*. The bidirectional exchange ensures stable coexistence as the two populations grew to a common equilibrium from starting ratios spanning over twelve orders of magnitude. Growth and metabolic trends were highly reproducible over serial transfers, motivating us to develop an ecological model that accurately describes these trends.

Exchange of metabolites is an important basis for many mutualistic systems. However, the impact of metabolite exchange rates on the bacteria involved is typically difficult to address. Using our model, we simulated the effects of higher *R. palustris* NH_4^+ excretion rates. The model predicted that more NH_4^+ excretion would result in a less efficient utilization of feedstock and a lower final *R. palustris* cell density. Essentially, higher levels of NH_4^+ would stimulate faster *E. coli* growth and organic acids would be produced faster than *R. palustris* could consume them, eventually resulting in a growth-inhibiting acidic pH. By cooperating more, *R. palustris* would change the nature of organic acids from a carbon source to a growth inhibitor. To test these predictions, we generated a ‘hyper-cooperator’ strain of *R. palustris* that excretes 3-fold more NH_4^+ than the Nx parent. Cocultures with the hyper-cooperator confirmed the predictions, as there were fewer *R. palustris* cells, a higher residual organic acid concentration, and a more acidic pH. Nonetheless, the hyper-cooperator and *E. coli* stably coexisted over serial transfers.

Our results have implications for the deployment of engineered NH_4^+-excreting diazotrophs as biofertilizers for agricultural and industrial applications and perhaps for other engineered cross-feeding systems. Varied NH_4^+ excretion rates can indirectly modify system behavior in potentially negative ways, such as altering the species ratio or misdirecting resources towards undesired products. In other words, optimizing metabolite excretion rates for cross-feeding does necessarily mean increasing the excretion rates.

This work was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, under award number DE-SC0008131.

Characterization of the Alginate Lyases and Laminarinases from *Vibrio sp.*

Ahmet Badur[1], Ehar Ammar[1], Matthew Plutz[1], Geethika Yalamanchili[1], Jan-Hendrik Hehemann[3], Martin Polz[2], and **Christopher V. Rao**[1*] (cvrao@illinois.edu)

[1]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign
[2]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology
[3]Center for Marine Environmental Sciences, University of Bremen

Project Goals: This project will harvest 'biomass to biofuel' pathways from algae-associated bacteria, and develop these as reusable genetic parts. Marine algae hold great promise for biofuel production and have advantages over terrestrial biomass and freshwater algae. Despite this potential, little effort has been made to date to harness the enzymatic machinery that bacteria use to convert marine algal carbohydrates into bioenergy substrates. Our project capitalizes on this unexplored opportunity via three distinct activities: bioprospecting for novel algal polysaccharide-degrading genes, functional screening for enzymes with desired biochemical properties, and repackaging pathways in reusable genetic modules.

Brown seaweeds are an attractive source of feedstocks for biofuel production, since they have advantages over terrestrial feedstocks. Brown seaweeds have higher growth rates than terrestrial plants, and they lack crystalline cellulose and lignin. Additionally, brown seaweeds do not impinge on arable land, thus negating the conflict between food and fuel. Two of the primary components of brown seaweeds are alginate and laminarin. Alginate is a copolymer consisting of 1,4 linked epimers α- L-guluronate (G) and β-D-mannuronate (M). The local structure of alginate can take one of three forms: short stretches of polyguluronate (polyG), short stretches of polymannuronate (polyM), or alternating sequences of guluronate and mannuronate. The enzymes that can degrade the linkages within alginate are called alginate lyases. Alginate lyases are classified based on their specific dyad G-G (EC 4.2.2.11), M-M (EC 4.2.2.3), and M-G/G-M bonds that they cleave. Additionally, alginate lyases are classified based on whether they have exolytic or endolytic cleavage. Laminarin is a polysaccharide consisting of β-1,3 and β-1,6 linked glucose. The enzymes that can degrade these linkages are called glycoside hydrolases (GHs). More specifically, the β- 1,3 linkage is degraded by enzymes belonging to seven GH families: GH3, GH5, GH16, GH17, GH55, GH64, and GH81. β-1,6 degrading GHs are remain unknown.

We are investigating the mechanism of alginate and laminarin metabolism within marine *Vibrio sp.* To this end, we next cloned, purified, and characterized the alginate lyases within *Vibrio splendidus* 12B01 and 13B01 and *Vibrio breoganii* 1C10. We found that these enzymes are most active between pH 7.5 and 8.5, 20°C and 25°C, and 250 and

1000 mM NaCl. We then determined the enzyme kinetics for each enzyme. We found that each enzyme had a V_{max} between 0.090 and 1.7 μM s^{-1}, K_M between 22 and 300 μM laminarin, and a turnover number between 0.19 and 4.9 s^{-1}. We also determined the dyad specificity of each lyase; we found G-M, G-G, M-M, and M-G specificity. Between the three organisms, we have characterized 18 alginate lyases, which allows for a broad sampling of how *Vibrio sp.* degrade alginate. We have begun to elucidate synergies between lyases. Since a single organism contains multiple lyases, we want to find how alginate lyases can be used simultaneously to degrade alginate to completion and with faster kinetics.

We are also investigating the mechanism of laminarin metabolism within marine *Vibrio sp.* Firstly, we have found that laminarinases are induced by growth on laminarin. We next sought to clone, purify, and characterize the four laminarinases within *Vibrio breoganii* 1C10. We found that these enzymes are most active between pH 6.5 and 8.0, 20°C and 25°C, and 50 and 400 mM NaCl. We then determined the enzyme kinetics for each enzyme. We found that each enzyme had a Vmax between 0.148 and 0.92 μM s-1, KM between 3.4 and 6.0 mM laminarin, and a turnover number between 0.69 and 6.1 s-1. These results now allow for metabolic engineering of microorganisms that degrade laminarin as their sole carbon source.

This project is a part of the Biosystems Design Program supported by the Office of Biological and Environmental Research in the DOE Office of Science

Bacterial Succession Patterns in Algal Degradation: Identifying the Most Efficient Algal Degrading Communities

Xiaoqian Yu[1*](xy43@mit.edu), Christopher Corzett[1] (corzett@mit.edu), Jan-Hendrik Hehemann[2], **Martin Polz**[1] and **Eric Alm**[1]

[1] Massachusetts Institute of Technology, Cambridge, MA, USA

[2] Max Planck Institute for Marine Microbiology, Bremen, Germany

Project Goals: Marine algae hold great promise for biofuel production and have advantages over terrestrial biomass and freshwater algae. Despite this potential, little effort has been made to date to harness the enzymatic machinery that bacteria use to convert marine algal carbohydrates into bioenergy substrates. Our project capitalizes on this unexplored opportunity via three distinct activities: bioprospecting for novel algal polysaccharide-degrading genes, functional screening for enzymes with desired biochemical properties, and repackaging pathways in reusable genetic modules.

In the marine ecosystem, macroalgae are both fast growing carbon producers and major carbon sinks; they represent half of the carbon production and three-quarters of the biomass in coastal oceans. Studying the organisms and metabolic pathways that are involved in natural degradation of macroalgae could help improve our understanding of marine energy flow. This research also has potential to provide guidelines for synthetic community design in using macroalgae as a feedstock for biofuel production. In a microcosm experiment with *Fucus* extract as the sole carbon source provided, we found that different natural marine communities (surface seawater/sediment) followed similar community succession patterns. Only five bacterial families (*Halomonadaceae, Oceanospirillaceae, Pseudoalteromonadaceae, Rhodobacteraceae, Vibrionaceae*) increased to a relative abundance of 10% or more in the microcosm communities. The expansion of these families followed a strict sequential order. In another microcosm experiment performed, random subsets of a natural seawater community were generated by a removal-of-species-by-dilution method. Some of these subset communities were found to be 2-4 fold more productive than the original natural seawater community in terms of degrading *Fucus* extract. These two microcosm experiments suggest that specific combinations of bacteria from certain families may be optimal at degrading a complex carbon source such as *Fucus* material. These specific combinations will be further identified and characterized using a combination of culture-free and culture-based methods.

This work is supported by the Office of Biological and Environmental Research in the Department of Energy Office of Science (DE-SC0008743).

The Ecology of Algal Polysaccharide Utilization: Novel *Verrucomicrobia* Isolates Efficiently Degrade Fucoidan

Christopher Corzett[1*] (corzett@mit.edu), Jan-Hendrik Hehemann,[2] **Eric Alm**,[1] and **Martin Polz**[1]

[1] Massachusetts Institute of Technology, Cambridge, MA, USA
[2] Max Planck Institute for Marine Microbiology, Bremen, Germany

Project Goals: Marine algae hold great promise for biofuel production and have advantages over terrestrial biomass and freshwater algae. Despite this potential, little effort has been made to date to harness the enzymatic machinery that bacteria use to convert marine algal carbohydrates into bioenergy substrates. Our project capitalizes on this unexplored opportunity via three distinct activities: bioprospecting for novel algal polysaccharide-degrading genes, functional screening for enzymes with desired biochemical properties, and repackaging pathways in reusable genetic modules.

Marine macroalgae are vital players in the global carbon cycle, and polysaccharides represent a significant output of their primary production. Identifying the microbes and metabolic pathways responsible for degrading these sugars is not only crucial to understanding marine carbon flow, but also offers potential for biofuel production using seaweed feedstocks. Fucoidans are an important class of structurally heterogeneous sulfated polysaccharides found in brown seaweeds, yet few organisms have been shown to metabolize this abundant carbohydrate. Enrichment cultures from coastal samples yielded numerous isolates from diverse genera (*Vibrio*, *Stappia*, *Neptunomonas*, *Alteromonas*, *Tenacibaculum*) implicated in the breakdown of fucoidan polysaccharides, and novel *Verrucomicrobia* isolates performing an especially critical role in facilitating this process were sequenced. Draft genomes reveal Polysaccharide Utilization Loci (PULs) enriched with numerous and diverse Carbohydrate-Active Enzymes (CAZymes), with some isolates encoding as many as 60 genes with homology to established fucosidases. Differences in the dynamics and extent of fucoidan utilization among closely related isolates suggest variation in enzymatic capabilities and may reflect resource partitioning. Specific combinations of natural isolates appear to complement one another and yield greater overall biomass accumulation, suggesting engineered organisms or communities with a full repertoire of enzymatic machinery may facilitate the efficient conversion of algal biomass.

This work is supported by the Office of Biological and Environmental Research in the Department of Energy Office of Science (DE-SC0008743).

Isotopically nonstationary ^{13}C flux analysis of isobutyraldehyde production in Synechococcus elongatus

Lara J. Jazmin[1], Yao Xu[2], Adeola O. Adebiyi[1], Yi Ern Cheah[1]*(yi.ern.cheah@vanderbilt.edu), Carl H. Johnson[2], **Jamey D. Young**[1,3]

[1] Chemical & Biomolecular Engineering, Vanderbilt University, Nashville, TN
[2] Biological Sciences, Vanderbilt University, Nashville, TN
[3] Molecular Physiology & Biophysics, Vanderbilt University, Nashville, TN

http://www.vanderbilt.edu/younglab

Project Goals: This project aims to develop strains of cyanobacteria that are optimized for producing advanced biofuels. This will be done by applying isotopically nonstationary ^{13}C metabolic flux analysis (^{13}C-MFA) to quantify photoautotrophic metabolism in engineered cyanobacterial strains and then redirecting carbon flux toward biofuel production using rational pathway manipulations identified by ^{13}C-MFA.

Photosynthetic microorganisms are promising systems for converting energy from sunlight and carbon from CO_2 directly into renewable fuels and chemicals. Despite recent advances in cyanobacterial biofuels production, the productivities achieved are yet to be economicaly feasible and few tools are available that specifically address the challenges of determining and redirecting metabolic flux in photosynthetic microbes.

Our group is developing novel approaches that use ^{13}C-MFA to quantitatively assess *in vivo* metabolic phenotypes of photoautotrophs. Previously, we used this approach to map carbon fluxes in wild-type *Synechococcus elongatus* PCC7942 (WT) and a mutant (SA590) engineered to convert pyruvate to isobutyraldehyde (IBA) [1]. Compared to WT, ^{13}C-MFA revealed an increased flux through a pyruvate kinase (PK) bypass pathway in SA590. As a result, we generated SA590 mutants that singly overexpress each gene in the PK bypass pathway: PEP carboxylase (PEPC), malate dehydrogenase (MDH), and malic enzyme (ME). These mutants showed significant improvements in IBA production while maintaining comparable growth rates to SA590.

We recently examined the intracellular metabolite pool sizes of our mutants to elucidate the effects of the singly overexpressed PK bypass genes. Compared to SA590, a large increase in malate pool size was observed in strain SA590-MDH. We hypothesize that the flux to malate could be redirected toward pyruvate (and hence toward increasing IBA production) by tandem overexpression of ME and PEPC. Hence, two additional mutants were generated to test this hypothesis: SA590-MDH-ME and SA590-MDH-ME-PEPC. This presentation summarizes our efforts to date and demonstrates the utility of ^{13}C-MFA in guiding rational pathway engineering of photosynthetic microorganisms for biofuel production.

This work is funded by the U.S. Department of Energy (DOE) Award DE-SC008118.

[1] S. Atsumi, W. Higashide, J.C. Liao, Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde, Nature Biotechnology, 27 (2009) 1177-U1142.

Cellular Impact of Inactivation of the Nitrate Reductase Gene in the Marine Diatom*, Phaeodactylum tricornutum*

Andrew. E. Allen (PI)[1,2*], McCarthy, James[1], Christopher L. Dupont (Co-PI)[1], Ursula Goodenough[3]

[1]Microbial and Environmental Genomics, J. Craig Venter Institute, San Diego, CA, USA
[2]Integrative Oceanography Division, Scripps Institution of Oceanography, University of California, San Diego, CA 92093, United States
[3]Department of Biology, Washington University, St. Louis, MO 63130

A knockout of *Phaeodactylum tricornutum's* nitrate reductase (NR) gene, resulting in the absence of the NR protein, was produced by TALEN-enabled homologous recombination. In place of NR, a stable, selectable marker was introduced into the *P. tricornutum* genome. Investigations of the NR knockouts by confocal microscopy and deep-etch freeze fracture EMs show a dramatic reorganization of normal pennate diatom physiology. KO cells are swollen and deformed with chloroplasts reduced and pushed to the sides or ends of the cells. FTIR results indicate a dramatic rearrangement of protein and carbohydrate fractions in the scans of WT and NR-KO samples. UV-Spectroscopy showed that NR-KO cells had, within 3 days, accumulated greater than 1 mM NO_3^-, whereas in WT cells, by assimilating the NO_3^-, there was almost no NO_3^- in the cell extracts. An increase was also observed in the NR-KO lipid fraction as compared to the wild type. FAME analysis of cell lipids showed an immediate, substantial redistribution of TAG fatty acids in the knockout line. Whole transcriptome analysis, comparing NR-KO lines vs WT expression profiles and differential expression over a 10-day time series, confirmed and helped characterize the physiological and biochemical transformation of *P. tricornutum* caused by the loss of function of nitrate reductase.

Influence of DNA Delivery on Mechanism of CRISPR-Cas Genome Editing in Marine Diatoms

Andrew. E. Allen (PI)[1,2*], Mark Moosburner[1,2]

[1]Microbial and Environmental Genomics, J. Craig Venter Institute, San Diego, CA, USA
[2]Scripps Institution of Oceanography, University of California, San Diego, CA 92093, United States

Diatoms, a major phylogenetic group of phytoplankton, play a significant role in shaping marine systems in a wide range of aquatic environments. They greatly contribute to the pool of organic carbon fixed via photosynthesis in contemporary oceans, producing between 25% and 40% of the 40-50 billion tons of organic matter generated annually. They also flourish in conditions where nutrient availability is variable, such as coastal upwelling and polar regions. Along with their ecological significance, diatoms have unique metabolic properties that hold promising biotechnology potential in sustainable biofuels and nano-materials. Sequencing data from diatoms, most notably *Phaeodactylum tricornutum (Pt)* and *Thalassiosira pseudonana (Tp),* greatly aids in determining the genetic and genomic basis for diatoms' ecological and biotechnological importance, yet genetic manipulation tools essential to explore such areas are not well developed in diatoms. To date, tools such as RNAi and TALENs have been employed in the laboratory strains *Pt* and *Tp*, but the more powerful, sophisticated, and flexible genome engineering tool, CRISPR-Cas, has yet to be implemented. Here, CRISPR-Cas genome engineering was utilized to functionally investigate and disrupt the urease and nitrate reductase genes. Two DNA repair mechanisms, non-homologous end joining (NHEJ) and homology directed repair (HDR), were used to induce mutations randomly and introduce scar-less mutations, respectively, at the CRISPR-Cas target locus. Also, two genetic transformation techniques available, micro-particle bombardment and bacterial conjugation, where harnessed to deliver the CRISPR-Cas machinery to diatom *Pt*. This study demonstrates that precise targeted mutagenesis via HDR can be accomplished when employing micro-particle bombardment delivery, and NHEJ-mediated mutations can be induced using either transformation avenues. The development of the CRIPSR-Cas technology using both micro-particle bombardment and bacterial-conjugation vastly widens the possibilities in the exploration diatom genetics.

Mechanisms of Carbon Partitioning into Chrysolaminarin, the Storage Polysaccharide of Diatoms

Michael A. Caballero[1]* (mike.caballero@colostate.edu), Graham Peers,[1] and **Andrew E. Allen[2]**

[1]Colorado State University, Fort Collins; [2]J. Craig Venter Institute, La Jolla, CA

Project Goals: Our overall goal is to reprogram metabolic networks using i*n vivo* synthetic modules to increase the flux of energy and carbon into biofuel precursors in the marine diatom *Phaeodactylum tricornutum*. This is broken down into 3 sub-goals: 1) Profile the transcriptome, proteome and metabolome to investigate cell responses to physiologically relevant conditions. 2) Identify and manipulate key factors involved in the control of inorganic C assimilation, photosynthetic efficiency and carbon partitioning. 3) Create screen and genotype a forward genetic library generation. These approaches complement our development of *Phaeodactylum* genome reconstruction modeling and our development of novel synthetic genomic tools to achieve our overall goal of increasing photosynthetic productivity.

Diatoms contribute to global carbon cycles, accounting for about one-fifth of global primary productivity. Partitioning photosynthate into storage metabolites enables flexible cellular metabolism by providing a reservoir of carbon and energy. Diatoms store carbohydrate as chrysolaminarin, a β-1,3 glucan, instead of starch or glycogen. Disrupting chrysolaminarin metabolism may direct diatom carbon partitioning from storage carbohydrates to triacylglycerol, an important metabolite for biodiesel production. However, the genes responsible for producing proteins important for chrysolaminarin metabolism remain broadly unknown. We are interested in identifying these genes, disrupting their expression, and investigating their impact on carbon partitioning.

The lack of a robust method for chrysolaminarin quantification was an initial challenge for this study. We developed a method to selectively quantify chrysolaminarin from *Phaeodactylum tricornutum* cell extracts with selective hydrolysis. Chrysolaminarin reserves were depleted overnight, decreasing from 1.69±0.27 to 0.06±0.03 pg glucose equivalents per cell, while structural carbohydrates did not significantly decrease. H-NMR structural analysis of *Phaeodactylum's* chrysolaminarin indicates a smaller polymer with less branching than commercially available laminarin. This method enables phenotyping *Phaeodactylum* mutants by chrysolaminarin accumulation, rather than established total-carbohydrate methods.

We have identified chrysolaminarin-related targets through several genomics-enabled approaches. Comparative genomics has identified several conserved UGPases in the *Phaeodactylum* genome, and are thought to be the initiating step of chrysolaminarin synthesis. We are investigating the contribution of a putatively chloroplast-localized UGPase to chrysolaminarin biology. In a complementary approach, we identified a putative, chrysolaminarin-related, phosphatase-like protein (PCP) from the *Phaeodactylum* proteome using 2D-affinity electrophoresis. PCP RNAi lines accumulated two-fold more chrysolaminarin

per cell volume than wild type.

This research seeks to identify and study new chrysolaminarin-related proteins and expands the available biochemical toolkit to characterize carbon partitioning in algae. Investigating chrysolaminarin biology will improve our understanding of diatom central carbon metabolism and informs our future bioengineering efforts to produce metabolites of interest, such as triacylglycerol for biodiesel.

Project supported by the DOE Office of Science-BER (DE-SC0008595).

A Systems-Level Investigation of Low-Light Acclimation in the Marine Diatom *Phaeodactylum tricornutum*.

Graham Peers*[1] (graham.peers@colostate.edu), Nathan Sindt,[1] Wen Zhou,[1] Jessica Prenni,[1] Chris L. Dupont,[2] **Andrew E Allen[2]**

[1]Colorado State University, Ft. Collins; [2]J. Craig Venter Institute, San Diego, CA

Project Goals: Our overall goal is to reprogram metabolic networks using i*n vivo* synthetic modules to increase the flux of energy and carbon into biofuel precursors in the marine diatom *Phaeodactylum tricornutum*. This is broken down into 3 sub-goals: 1) Profile the transcriptome, proteome and metabolome to investigate cell responses to physiologically relevant conditions. 2) Identify and manipulate key factors involved in the control of inorganic C assimilation, photosynthetic efficiency and carbon partitioning. 3) Create screen and genotype a forward genetic library generation.These approaches complement our development of *Phaeodactylum* genome reconstruction modeling and our development of novel synthetic genomic tools to achieve our overall goal of increasing photosynthetic productivity.

The natural light environment is highly dynamic. Photosynthetic metabolism is regulated at the physiological and genetic level to respond to these changes. Diatoms, which are eukaryotic algae, are capable of adjusting their metabolism to maximize light harvesting efficiency in low light and minimize oxidative damage in excess light. Most studies have focused on the stress response during acclimation to excess light, but low light photoacclimation has received less attention. We observed changes in the transcriptome, proteome, metabolome and in the photophysiology of *Phaeodactylum tricornutum* following a shift from excess light fluxes to low light (880 to 80 μmol photons/m^2/s) over 24 hours. We observed an increase in cell pigments and in pigment precursors using established visible spectroscopy methods and by our non-targeted metabolomics methods. We developed novel statistical methods for analyzing time-series based –omics data. In general, we saw increases in light harvesting complex (LHC) transcripts following the shift to low light this was trailed by a significant increase in total LHC protein abundance. However, changes in an individual LHC's protein abundance were not necessarily predicted by changes in transcript. This pattern was frequently observed when comparing transcript and protein abundance and we will discuss the implications of inferring metabolic changes from transcriptomics alone. Catabolic pathways quickly responded to lower light. Energy storage molecules, such as triacylglyercols and chrysolaminarin (carbohydrate storage polymer) decreased. The relative abundance of many hexoses, disaccharides and fatty acids began to increase within 20-40 minutes and this was accompanied by rapid up-regulation of the TCA cycle, but without major increases in measured respiration rates. Lipid classes associated with both the thylakoid membrane and other structural lipids increased in relative abundance during acclimation to low light. This correlated with an increase in transcripts encoding components associated with fatty acid elongation. Interestingly, organic carbon per cell only

dropped by 15% during photoacclimation and growth rates remain unchanged. These results suggest that *Phaeodactylum* can rapidly adjust its metabolism to maintain fitness a rapidly changing environment. These findings will be integrated into a new genome reconstruction model in order to improve our understanding of energy and carbon partitioning. Ultimately, this study will provide guidance for our engineering efforts to increase the production of biofuel precursors.

Project supported by the DOE Office of Science-BER (DE-SC0008595).

Towards Repurposing the Yeast Peroxisome for Compartmentalizing Heterologous Metabolic Pathways

Zachary Russ,[1] William DeLoache, [1] Jennifer Samson,[1] **John Dueber[1,2]***

[1]University of California, Berkeley; [2]Lawrence Berkeley National Laboratory, Berkeley, CA

Our long-term goal is to repurpose the peroxisome to be a synthetic organelle that can insulate engineered multi-enzyme pathways from undesired cross-talk with the rest of the cell production host. Ultimately, we endeavor to control the chemical environment of the organelle such that catalysis can be conducted that would not be feasible in the cytoplasm. Towards this long-term goal, we aim to learn about natural peroxisome biology, increase the efficiency of protein cargo import, and characterize small molecule permeability.

Compartmentalization of enzymes into organelles is a promising strategy for limiting metabolic crosstalk and improving pathway efficiency, but improved tools and design rules are needed to make this strategy available to more engineered pathways. Here we focus on the Saccharomyces cerevisiae peroxisome and develop a sensitive high-throughput assay for peroxisomal cargo import. We identify an enhanced peroxisomal targeting signal type 1 (PTS1) for rapidly sequestering non-native cargo proteins. Additionally, we perform the first systematic in vivo measurements of nonspecific metabolite permeability across the peroxisomal membrane using a polymer exclusion assay. Finally, we apply these new insights to compartmentalize a two-enzyme pathway in the peroxisome and characterize the expression regimes where compartmentalization leads to improved product titer. This work builds a foundation for using the peroxisome as a synthetic organelle, highlighting both promise and future challenges on the way to realizing this goal.

Publications

1. DeLoache, WC, Russ, ZN, and Dueber JE. Towards Repurposing the Yeast Peroxisome for Compartmentalizing Heterologous Metabolic Pathways. Nature Communications. 2016. In press.
2. Deloache, W. C., Russ, Z. N., Narcross, L., Gonzales, A. M., Martin, V. J. J., & Dueber, J. (2015). An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. Nature Chemical Biology, 11(7), 465–471. http://doi.org/10.1038/nchembio.1816

Functional Characterization and Regulatory Modeling of Lignocellulose Deconstruction in the Saprophytic Bacterium *Cellvibrio japonicus*

Cassandra E. Nelson,[1] Nina R. Beri,[1] and **Jeffrey G. Gardner**[1]* (jgardner@umbc.edu)

[1]Department of Biological Sciences, University of Maryland – Baltimore County

Project Goals: Understanding polysaccharide degradation by microbes is of great importance to unraveling environmental processes driving global nutrient cycles, determining gut microbiota nutritional contributions, and alleviating bottlenecks in renewable fuel and chemical production. However, one current knowledge gap is how these microbes are able to degrade and consume what is effectively an unreactive and inert substrate (lignocellulose). To that end, completion of this project will facilitate the establishment a fundamental systems-level model of lignocellulose deconstruction by saprophytic soil bacteria. Additionally, over the course of this project we will identify and characterize novel enzymes that have the potential to accelerate the advancement of renewable fuel and chemical technologies.

Substantial engineering of industrially relevant bacteria allows a synthetic metabolism in these microbes to make desirable compounds. However, one challenge that still exists is obtaining low cost substrates that feed into this synthetic metabolism. Cellulosic materials represent a deep reservoir of sugars, however they are locked in a recalcitrant polymeric form. Decades of enzymatic studies have sought to overcome the recalcitrance of environmental polysaccharides such as lignocellulose. Recently several biochemical studies have commented that *in vivo* studies of recalcitrant polysaccharide degradation will be required to fully understand the process, as exclusively *in vitro* studies will not necessarily identify more efficient enzymes, especially in a physiologically or biotechnologically relevant context [1-3]. Therefore, returning to fundamental studies of environmental bacteria that are proficient at lignocellulose degradation presents a promising area of study. Novel enzymes with desirable properties uncovered by this approach can then be further enhanced with synthetic biology techniques for renewable fuel and chemical production.

We have employed systems approaches and interdisciplinary studies to characterize in a physiologically relevant manner the cellulose degradation capabilities in saprophytic bacteria [4]. Our approach is comprehensive and incorporates a diversity of techniques. Briefly, we use transcriptomics (RNAseq) to determine both up-regulated and highly constitutive genes under conditions of interest (*i.e.* degrading lignocellulose). This approach yields a gene set to study with classical bacterial genetics. Using an in-frame deletion strategy developed by our group [5], we can evaluate sets of genes in a physiologically relevant manner and determine genes with essential function. This approach has also uncovered novel genes of previously unknown function [4]. Enzyme kinetic analysis further characterizes essential genes, while determining if they are candidates for industrial processes. Completing our analysis toolkit are computational methods for determining suites of co-regulated genes and regulatory networks. A synthesis of these data results in a comprehensive and predictive model of polysaccharide degradation by saprophytic bacteria.

Here we present our recent progress examining recalcitrant polysaccharide degradation using the model saprophytic bacterium *Cellvibrio japonicus*. We will discuss our methods to use exclusively physiologically relevant substrates (*e.g.* minimally processed corn stover or switchgrass) to discover novel enzyme targets. Additionally, we will describe our approach to modeling the complex regulatory networks required to detect and consume environmental polysaccharides.

References

1. Cartmell A., McKee L.S., Pena M.J., Larsbrink J., Brumer H., Kaneko S., Ichinose H., Lewis R.J., Vikso-Nielsen A., Gilbert H.J., and Marles-Wright J. 2011. The structure and function of an arabinan-specific alpha-1,2-arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases. *Journal of Biological Chemistry*. 286(17):15483-95.
2. Naas A.E., Mackenzie A.K., Mravec J., Schuckel J., Willats W.G.T., Eijsink V.G.H., and Pope P.B. 2014. Do rumen Bacteroidetes utilize an alternative mechanism for cellulose degradation? *MBio*. 5(4): e01401-14. doi:10.1128/mBio.01401-14.
3. Zhang X., Rogowski A., Zhao L., Hahn M.G., Avci U., Knox J.P., and Gilbert H.J. 2014. Understanding how complex molecular architecture of mannan-degrading hydrolysis contributes to plant cell wall degradation. *Journal of Biological Chemistry*. 289(4): 2002-12.
4. Gardner J.G., Crouch L., Labourel A., Forsberg Z., Bukhman Y.V., Vaaje-Kolstad G., Gilbert H.J., and Keating D.H. 2014. Systems biology defines the biological significance of redox-active proteins during cellulose degradation in an aerobic bacterium. *Molecular Microbiology*. 94(5): 1121-33.
5. Nelson C.E. and Gardner J.G. 2015. In-frame deletions allow functional characterization of complex cellulose degradation phenotypes in *Cellvibrio japonicus*. *Applied and Environmental Microbiology*. 81(17): 5968-75.

This work is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0014183.

Defining the Minimal Set of Microbial Genes Required for Valorization of Lignin Biomass

Nikita Khylsov[1], Russ Li[1] and **Elizabeth Sattely**[1]* (sattely@stanford.edu)

[1]Stanford University, Stanford, CA

Project Goals: Lignin is the second most abundant biopolymer on earth and represents a critically underutilized biomass resource for hydrocarbon feedstocks. Despite substantial effort, there is still no efficient process to convert lignin to useable carbon-based platform chemicals and materials. The goal of this project is identify a minimal set of microbial enzymes necessary for lignin breakdown and sufficient for the synthesis of valuable chemical intermediates from lignin isolated as a byproduct of lignocellulosic ethanol production. These genes will be then used to engineer functional whole cell biocatalysts for tunable lignin metabolism.

To date, although a number of enzymes have been associated with lignin degradation, most have been tested in isolation (as individual enzymes) and on drastically different substrates -- often dyes that are not related to lignin. In contrast, lignin utilization in nature likely occurs by microbial consortia with multiple enzymes acting synergistically. We propose to examine two separate stages of lignin breakdown carried out by the microbes that do it best: (1) early breakdown of native polymeric lignin into soluble fragments by a set of sequenced wood-rotting fungal species, and (2) downstream metabolism of these soluble lignin fragments to useful chemical intermediates by a panel of sequenced soil saprophytes. Our approach involves testing sets of genes that will be assayed combinatorially in the context of a heterologous expression host. The resulting engineered strains will be systematically assayed using soluble lignin fragments, synthetic defined polymeric lignin, and finally lignin directly sourced from lignocellulosic processing streams. In addition to resulting in a functional whole cell biocatalyst for lignin utilization, we anticipate that this approach will allow us to address key unanswered questions about lignin metabolism in nature, including: (1) Why does the *Trametes versicolor* genome contain 25 different class II peroxidases? (2) What is the role of laccases in lignin metabolism? Why do some aggressive lignin degraders have many laccases (e.g. >7 in *T. versicolor*) while others have none (e.g. *P. chrysosporium*)? (3) How is peroxide provided in a controlled manner to drive peroxidase activity without causing the enzyme inhibition that is so often observed in vitro? (4) What strategies do microbial lignin degraders use to avoid the problem of repolymerization during active lignin degradation? and (5) Can microbial lignin metabolism be diverted for high level production of defined aromatics? A final critical question is whether combining key minimal sets of enzymes from a wide range of organisms will result in engineered strains capable of highly efficient, streamlined pathways for lignin utilization that can be tuned for a specific carbon output. This effort will leverage DOE investments in microbial genome sequencing, and secure a critical channel for lignin biomass utilization that will also help to render lignocellulosic a viable feedstock for the production of renewable liquid biofuels.

Funding statement. DE-SC0014112- Defining the Minimal Set of Microbial Genes Required for Valorization of Lignin Biomass

Engineering Anaerobic Gut Fungi for Lignocellulose Breakdown

John K. Henske[1]* (johnhenske@enginering.ucsb.edu), Sean P. Gilmore[1], Kevin V. Solomon[1], Michael K. Theodorou[2], Heather M. Brewer,[3] Samuel O. Purvine,[3] Aaron T. Wright,[3] Alan Kuo,[4] Stephen Mondo,[4] Asaf Salamov,[4] Kurt LaButti,[4] Dawn Thompson[5], Aviv Regev[5], Igor Grigoriev,[4] **Michelle A. O'Malley[1]**

[1]Department of Chemical Engineering, University of California, Santa Barbara; [2]Animal Production, Welfare and Veterinary Sciences, Harper Adams University, Newport, Shropshire, TF10 8NB, United Kingdom; [3]Enivronmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99352; [4]US DOE Joint Genome Institute, 2800 Mitchell Dr., Walnut Creek, CA 94598; and [5] Broad Institute of MIT and Harvard, Cambridge, MA 02142

http://omalleylab.weebly.com

Project Goals: The goal of this project is to engineer anaerobic gut fungi as novel platform organisms for biofuel production from plant material. To accomplish this goal, a panel of anaerobic fungi will be isolated from different herbivores and screened for their ability to degrade lignocellulose. The basic metabolic networks that govern lignocellulose hydrolysis within anaerobic fungi will also be determined, and models will be generated to describe how important enzyme groups are coordinated during breakdown. Using this information, genetic transformation strategies to manipulate gut fungi will be developed, which would endow them with enhanced functionality against a range of industrially relevant substrates. Collectively, this information will establish the molecular framework for anaerobic fungal hydrolysis, and will guide in the development of lignocellulosic biofuels.

To support renewable technologies, it is necessary to develop more efficient methods to extract sugars from crude plant biomass (lignocellulose). While plants contain cellulose that depolymerizes into fermentable sugars for microbial biofuel production, it is trapped within lignin, hemicellulose and other biopolymers that complicate its hydrolysis. To address this issue, one can turn to nature, particularly to microbes that routinely degrade plant biomass. Many large herbivores, such as cows and horses, harbor a consortium of microbes in their digestive tracts that convert recalcitrant biomass into sugars. Within this consortium, anaerobic gut fungi are the primary colonizers of plant material, and represent a rich source of biomass degrading enzymes. We have used transcriptomics to identify biomass degrading enzymes produced by the gut fungi and, moreover, have examined regulation patterns to determine how the suite of enzymes is tailored to the carbon source present.

By providing a pulse of a simple sugar during growth on un-pretreated biomass we have studied how the transcriptome remodels to match the perturbation. This reveals how quickly gut fungi

respond to alter the expression of biomass degrading enzymes when they are no longer necessary. Additionally, patterns of co-regulated transcripts may provide insight into the putative function of transcripts with unknown function. Cluster analysis of transcriptomic regulation data reveals co-regulated groups enriched in biomass degrading function, primarily consisting of cellulolytic and/or hemicellulolytic enzymes. Further work examining the transcriptome on a variety of substrates ranging in complexity, including glucose, cellobiose, cellulose, and biomass provides further insight to how these genes are regulated. Gene set enrichment analysis reveals not only how the expression of biomass degrading genes changes across these substrates, but also how the variety of biomass degrading enzymes changes. In the case of two isolated strains, *Neocallimastix californiae* and *Piromyces finnis*, there is a gradual increase in variety of biomass degrading enzymes with the increase in complexity of the carbon source. However, in the case of a third isolated strain, *Anaeromyces robustus*, GSEA reveals that cellobiose triggers the increase in expression of a wide variety of enzymes, not isolated to those involved in hydrolyzing cellobiose and cellulose. Further, we can examine the changes in expression of core metabolic proteins to determine how the fungal cells tailor their metabolism across these different conditions.

Recently, in partnership with the JGI, we have acquired genomic information for each of these isolates. Combining the regulation information with genomic localization, we can begin to identify potential promoters that control expression of these important genes. By investigating the DNA regions upstream of genes of similar functions and similar regulation patterns, we are currently working to identify consensus sequences that may be controlled by the same transcription factors. Together, this approach will enhance our efforts to develop new tools to genetically engineer the anaerobic fungi.

This Project is supported by the Office of Biological and Environmental Research through the DOE Office of Science (Early Career Program)

Engineering Synthetic Systems Inspired by Anaerobic Gut Fungi

Sean P. Gilmore[1*] (SeanGilmore@umail.ucsb.edu), Charles H. Haitjema,[1] John K. Henske,[1] Jessica A. Sexton,[1] Kevin V. Solomon,[1] Heather M. Brewer,[2] Samuel O. Purvine,[2] Aaron T. Wright,[2] Alan Kuo,[3] Stephen Mondo,[3] Asaf Salamov,[3] Kurt LaButti,[3] Michael K. Theodorou,[4] Igor Grigoriev,[3] and **Michelle A. O'Malley**[1]

[1]Department of Chemical Engineering, University of California, Santa Barbara; [2]Enivronmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99352; [3]US DOE Joint Genome Institute, 2800 Mitchell Dr., Walnut Creek, CA 94598; and [4]Animal Production, Welfare and Veterinary Sciences, Harper Adams University, Newport, Shropshire, TF10 8NB, United Kingdom

http://omalleylab.weebly.com

Project Goals: The overall goal of this project is to engineer novel anaerobes as platform organisms for biofuel production from plant material. To accomplish this goal, anaerobic consortia were isolated from different herbivores and screened for their ability to degrade lignocellulose. From these native consortia, we seek to assemble a "parts list" that comprise multi-enzyme fungal cellulosomes – this includes identification of the fungal scaffolding system, and cohesin domain, which bind to fungal dockerin-fused enzymes. Additionally, we aim to characterize native fungal-containing microbial consortia to "mix-and-match" fungi with other anaerobes to enhance biomass hydrolysis and tune bioproduction.

Anaerobic fungi in the hindgut of large herbivores are among the most robust organisms at degrading crude lignocellulose. Their remarkable cellulolytic capabilities have great potential for use in biomass breakdown and biofuel processing. Anaerobic fungi achieve cellulolytic efficiency through the production of large, multi-enzyme complexes called fungal cellulosomes. In isolation, anaerobic fungi metabolize some of the released sugars and convert them into fermentation products. In nature, however, they exist in a community with archaea, bacteria, and protozoa, which drastically alter the behavior of the fungi. By elucidating the parts responsible for efficient biomass degradation at both the protein and cellular level, we seek to replicate this efficiency in synthetic systems.

Fungal cellulosomes are similar to bacterial cellulosomes in that the protein-protein interactions are mediated through parts termed the dockerin and cohesin. However, many differences exist. The dockerin domains exist in tandem repeats and bear no species specificity like those in the bacterial systems. Furthermore, the exact sequence for the cohesin module has yet to be established. Through a combination of –OMICs approaches and traditional biochemical assays, a large putative scaffoldin molecule was identified. The scaffoldin was heterologously expressed and screened for interaction with recombinant dockerin through an ELISA. The K_D^{app} was determined using Equilibrium Surface Plasmon Resonance. A transcriptomic survey of dockerin domain-containing proteins revealed some degree of conservation in dockerin location on classes

of CAZymes. Using this observation, the dockerin domains were adapted to thermostable cellulases, demonstrating its applicability as a novel protein scaffolding systems and suggesting the possibility of synthetic cellulosomes for biomass degradation.

Additionally, anaerobic fungi have been shown to interact closely with methane producing archaea (methanogens). The methanogens siphon hydrogen and other metabolites from the fungi, allowing the fungi to more efficiently produce energy by increasing the flux through their hydrogenosomes. To further investigate this mechanism, native fungal/methanogen consortia were isolated from herbivore fecal materials. Consortia were maintained together and also separated into monocultures for comparison. Genomic sequencing revealed the presence of one fungus, two methanogens, and one bacterium in one consortium, which was stable under continuous passage for over 20 months. The consortium demonstrated faster and more complete degradation of cellulosic substrates, as well as a wider range of utilized substrates compared to the monocultured fungus alone. By introducing the methanogens into cultures of other well-characterized anaerobic fungi, stable synthetic co-cultures were established. These stable synthetic consortia demonstrated similar efficiency, and suggest a promising option for conversion of crude biomass into sustainable chemicals.

This Project is supported by the Office of Biological and Environmental Research through the DOE Office of Science (Early Career Program)

Title: A proteomic survey of diverse gut microbes

Ernesto S Nakayasu[1], Meagan C Burnett[1], Stephen J Callister[1], **Samuel H Payne[1]* (samuel.payne@pnnl.gov)**

1. Biological Sciences Division, Pacific Northwest National Laboratory, Richland WA

Project Goals: This project is focused on improving algorithms and methods for mass spectrometry data analysis of metaproteomics data. Recent advances in mass spectrometry and biological separations have dramatically increased the depth of proteomic discovery. Unfortunately, traditional computational workflows are in many cases preventing researchers from realizing these benefits for microbial communities. We propose to create a new generation of computational workflows to overcome the sensitivity limitations inherent in status quo data processing schemes.

To advance our ability to annotate tandem mass spectrometry data from microbial communities, our project has been developing algorithms to match spectra from metaproteomics experiments to a library of annotated spectra. With significant improvements having been achieved in the algorithms, the next step towards a fully functioning pipeline is to greatly expand our library of annotated spectra, specifically the diversity of the microbes present in the library. Consistent with our focus on biofuels and microbial communities that degrade plant feedstock, we are specifically targeting gut microbes such as those that live in cow rumen or beetle gut. Our first release of the PNNL Biodiversity Library[1] contained over 100 bacteria and archaea from 15 phyla, but had few organisms representing this important ecological niche. Therefore this year, we have collected global proteomics data from 10 organisms in firmicute and bacteriodetes and are in the process of identifying an additional 40 organisms for future data collection. Several organisms have been analyzed with a variety of complex media conditions to understand the metabolic adaptation to different nutrients. Dramatic advances in mass spectrometry instrumentation allow us to sample the proteome more deeply with less instrument time, enabling a greater survey of biodiversity. We have begun to analyze the data by annotating proteins into KEGG, both to look at expression across pathways, but also to help identify orthologs across species.

References:

1. The Pacific Northwest National Laboratory library of bacterial and archaeal proteomic biodiversity. Payne SH, Monroe ME, Overall CC, Kiebel GR, Degan M, Gibbons BC, Fujimoto GM, Purvine SO, Adkins JN, Lipton MS, Smith RD. Sci Data. (2015) 2:150041. doi: 10.1038/sdata.2015.41. PMID: 26306205

This project was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Early Career Research Program.

Progress towards defining a minimal parts list for the biodesign of Crassulacean acid metabolism (CAM): identification, characterization and ground-truthing of candidate CAM genes using the model species *Kalanchoë fedtschenkoi* and *K. laxiflora*

Susanna F. Boxall[1], Louisa V. Dever[1], Jack Davies[1], Richard Heaton[1], Nirja Kadu[1] and **James Hartwell**[1*] (james.hartwell@liverpool.ac.uk)

[1]Institute of Integrative Biology, University of Liverpool, UK

http://cambiodesign.org

Project Goals: The long-term goal of this project is to enhance the water-use efficiency (WUE) and adaptability to hotter, drier climates of species that normally perform C_3 photosynthesis by introducing crassulacean acid metabolism (CAM). Photosynthetic performance and WUE will be enhanced in *Arabidopsis* and *Populus* by: 1) defining the genetic basis of key CAM modules in both eudicot and monocot CAM species, 2) characterizing the regulation of 'carboxylation', 'decarboxylation', and 'inverse stomatal control' gene modules of CAM via loss-of-function studies in model CAM species, 3) deploying advanced genome engineering technologies to enable transfer of fully functional CAM modules into C_3 species and 4) analyzing the effects of these transgenic modules on 'stomatal control', CO_2 assimilation and transpiration rates, biomass yield, and WUE in *Arabidopsis* and *Populus*.

CAM plants are characteristic of arid, semi-arid and seasonally dry habitats, suggesting that a key driver for CAM evolution was likely to have been high temperatures and drought-prone conditions. CAM assists plants to achieve enhanced WUE by virtue of a temporal separation of primary and secondary CO_2 fixation in individual photosynthetic leaf mesophyll cells. Stomatal opening and associated primary atmospheric CO_2 fixation occurs during the cooler more humid dark period, thereby keeping transpirational water loss to a minimum. Dark CO_2 fixation is catalyzed by phosphoenolpyruvate carboxylase (PPC) in photosynthetic mesophyll cells. PPC has a much higher affinity for CO_2 than ribulose 1,5-bisphosphate carboxylase oxygenase (RuBisCO), and lacks oxygenase activity. CO_2 fixation by PPC generates oxaloacetate, which is converted to malate by malate dehydrogenase (MDH). Malate is stored in the vacuole as malic acid until dawn when it is decarboxylated by either NAD(P)-malic enzyme, or a combination of MDH and phosphoenolpyruvate carboxykinase, generating an internal supply of CO_2 inside the leaf. High internal [CO_2] is believed to signal stomatal closure in the light, minimizing water loss during the hottest and driest part of the 24 h cycle. The carbon dioxide from malate decarboxylation is refixed by RuBisCO in the Calvin-Benson cycle in the chloroplast, yielding sugars to fuel growth and reproduction. The high concentration of CO_2 that builds up inside the leaf during malate decarboxylation behind closed stomata minimizes the oxygenase activity of RuBisCO, thus preventing the inefficient side reaction of photosynthesis known as photorespiration. Temporal coordination of stomatal opening and primary CO_2 fixation in the dark, and stomatal closure and secondary refixation of CO_2 in the light is achieved through tight coupling of the entire CAM pathway to the central circadian clock. Clock-control allows dawn and dusk to be anticipated and CAM biochemical steps to be optimized to prevent futile cycling.

Forward engineering of CAM into non-CAM crop species in order to enhance their WUE requires a comprehensive knowledge of the minimal set of genes and proteins required for the efficient operation of CAM. Here we describe progress with functional genomics research that aims to define and characterize the complete CAM genetic blueprint from the model species *Kalanchoë fedtschenkoi* and *Kalanchoë laxiflora*. A draft assembly of the 246 Mbp *K. fedtschenkoi* genome has been assembled and annotated and recently further improved through the addition of 100X coverage of PacBio long reads. Quantitative RNA-seq analysis of light/ dark time course samples from C_3 and CAM leaves identified candidate CAM-associated genes that increase in transcript abundance concomitant with CAM. Many of these CAM-induced genes also undergo robust oscillations in transcript abundance over the light/ dark cycle, consistent with a role in the circadian optimization of the pathway. The RNA-seq data has guided the reconstruction of a comprehensive model of CAM for which candidate CAM-recruited genes are allocated to each step in the pathway. This in turn has allowed targeted RNAi gene silencing and over-expression approaches to be applied to each candidate CAM gene through the generation of stable transgenic lines of *K. fedtschenkoi* and *K. laxiflora*. Detailed analysis of CAM-associated phenotypes in the transgenic lines is revealing which genes are essential for efficient CAM, and which genes are dispensable (1). We will present data on the phenotypic characterization of RNAi lines of *K. laxiflora* lacking key CAM genes. Several lines fail to fix atmospheric CO_2 in the dark period, and the phenotypic consequences of this will be described.

In addition, data will be presented from quantitative Illumina RNA-seq analysis of a light/ dark timecourse of transcript levels in epidermal peels (enriched for guard cells) and separated mesophyll tissue from CAM leaves of *K. fedtschenkoi*. This data provides unique insights into the temporal coordination of known guard cell signaling genes including those known to be involved in CO_2-, light-, and ABA-dependent regulation of stomatal aperture. Several of these guard cell-signaling genes display a 6 to 12 h shift in the timing of their transcript abundance peak in CAM leaves of *K. fedtschenkoi* when compared to the temporal regulation of the orthologous transcripts in C_3 leaves of Arabidopsis. These unique discoveries are providing unprecedented insights into the gene regulatory networks that associate with reverse stomatal opening required for CAM.

Finally, an overview will be presented of work on the characterisation of light/ dark- and circadian clock-controlled transcription factors (TFs) that are induced coincident with CAM. These TFs are candidates for controlling key regulatory steps in the circadian coordination and optimization of the daily CAM cycle, potentially linking the central circadian clock to CAM as part of a CAM-specific clock output pathway. Highlights will be presented from RNAi silencing of novel CAM-associated TFs, and progress towards ChIP-seq analysis of TF genomic targets will be summarized.

References

1) Dever LV, Boxall SF, Knerova J, Hartwell J (2015) Transgenic perturbation of the decarboxylation phase of crassulacean acid metabolism alters physiology and metabolism but only has a small effect on growth. Plant Physiol. **167**: 44-59.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Genomic Science Program under Award Number DE-SC0008834.

Engineering Crassulacean Acid Metabolism (CAM) to Improve Water-use Efficiency of Bioenergy Feedstocks

John C. Cushman[1]* (jcushman@unr.edu), Rebecca Albion[1], Jungmin Ha[1], Sung Don Lim[1], Bernard W. M. Wone[1], Won Cheol Yim[1], Travis Garcia[1], Pradeep K. Yerramsetty[1], Juli Petereit[2], Richard Tillett[2], Karen A. Schlauch[2], Xiaohan Yang[3], Anne M. Borland[3,4], Alistair Leverett[4], Henrique C. De Paoli[3], Rongbin Hu[3], Degao Liu[3], Kaitlin Palla[3,5], Robert C. Moseley[3,5], Timothy J. Tschaplinski[3], David J. Weston[3], Gerald A. Tuskan[3], Jay Chen[3], James Hartwell[6], Louisa V. Dever[6], Susanna F. Boxall[6], Jade Waller[6], Jack Davies[6], Phaitun Bupphada[6], Nirja Kadu[6], Hong Guo[7], Hao-Bo Guo[7], Paul Abraham[8], and Robert Hettich[8].

[1]Department of Biochemistry and Molecular Biology, University of Nevada, Reno. [2]Nevada Center for Bioinformatics, University of Nevada, Reno; [3]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [4]School of Biology, Newcastle University, Newcastle upon Tyne; [5]Bredesen Center, University of Tennessee, Knoxville, TN; [6]Department of Plant Sciences, Institute of Integrative Biology, University of Liverpool; [7]Department of Biochemistry & Cellular and Molecular Biology, University of Tennessee, Knoxville; [8]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN.

http://cambiodesign.org

Project Goals: The long-term goal of this project is to enhance the water-use efficiency (WUE) and adaptability to hotter, drier climates of species that normally perform C_3 photosynthesis by introducing crassulacean acid metabolism (CAM). Photosynthetic performance and WUE will be enhanced in *Arabidopsis* and *Populus* by: 1) defining the genetic basis of key CAM modules in both eudicot and monocot CAM species, 2) characterizing the regulation of 'carboxylation', 'decarboxylation', and 'inverse stomatal control' gene modules of CAM via loss-of-function studies in model CAM species, 3) deploying advanced genome engineering technologies to enable transfer of fully functional CAM modules into C_3 species and 4) analyzing the effects of these transgenic modules on 'stomatal control', CO_2 assimilation and transpiration rates, biomass yield, and WUE in *Arabidopsis* and *Populus*.

In order to meet the grand challenges of overcoming the negative effects of global climate change on crop productivity, an increased reliance on crassulacean acid metabolism (CAM) crops might serve as a useful component of an integrated scheme to develop sustainable agroecosystems for dryland reclamation to provide both bioenergy feedstocks and ecosystem services (1, 2). However, traditional food and bioenergy crops with greater heat and drought durability and greater water-use efficiency (WUE) will also be crucial for sustainable biomass production systems in the future. Thus, one approach to increase crop WUE is to move the water-wise photosynthetic machinery of CAM into C_3 food and bioenergy crops (2, 3). CAM features inverse stomatal behavior, in which stomata are open at night for CO_2 uptake when evapotranspiration rates are reduced compared with the daytime and closed during all or part of the day, thereby maximizing WUE. CAM also features a temporal separation of nocturnal CO_2 uptake and fixation by phospho*enol*pyruvate carboxylase (PEPC), which leads to the formation of C_4 organic acids that are stored in the vacuole. The subsequent decarboxylation of these

organic acids during the day releases CO_2 and concentrates it around ribulose-1-5-bisphosphate carboxylase/oxygenase (RUBISCO), suppressing photorespiration, while resulting in carbohydrate production *via* the C_3 Calvin–Benson cycle.

Comparative transcriptomic and genomic sequencing projects were used to identify CAM-related genes by examining developmentally regulated or water-deficit stress-inducible gene expression patterns from obligate and facultative CAM species, respectively (3). Clues to the transcriptional control of CAM expression networks are being derived from the analysis of candidate *cis*-regulatory elements and cognate transcription factors controlling circadian and mesophyll expression patterns of CAM genes. For example, yeast one-hybrid screens using candidate consensus *cis*-regulatory elements and proximal 5' regulator regions of the phospho*enol*pyruvate carboxylase (*Ppc*) gene as bait sequences have been completed. A collection of candidate transcription factors including AP2 domain, homeobox-leucine zipper domain, Myb-like, C2H2-type zinc finger, and bZIP family transcription factors have been identified to date.

In addition, loss-of-function studies of individual enzymes, metabolite transporters, and regulatory proteins or transcription factors are being used to provide critical insights into the basic genetic requirements for CAM. For example, RNAi-mediated silencing of mitochondrial NAD-malic enzyme and cytosolic/plastidic pyruvate orthophosphate dikinase not only impaired nocturnal CO_2 uptake, but also reduced the circadian clock-controlled phosphorylation of PPC (4). Other studies using RNAi lines of *K. fedstchenkoi* have shown that the route of nocturnal starch degradation is a key point of divergence between C_3 and CAM. Data have been obtained which indicate that phosphorolytic degradation of starch produces substrate for production of PEP, while the hydrolytic production and nocturnal export of glucose from the chloroplast primarily directs substrate towards provision of sucrose for growth. Such information is critical for the selection of genes and gene networks for reconstructing and validating of carboxylation and decarboxylation modules.

Tissue succulence engineering in C_3 photosynthesis plants may be a key anatomical attribute for enhancing the efficient operation of engineered CAM in C_3 photosynthesis species. Thus, increased tissue succulence has been accomplished in the C_3 photosynthesis model species, *Arabidopsis thaliana*, in order to increased mesophyll cell size for increased malate storage capacity in the vacuole and reduced intercellular air space (IAS) to limit the diffusion of CO_2 out of the leaf during the day for refixation by ribulose 1,5-bisphosphate carboxylase oxygenase (RUBISCO). For CAM Biodesign, strategies for stacking of multi-gene circuits with appropriate circadian and drought-inducible expression patterns for reconstitution of CAM in *A. thaliana* have been developed using the Gibson assembly process. Mesophyll-specific expression patterns of transgenes were verified by promoter::GUS-LUC reporter constructs and will be tested for circadian expression patterns. Subcellular localization of targeted enzymes and transporters of the carboxylation and decarboxylation modules were verified by GFP-fusion protein localization studies in stably transformed *A. thaliana* lines. To facilitate the construction of gene circuits for CAM biodesign, a novel platform was developed for high-throughput assembly of DNA parts (5). Cutting-edge genome editing tools are being applied to functional genomics research and CAM engineering (6). Design and construction of plant-specific vector systems for Gibson isothermal assembly of gene circuits containing 9 and 15 genes CAM-specific genes has been completed. Lastly, phenotypic testing of various *Populus* varieties with respect to transformability, leaf anatomy, stomatal responsiveness to CO_2, and non-structural carbohydrate resources for nocturnal CO_2 fixation has revealed potentially suitable candidates for CAM engineering.

References

1) Cushman JC, Davis SC, Yang X, Borland AM. (2015) Development and use of bioenergy feedstocks for semi-arid and arid lands. J. Exp. Bot. 66:4177–4193.
2) Yang X, Cushman JC, Borland AM, Edwards EJ, Wullschleger SD, Tuskan GA, Owen NA, Griffiths H, Smith JAC, De Paoli HC, Weston DJ, Cottingham R, Hartwell J, Davis SC, Silvera K, Ming R, Schlauch KA, Abraham P, Stewart JR, Guo H-B, Albion RA, Ha J, Lim SD, Wone BWM, Yim WC, Garcia T, Mayer JA, Petereit J, Nair SS, Casey E, Hettich RL, Ceusters J, Ranjan P, Palla KJ, Yin H, Reyes-García C, Andrade JL, Freschi L, Beltran JD, Dever LV, Boxall SF, Waller J, Davies J, Bupphada P, Kadu N, Winter K, Sage RF, Aguilar CN, Schmutz J, Jenkins J, Holtum, JAC. (2015) A roadmap for research on crassulacean acid metabolism to enhance sustainable food and bioenergy production in a hotter, drier world. New Phytologist. 207*:* 491–504.
3) Borland AM, Hartwell J, Weston D, Schlauch KA, Tschaplinski TJ, Tuskan GA, Yang X, Cushman JC. (2014) Engineering Crassulacean acid metabolism to improve water-use efficiency. Trends Plant Sci. 19: 327–338.
4) Dever LV, Boxall SF, Knerova J, Hartwell J. (2015) Transgenic perturbation of the decarboxylation phase of crassulacean acid metabolism alters physiology and metabolism but only has a small effect on growth. Plant Physiol. 167: 44–59.
5) De Paoli HC, Tuskan GA, Yang X (2016) An innovative platform for quick and flexible joining of assorted DNA fragments. Scientific Reports. 6: 19278.
6) Liu D, Hu R, Palla KJ, Tuskan GA, Yang X (2016) Advances and perspectives on the use of CRISPR/Cas9 systems in plant genomics research. Current Opinion in Plant Biology. *In press.*

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Genomic Science Program under Award Number DE-SC0008834. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under Contract Number DE–AC05–00OR22725.

High Quality *De Novo* Genome Sequencing of the Common Ice Plant (*Mesembryanthemum crystallinum* L.) - a Functional Genomics Resource for Crassulacean Acid Metabolism (CAM) Biodesign

Won Cheol Yim[1] (wyim@unr.edu), Sung Don Lim[1], Bernard W.M. Wone[1], Jungmin Ha, Rebecca L. Albion[1] Hengfu Yin[2], Xiaohan Yang[2], and John C. Cushman[1]

[1]Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV
[2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

http://cambiodesign.org

Project Goals: Crassulacean acid metabolism (CAM) is a specialized mode of photosynthesis that exploits a temporal CO_2 pump with nocturnal CO_2 uptake to improve the water-use efficiency (WUE) and the adaptability of plants to hotter, drier climates. The long-term goal of the CAM Biodesign project is to introduce CAM into C_3 photosynthesis plants, such as *Arabidopsis* and *Populus,* and thereby enhance WUE and photosynthetic performance. Major project goals include: 1) defining the genetic basis of key CAM modules in both eudicot and monocot CAM species; 2) characterizing the regulation of 'carboxylation', 'decarboxylation', and 'inverse stomatal control' gene modules of CAM using a wide variety of functional genomic approaches including loss-of-function studies and transcriptome profiling in model CAM species; 3) deploying advanced genome engineering technologies to enable transfer of fully functional CAM modules into C_3 species; and 4) analyzing the effects of these transgenic modules on 'stomatal control', CO_2 assimilation and transpiration rates, biomass yield, and WUE in *Arabidopsis* and *Populus*. Successful transfer of CAM into target C3 photosynthesis species might allow expansion of biofuel feedstock product into water-limited, semi-arid or seasonally dry environments.

The common ice plant (*Mesembryanthemum crystallinum* L., Azoaceae, Caryophyllales) is an important halophytic and facultative crassulacean acid metabolism (CAM) model in which CAM can be induced by salinity or water-deficit stress (1, 2). To improve our understanding of the molecular genetic basis of CAM, water-use efficiency, and salinity tolerance in plants, a high-quality, annotated genome assembly was generated for ice plant using two different genome sequencing platform, Illumina and PacBio, resulting in 95,569,388,632 and 33,865,183,009 bases, with read depths of coverage of 254 X and 90 X, respectively. We assembled the reads in two steps, first with ABySS, SparseAssembler, String Graph Assembler, and SOAPdenovo2 for Illumina reads and SMRTanalysis for PacBio reads, independently. The resulting assemblies were compared with each assembler output with the Illumina mate-pair reads used for scaffolding. The scaffold from the PacBio was combined with the Illumina assembly resulted in the longest scaffolds of 11.6 Mb, N50 & NG50 of 3.2 Mb, 957 scaffolds, and a total size is 382.6 Mb, which is close to the ice plant genome size estimate of 390 Mb from flow cytometry analysis. This accuracy and completeness of the genome assembly was improved further using BioNano Genomics Irys® single-molecule next-generation mapping (NGM) with an increase in scaffold N50 from 3.2 Mb to 4.0 Mb. A total of 44,208 open reading frames were annotated using this genome assembly. This high-quality draft genome assembly will enable a comprehensive identification of genes required for the operation of CAM and for the adaptation

of this halophytic species to salinity stress conditions. This high-quality ice plant genome sequence will also serve as a reference for comparative CAM genomics, tracing the evolution of CAM, and CAM Biodesign efforts aimed at introducing the improved WUE of CAM into bioenergy crops.

References

1. Borland AM, Hartwell J, Weston D, Schlauch KA, Tschaplinski TJ, Tuskan GA, Yang X, Cushman JC. (2014) Engineering Crassulacean acid metabolism to improve water-use efficiency. Trends Plant Sci. 19: 327-338.

2. Yang X, Cushman JC, Borland AM, Edwards EJ, Wullschleger SD, Tuskan GA, Owen NA, Griffiths H, Smith JAC, De Paoli HC, Weston DJ, Cottingham R, Hartwell J, Davis SC, Silvera K, Ming R, Schlauch KA, Abraham P, Stewart JR, Guo H-B, Albion RA, Ha J, Lim SD, Wone BWM, Yim WC, Garcia T, Mayer JA, Petereit J, Nair SS, Casey E, Hettich RL, Ceusters J, Ranjan P, Palla KJ, Yin H, Reyes-García C, Andrade JL, Freschi L, Beltran JD, Dever LV, Boxall SF, Waller J, Davies J, Bupphada P, Kadu N, Winter K, Sage RF, Aguilar CN, Schmutz J, Jenkins J, Holtum, JAC. (2015) A roadmap for research on crassulacean acid metabolism to enhance sustainable food and bioenergy production in a hotter, drier world. New Phytologist. 207*:* 491–504.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Genomic Science Program under Award Number DE-SC0008834. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under Contract Number DE–AC05–00OR22725.

Engineering Tissue Succulence to Improve Water-use Efficiency of Bioenergy Feedstocks

Sung Don Lim[1], (sunglim@unr.edu), Won Cheol Yim[1], and John C. Cushman[1]

[1]Department of Biochemistry and Molecular Biology, University of Nevada, Reno

http://cambiodesign.org

Project Goals: Crassulacean acid metabolism (CAM) is a specialized mode of photosynthesis that exploits a temporal CO_2 pump with nocturnal CO_2 uptake to improve the water-use efficiency (WUE) and the adaptability of plants to hotter, drier climates. The long-term goal of the CAM Biodesign project is to introduce CAM into C_3 photosynthesis plants, such as *Arabidopsis* and *Populus,* and thereby enhance WUE and photosynthetic performance. Major project goals include: 1) defining the genetic basis of key CAM modules in both eudicot and monocot CAM species; 2) characterizing the regulation of 'carboxylation', 'decarboxylation', and 'inverse stomatal control' gene modules of CAM using a wide variety of functional genomic approaches including loss-of-function studies and transcriptome profiling in model CAM species; 3) deploying advanced genome engineering technologies to enable transfer of fully functional CAM modules into C_3 species; and 4) analyzing the effects of these transgenic modules on 'stomatal control', CO_2 assimilation and transpiration rates, biomass yield, and WUE in *Arabidopsis* and *Populus*. Successful transfer of CAM into target C3 photosynthesis species might allow expansion of biofuel feedstock product into water-limited, semi-arid or seasonally dry environments.

Introducing the CAM photosynthetic machinery into C_3 plants (CAM biodesign) is expected to confer improved WUE in order to assist in the expansion of crop production into semi-arid regions (1, 2, 3). Tissue succulence is an important plant trait that allows plants to withstand long episodes of drought. Some degree of tissue succulence is typically correlated with the optimal performance of CAM. Thus, for CAM biodesign, tissue succulence engineering in C_3 photosynthesis plants may be a key anatomical attribute for enhancing the efficient operation of engineered CAM in C_3 photosynthesis species. Specifically, increased tissue succulence is expected to afford increased mesophyll cell size in order to increase malate storage capacity in the vacuole and to reduce intercellular air space (IAS) to limit the diffusion of CO_2 out of the leaf during the day for refixation by ribulose 1,5-bisphosphate carboxylase oxygenase (RUBISCO). A method for the genetic engineering of tissue succulence was developed involving the overexpression of a modified basic helix-loop-helix (bHLH) transcription factor from *Vitis vinifera*. The engineered tissue succulence in *Arabidopsis* resulted in up to a 2.2-fold increase plant leaf fresh weight, a 2.4-fold increase in leaf dry weight, root, flower biomass, and seed production relative to controls. The increased cell size does not appear to be associated with an increase in ploidy level of the plants. The increased size of all organs also resulted in up to a 1.6-fold increase in leaf thickness, up to a 1.8-fold increase in leaf succulence, up to a 2.9-fold increase in leaf water amount, and up to a 37% reduction in intracellular air space (IAS). This reduction in IAS is a key feature of the innovation because it limits CO_2 diffusion out of the leaf and is thought to be critical for recapture of photorespiratory CO_2 loss and CO_2 recapture by CAM during the day. Importantly, plants with engineered succulent exhibited up to a 35% increase in seed number per silique, up to a 21% increase in seed area, up to a 38% increase in

100-seed weight, and up to a 2.6-fold increase in overall seed yield per plant. Lastly, the engineered succulent plants displayed greater tolerance to salinity and osmotic stress, and water-deficit stress and greater survival and regrowth following acute water-deficit stress likely due to their ability to retain and store water within their tissues.

References

1. Borland AM, Hartwell J, Weston D, Schlauch KA, Tschaplinski TJ, Tuskan GA, Yang X, Cushman JC. (2014) Engineering Crassulacean acid metabolism to improve water-use efficiency. Trends Plant Sci. 19:327-338.

2. Borland AM, Wullschleger SD, Weston DJ, Tuskan GA, Hartwell J, Yang X, Cushman JC. (2015) Climate-resilient agroforestry: Physiological responses to climate change and engineering of crassulacean acid metabolism (CAM) as a mitigation strategy. Plant Cell & Environ. 38:1833–1849.

3. Yang X, Cushman JC, Borland AM, Edwards EJ, Wullschleger SD, Tuskan GA, Owen NA, Griffiths H, Smith JAC, De Paoli HC, Weston DJ, Cottingham R, Hartwell J, Davis SC, Silvera K, Ming R, Schlauch KA, Abraham P, Stewart JR, Guo H-B, Albion RA, Ha J, Lim SD, Wone BWM, Yim WC, Garcia T, Mayer JA, Petereit J, Nair SS, Casey E, Hettich RL, Ceusters J, Ranjan P, Palla KJ, Yin H, Reyes-García C, Andrade JL, Freschi L, Beltran JD, Dever LV, Boxall SF, Waller J, Davies J, Bupphada P, Kadu N, Winter K, Sage RF, Aguilar CN, Schmutz J, Jenkins J, Holtum, JAC. (2015) A roadmap for research on crassulacean acid metabolism to enhance sustainable food and bioenergy production in a hotter, drier world. New Phytol. 207:491–504.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Genomic Science Program under Award Number DE-SC0008834. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under Contract Number DE–AC05–00OR22725.

The *Kalanchoe* genome -- An important model for systems biology and synthetic biology of crassulacean acid metabolism

Xiaohan Yang[1](yangx@ornl.gov), John C. Cushman[2], Rongbin Hu[1], Jerry Jenkins[3], Jeremy Schmutz[3,4], Gerald A. Tuskan[1], Hengfu Yin[1], Shengqiang Shu[4], Daniel Rokhsar[4], David Goodstein[4], Uffe Hellsten[4], Anne M. Borland[1,5], Ray Ming[6,7], Robert Van Buren[8], Ching Man Wai[6], Haibao Tang[7], Paul Abraham[9], Hao-Bo Guo[10], Hong Guo[10], Rebecca L. Albion[2], Travis Garcia[2], Jungmin Ha[2], Sung Don Lim[2], Jesse A. Mayer[2], Bernard W. M. Wone[2], Won Cheol Yim[2], Daniel A. Jacobson[1], Deborah A. Weighill[1], James Hartwell[11], Xueqiong Li[12], Yanbin Yin[12]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV; [3]HudsonAlpha Institute for Biotechnology, Huntsville, AL; [4]US Department of Energy Joint Genome Institute, Walnut Creek, CA; [5]School of Biology, Newcastle University, Newcastle, UK; [6]Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL; [7]FAFU and UIUC-SIB Joint Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fuzhou, China; [8]Donald Danforth Plant Science Center, St. Louis, Missouri; [9]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [10]Department of Biochemistry & Cellular and Molecular Biology, University of Tennessee, Knoxville, TN; [11]Department of Plant Sciences, University of Liverpool, Liverpool, UK; [12]Department of Biological Sciences, Northern Illinois University, DeKalb, Illinois

http://cambiodesign.org/

Project Goals: Crassulacean acid metabolism (CAM) is a specialized mode of photosynthesis that features a temporal CO_2 pump with nocturnal CO_2 uptake, facilitates increased water-use efficiency (WUE), and enables CAM plants to inhabit water-limited semi-arid or seasonally dry environments. CAM provides an excellent opportunity for engineering both enhanced WUE and photosynthetic performance into bioenergy crops. This project has two main goals: 1) to identify the CAM-associated genes and gene networks using systems biology approaches and 2) to engineer CAM gene modules into C_3 species using synthetic biology approaches. The success of the project could allow biomass production on semi-arid, abandoned, or marginal agricultural lands.

Kalanchoe laxiflora is an important model species for systems biology research to understand the molecular basis of the CAM pathway (Yang *et al*. 2015) and identify the key genes necessary for engineering CAM into C_3 species. *K. laxiflora* features a relatively small genome (~256 Mb), an established transformation system and a short life cycle. Recently, we constructed a high-

quality *K. laxiflora* genome assembly consisting of 1,324 scaffolds, with a total length of ~256 Mb and an N50 of 2.45 Mb. A total of 30,964 genes were annotated in this genome assembly. Our comparative analysis of the *K. laxiflora* genome and 24 other plant genomes, including two CAM species, three C_4 photosynthesis species and 19 C_3 photosynthesis species, revealed CAM-specific orthologous gene groups that were shared among the three CAM species [*K. laxiflora*, *Ananas comosus* (pineapple) and *Phalaenopsis equestris* (orchid)] but absent in C_3 or C_4 photosynthesis species. We analyzed the synteny between the *K. laxiflora* genome and several other plant genomes and found that the *K. laxiflora* genome has a distinctive whole-genome duplication history. Based on the comparative analysis of day-night time-courses of gene expression between CAM and C_3 photosynthesis genes, we identified several genes that are involved in the CAM pathway in *Kalanchoe*. We also performed an in-depth analysis of selected gene families relevant to CAM physiology and determined that positive selection has contributed to the molecular evolution of CAM plants. The results highlight the potential of the *K. laxiflora* genome as a model for CAM genomics research. The CAM-related genes identified in the *K. laxiflora* genome provide a solid foundation for the ongoing effort to engineer CAM into C_3 photosynthesis bioenergy crops.

References

Yang X, Cushman JC, Borland AM, Edwards EJ, Wullschleger SD, Tuskan GA, *et al*. (2015) A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. New Phytologist 207: 491-504.

This material is based upon work supported by the Department of Energy, Office of Science, Genomic Science Program (under award number DE-SC0008834). Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the US Department of Energy (under contract number DE-AC05-00OR22725).

Crassulacean Acid Metabolism (CAM) Informatics Tools

Karen A. Schlauch[1*] (schlauch@unr.edu), Robert C. Moseley[2,3], Juli Petereit[4], David Weston[2], Xiaohan Yang[2], and **John C. Cushman[1]**.

[1]Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV; [2]Biosciences Division, Oak Ridge National Laboratories, Oak Ridge, TN; [3]Bredesen Center, University of Tennessee, Knoxville, TN; [4]Department of Biomedical Engineering, University of Nevada, Reno, NV.

http://cambiodesign.org/

Project Goals: The long-term goal of this research is to enhance the water-use efficiency (WUE) and adaptability to hotter, drier climates of species that normally perform C_3 photosynthesis by introducing crassulacean acid metabolism (CAM). Four major objectives will be pursued to enhance photosynthetic performance and WUE in *Arabidopsis* and *Populus*: 1) Define the genetic basis of key CAM modules in eudicot and monocot CAM species *via* co-expression network models; 2) characterize the regulation of 'carboxylation', 'decarboxylation', and 'inverse stomatal control' modules of CAM using comparative genomics, network models, and loss-of-function testing; 3) deploy advanced genome engineering technologies to enable stacking of a large number of transgenes into a single genomic locus to improve transgene persistence and transfer fully functional 'carboxylation' and 'decarboxylation' modules from CAM species to C_3 species that can accommodate overnight malic acid storage in the vacuole; and 4) analyze the effects of these transgenic modules on 'stomatal control', CO_2 assimilation and transpiration rates, biomass yield, and WUE in *Arabidopsis* and *Populus*.

CAM is a specialized mode of photosynthesis that provides increased plant water-use efficiency (WUE). This adaptation has prompted research to examine the genetic and metabolic networks governing CAM physiology as part of a biodesign strategy to produce bioenergy feedstocks with enhanced WUE. To assist this effort, two informatics tools were developed: (1) **petal** is a co-expression network modeling and analysis program offering novel analytical approaches for CAM-specific whole-omics data; (2) **CAMbase** is an integrative resource platform where tools are deployed for intuitive analysis and visualization of genomic data.

petal is a whole-omics co-expression network analysis algorithm that offers two approaches not currently provided by most other available methods. Firstly, as most high-throughput expression data are not normally distributed, network models generated by **petal** can be based on either parametric and non-parametric similarity measures to provide more statistically sound network construction. Secondly, as most biological networks are known to have approximate scale-free and small-world structures, **petal** generates network models following these two properties. By allowing the network construction to be governed by these

properties, our networks are guaranteed to be both scale-free and small-world without user intervention (e.g., specifying a threshold/cut-off value). Connectivity structures in the co-expression network models such as cliques, modules, clusters, and paths represent gene or protein interactions. Cliques are of particular interest, as their structure represents more robust connectivity based upon their unique mathematical properties. **petal** offers downstream network analysis by identifying and examining the cliques in the networks, which provide putative functional groupings within the network. The petal system is implemented in **R** and easily processes whole-omics data.

CAMbase is an integrative platform in which tools are deployed in an infrastructure for intuitive analysis and visualization of genomic data. Currently, the database includes high-resolution, time-course RNA-seq data for *Kalanchoe laxiflora*, *Arabidopsis thaliana*, and *Agave americana*, as well as reference genomes for *Kalanchoe laxiflora*, orchid, and pineapple. Additionally, CAMbase includes gene ontology and homologs identified from other plant genomes for available CAM plants. Jbrowse and Apollo are available to explore genome features and annotation updating, respectively, of available CAM genomes. CAMbase also includes an eFP browser for comparative analysis of omics time-course data. Current progress of CAMbase development will be presented with a focus on the implementation of tools (e.g., eFP browser, Jbrowse, Apollo) for newly sequenced CAM model species. CAMbase data are being formatted in a KBase compliant manner, with the goal of pushing novel RNA-seq and genome data into the DOE-Kbase cyberinfrastructure for downstream modeling applications related to feedstock biodesign.

References

1. Borland AM, Hartwell J, Weston D, Schlauch KA, Tschaplinski TJ, Tuskan GA, Yang X, Cushman JC. (2014) Engineering crassulacean acid metabolism to improve water-use efficiency. Trends Plant Science 19: 327-338.
2. Petereit J, Harris, FC Jr, and Schlauch K, **petal**: A Novel Co-Expression Network Modeling System, Proceedings of the 2015 IEEE International Conference on Bioinformatics and Biomedicine (IEEE BIBM 2015) pp 234-241, November 9-12, 2015, Washington, DC.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Genomic Science Program under Award Number DE-SC0008834. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under Contract Number DE–AC05–00OR22725.

Expanding the breeder's toolbox for perennial grasses: Sequencing the genome of *Brachypodium sylvaticum* to establish a tractable perennial grass model system and the identification of transgenes conferring tolerance to abiotic stress.

Sean P. Gordon[1], Lifeng Liu[1,2], Nir Sade[3], Amy Cartwright[1], Jerry Jenkins[1,4], Jeremy Schmutz[1,4], Brian Tarape Hernandez[2], **Christian Tobias[2], Roger Thilmony[2], Eduardo Blumwald[3*] (eblumwald@ucdavis.edu), John Vogel[1,5]**

[1] DOE Joint Genome Institute;[2]USDA-ARS Western Regional Research Center;[3]UC Davis Department of Plant Sciences;[4]HusonAlpha Institute for Biotechnology;[5]UC Berkeley Department of Plant and Microbial Biology

Project Goals: The project aims at using a systems-based approach to develop new breeding tools for perennial grasses using the perennial grass *Brachypodium Sylvaticum* as a model, and apply these tools towards the improvement of switchgrass (*Panicum virgatum* L.). The objectives of the research project are: 1. Develop a CENH3-based method for creating doubled haploid grasses and demonstrate it in switchgrass; 2. Use the model perennial grass *Brachypodium sylvaticum* to identify combinations of transgenes that confer tolerance to multiple abiotic stresses; 3. Develop a gene containment system to minimize gene flow from transgenic switchgrass; 4. Genome sequencing of *B. Sylvaticum*

Most of our knowledge of grass biology and responses to abiotic stress is based on studies of annual grasses like rice, wheat and barley. However, most of the grasses being developed as biomass crops are perennial. Annuals and perennials differ in many important physiological and developmental aspects, some of which that may be particularly relevant to stress tolerance. Unfortunately, the perennial grasses being developed as biomass crops (e.g. switchgrass and Miscanthus) are difficult experimental subjects due to their large size, long generation time, complicated genetics and large genomes. Thus, there is a pressing need for a tractable perennial grass model to study areas like abiotic stress tolerance and, in particular, to rapidly test transgenic approaches before moving into biomass crops.

We previously established key tools that allow *Brachypodium sylvaticum* to be used as perennial model grass including a highly efficient transformation protocol and inbred lines. Missing from this list is a high quality genome sequence, a prerequisite for a modern model organism. To fill this need, we sequenced the genome of *B. sylvaticum* to ~80x depth using PacBio long-read technology. The current assembly contains 358Mb of sequence in 1,118 contigs with a N50 contig length of 874Kb. The contigs are being ordered and orientated into high quality chromosome-level assemblies using a high-density genetic map created from 288 F2 individuals. A deep RNA-Seq expression atlas of 16 different tissues and developmental stages will aid gene annotation of the chromosomal-level assembly. A replicated set of another 92 samples, tested under a variety of abiotic stresses, is also being probed by RNA-Seq to identify stress responsive genes and their associated gene regulatory networks in this model perennial grass. Matched RNA-Seq samples have already been generated for the small annual grass, *Brachypodium distachyon*, allowing a unique comparison of stress response between closely related annual and perennial species.

We completed the transformation of *B. Sylvaticum* with constructs containing 33 genes shown to be associated with enhanced abiotic stress tolerance and source-sink relationships in monocots. Single and/or combination of these genes were overexpressed using either constitutive or stress inducible promoter. We have generated 5-7 single copy insert homozygous T2 lines for all constructs. The response of those transgenes lines to water deficit stress, salinity and a combination of stresses is being tested. Preliminary results revealed a higher biomass in the transgenic plants grown under stress. Our research indicates a great potential for the development of grasses with improved performance and yield under abiotic stress.

Transgene Containment in Perennial Grasses: Ablation of Transgenic Pollen in *Brachypodium sylvaticum* and *Panicum virgatum*.

Ray Collier[1], Bryan Tarape Hernandez[1], **Eduardo Blumwald**[2], **John Vogel**[3], **Christian Tobias**[1] **Roger Thilmony**[1*] (Roger.Thilmony@ars.usda.gov)

[1]USDA-ARS-WRRC, Crop Improvement and Genetics Research Unit, Albany, CA
[2]Department of Plant Sciences, University of California, Davis, CA
[3]DOE Joint Genome Institute, Walnut Creek, CA

Project Goal:
Develop a gene containment system to minimize transgene flow from transgenic switchgrass.

Interest in the use of perennial grasses, especially those utilizing C_4 metabolism, for biofuel production is increasing because of the low input cost and long-term land use benefits afforded by these plants relative to their annual counterparts. Efforts reliant on biotechnology for the improvement of stress tolerance and biomass production in species like switchgrass (*Panicum virgatum*) must consider that it is an obligate outcrossing, wind-pollinated species native to North America. Thus, a means to control transgene escape to wild plant populations is needed. We are developing a transgene containment system for perennial grasses to address this concern. We are using the model perennial grass *Brachypodium sylvaticum* and switchgrass to evaluate the utility of novel transformation constructs to block transmission of transgenes via pollen. Using *Agrobacterium*-mediated transformation, we generated transgenic plants that express *barnase* under the control of four rice pollen-specific promoters (*PS1*, *PS2*, *PS3* and *OsGEX2*). Multiple independent transgenic lines for each construct have been evaluated by pollen staining and genetic segregation analyses. Alexander's staining revealed that, relative to wildtype plants, >50% of the pollen collected from putative containment T_0 *B. sylvaticum* lines was dead or severely deformed. Analysis of selfed T_1 progeny showed that transgene heritability was 1:1, consistent with the expected segregation frequency for a male lethal gene, supporting the conclusion that successful ablation of transgenic pollen was achieved in these *Brachypodium sylvaticum* transgenic plants. Initial work with *PS2-barnase* transgenic switchgrass suggests that approximately 50% of the pollen produced is inviable. We have generated numerous independent transgenic switchgrass plants with each of the four pollen ablation constructs and will be examining their effectiveness at mediating transgene containment.

Uncovering genetic mechanisms for biological conversion efficiency in *Brachypodium distachyon*

Scott J. Lee[1]*, Yu Cui[2], Dave F. Garvin[3], Andrew O. Jackson[4], Sean P. Gordon[5], John P. Vogel[5], **Samuel P. Hazen**[1]

[1]Biology Department, University of Massachusetts, Amherst, MA; [2]State Key Laboratory of Agro-Biotechnology, China Agricultural University, Beijing, China; [3]USDA-ARS Plant Science Research Unit and Department of Agronomy and Plant Genetics, University of Minnesota; [4]Plant & Microbial Biology Department, University of California, Berkeley, CA; [5]Joint Genome Institute, Walnut Creek, CA

http://www.bio.umass.edu/biology/hazen/

Two major potential areas of improvement for plant-derived biofuels are increased biomass accumulation and enhanced conversion efficiency. Large grasses such as miscanthus, sorghum, and switchgrass possess many traits which make them desirable as potential energy crops but additional improvements must be made in order to compete with fossil fuels. In order to better understand important genetic contributors to biomass recalcitrance we characterized a recombinant inbred line population of the model grass species *Brachypodium distachyon* using a *Clostridium phytofermentans* based biological conversion assay to measure plant feedstock conversion to ethanol. Additionally, we measured several quantitative trait loci (QTL) for biomass yield traits including: three QTLs for plant height, one for dry stem weight, and one for number of tillers. One QTL for conversion efficiency was mapped to chromosome two and resulted in a significant 8% increase in ethanol yield. The associated interval is 1027 kb with 139 protein coding genes. Of those genes, 36 contain polymorphisms between the two mapping parents Bd21 and Bd3-1. We refined candidate genes to those having polymorphisms that result in non-synonymous amino acid changes or occur in their promoter sequence and could affect expression. Candidate genes were prioritized based on tissue-specific expression with the best candidates highly abundant in stem; the major source of aboveground biomass in grasses. This approach has allowed us to identify a putative glucosyltransferase (CAZy family GT61) as our top candidate. We have developed and are in the process of characterizing three heterozygous inbred families to further refine the *BIOFUEL1* interval through next-generation sequencing and conversion efficiency testing. Determining the genetic basis for improved conversion efficiency in *B. distachyon* should enable translatable improvements in feedstock quality for energy crops.

Quantitative trait loci analysis of leaf carbon isotopic composition in the C_4 grass *Setaria*

Patrick Z. Ellsworth (patrick.ellsworth@wsu.edu)[1*], Patricia V. Ellsworth[1], Max Feldman[2], Darshi Banan[3], Rachel Paul[3], Ivan Baxter[2], Andrew Leakey[3], Asaph B. Cousins[1], and **Thomas P. Brutnell**[2]

[1]Washington State University, Pullman, Washington, [2]Donald Danforth Plant Science Center, St. Louis, Missouri, and [3]University of Illinois, Champaign-Urbana, Illinois

http:// foxmillet.org/

Project Goals: Genetically tractable model systems closely related to bioenergy grasses need to be developed to drive the crop improvement required for large scale, ecologically sustainable bioenergy production. *Setaria viridis* is an ideal candidate C_4 panacoid grass. The objectives of this project are to utilize genomic, computational and engineering tools to begin the genetic dissection of drought and density response in *S. viridis*. This will be achieved through: 1) Quantitative trait and association genetics; 2) novel controlled environment and field phenotyping combined with molecular and chemical profiling; 3) development of metabolic and gene networks; 4) development of transformation technologies; 5) reverse genetic testing of candidate genes.

Abstract: Leaf carbon isotopic composition ($\delta^{13}C_{leaf}$) is a proxy for water use efficiency (WUE) in high throughput phenotyping. A population of 217 recombinant inbred lines (RIL) derived from the C_4 parental lines of A-10 (*Setaria viridis*) and B-100 (*Setaria italica*) was used to identify quantitative trait loci (QTL) controlling $\delta^{13}C_{leaf}$. To test the variation in WUE in this population, the RILs were grown under drought and density field experiments conducted at the University of Illinois Champaign-Urbana in 2013 and 2014. The leaf $\delta^{13}C$ values were significantly more negative in the drought and low-density treatments in both 2013 and 2014 ($P<0.0001$). Additionally, phenotypic correlation analysis of growth and leaf traits found that $\delta^{13}C$ values were most highly correlated with leaf N content/C:N ratio in both experiments and years (mean $r = -0.42$; $P < 0.0001$). Significantly QTLs were found for $\delta^{13}C_{leaf}$ at potentially 11 locations on the genome. The two QTLs with the largest additive effects were on chromosome 5 and 8 and were present in both the drought and density experiments in both years, contributing 0.28 ± 0.04‰ and 0.19 ± 0.04‰ to $\delta^{13}C_{leaf}$ values, respectively. These two QTLs were also found for leaf C:N ratio. The QTL on chromosome 5 is a pleiotropic

locus identified for several growth-related traits in the field experiments, but the QTL on chromosome 8 may be more specific to WUE. Having identified QTLs for $\delta^{13}C_{leaf}$ that are consistent across experiments, we move closer to using marker-assisted approaches to breed for WUE in C_4 plants.

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

High Throughput Phenotyping and Quantitative Genetics for Biomass Production in a C4 Grass Model

Max Feldman[1], Ivan Baxter[1,2] (ibaxter@danforthcenter.org), Pu Huang[1],Noah Fahlgren[1], Malia A. Gehan[1], Melinda S. Wilson[1], and **Thomas P. Brutnell[1]**

[1]Donald Danforth Plant Science Center, St Louis, Missouri and [2]USDA-ARS, St Louis, Missouri

http://foxmillet.org/
http://plantcv.danforthcenter.org

Project Goals:

Genetically tractable model systems closely related to bioenergy grasses need to be developed to drive the crop improvement required for large scale, ecologically sustainable bioenergy production. *Setaria viridis* is an ideal candidate C_4 panacoid grass. The objectives of this project are to utilize genomic, computational and engineering tools to begin the genetic dissection of drought and density response in *S. viridis*. This will be achieved through: 1) Quantitative trait and Association genetics; 2) novel controlled environment and field phenotyping combined with molecular and chemical profiling; 3) development of metabolic and gene networks; 4) development of transformation technologies; 5) reverse genetic testing of candidate genes.

Abstract:

Phenotyping has become the rate-limiting step in using large-scale genomic data to understand and improve agricultural crops. Our project has used the Bellwether Phenotyping platform for controlled-environment plant growth and automated, multimodal phenotyping to study how plant biomass traits change temporally in response to water availability and identify the genetic loci underlying those changes. We have analyzed two independent genetically structured populations of *Setaria* sp.: an interspecific *S. italica* x *S. viridis* recombinant inbred line population and to two grow outs of a *S. viridis* natural diversity panel. We developed Plant Computer Vision (PlantCV) as an open-source, platform independent quantitative image analysis community resource and have used it to quantify height, biomass, water-use efficiency, color, plant architecture, and near-infrared traits. Our studies identified several major genetic loci associated with variation in plant height. However, the contribution of these loci differs depending on environment and developmental time. We are currently extending our analysis to the other measured traits.

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

Rapid optical profilometry and computer vision of leaf epidermal structure applied to genetic and environmental control of stomatal patterning in model C_4 species

Andrew D.B. Leakey[1]* (leakey@illinois.edu), Gorka Erice[1], Nathan Miller[2], Dan Xie[1], Maximilian Feldman[3], Matt Kendzior[1], Parthiban[1], Miranda J Haus[1], Elizabeth Ainsworth[4], Patrick Brown[1], Michael Mickelbart[5], Edgar Spalding[2], Ivan Baxter[3] and **Thomas P. Brutnell[2]**

[1]University of Illinois, Urbana-Champaign; [2]Donald Danforth Plant Science Center, St Louis, Missouri; and [3]USDA-ARS, St Louis, Missouri

http://foxmillet.org/

Genetically tractable model systems closely related to bioenergy grasses need to be developed to drive the crop improvement required for large scale, ecologically sustainable bioenergy production. *Setaria viridis* is an ideal candidate C_4 panacoid grass. The objectives of this project are to utilize genomic, computational and engineering tools to begin the genetic dissection of drought and density response in *S. viridis*. This will be achieved through: 1) Quantitative trait and association genetics; 2) novel controlled environment and field phenotyping combined with molecular and chemical profiling; 3) development of metabolic and gene networks; 4) development of transformation technologies; 5) reverse genetic testing of candidate genes.

Leaf epidermal structures, including stomata and hairs, play key roles in leaf function. Stomatal and hair patterning are highly regulated developmental processes in response to both environmental and genetic signals. Modern quantitative genetics approaches have not been fully applied to understanding epidermal structures due the laborious nature of phenotyping methods. C_4 grasses are agriculturally and ecologically important, in large part due to their high water use efficiency. Yet, little is known about the mechanisms controlling stomatal and hair patterning in this key plant functional type. We have developed and applied a rapid method of assessing stomatal and hair patterning in two model C_4 species – maize and setaria. The leaf surface is scanned in less than two minutes with a modified confocal microscope, generating a quantitative measurement of a patch of the leaf surface. We have developed an algorithm for automatically detecting stomata in epidermal surfaces through training of a pattern-recognition neural network. We have validated this rapid phenotyping technique in: (1) diverse *Zea mays* inbreds grown at ambient and elevated ozone using free-air concentration enrichment (FACE) technology; and (2) in a recombinant inbred line (RIL) population resulting from the cross of *Setaria italica* x *Setaria viridis*. Variation in stomatal patterning among founder lines of the NAM population of *Z. mays* was reproducible between field and greenhouse conditions at Illinois and Purdue. QTL for stomatal patterning and epidermal surface roughness were identified in Setaria.

This research was funded through Subaward No. 23009-UI, CFDA # 81.049 between University of Illinois and Donald Danforth Plant Science Center Under Prime Agreement No. DE-SC0008769 from Department of Energy.

High Fidelity Detection of QTL for Biomass Production from Rapid Imaging of a C4 Grass Crop in the Field

Darshi Banan[1]*, Rachel Paul[1], Max Feldman[2], Ivan Baxter[2,3], Andrew D.B. Leakey[1] (leakey@illinois.edu), and **Thomas P. Brutnell[2]**

[1]University of Illinois, Urbana-Champaign; [2]Donald Danforth Plant Science Center, St Louis, Missouri; and [3]USDA-ARS, St Louis, Missouri

http://foxmillet.org/

Project Goals: The development of next generation bioenergy feedstocks that meet the demands of an emerging bioenergy economy requires exponentially accelerated crop improvement relative to the gains of conventional plant breeding. Recent advances in sequencing technologies have improved evaluation of genetic variance within germplasm collections. However, applying this genomic knowledge to detect genotype-to-phenotype associations is limited by our current ability to phenotype large mapping populations cheaply and quickly, especially for complex traits such as biomass. This necessitates the development and testing of diverse technologies that can assess plant physiology and growth in non-destructive and high-throughput ways. This study tested the use of hemispherical photography as a means to detect QTL for biomass production and demonstrates the fidelity and potential of rapid imaging as a high-throughput phenotyping technique for quantitative genetics studies on field grown bioenergy feedstocks.

Current rates of crop improvement are inadequate to meet demands for high productivity agriculture that is ecologically sustainable and resilient to global environmental change. High-throughput, non-destructive methods for assessing plant growth and function are a widely anticipated solution to the "phenotyping bottleneck" that currently limits discovery of genotype-to-phenotype relationships for crop breeding and biotechnology. However, proof of concept is needed to demonstrate that a high-throughput phenotyping technique can be applied to field trials of grass crops to detect the genetic architecture of a productivity trait as effectively as destructive harvest techniques. This study demonstrates the fidelity of hemispherical photography as a method to detect quantitative trait loci (QTL) for above-ground biomass production that co-localize with results from destructive harvests in a recombinant inbred population derived from a *Setaria italica* x *Setaria viridis* cross. Plant area index (PAI) was estimated from hemispherical photographs acquired in 8 person-hours. Total above-ground, vegetative, leaf, stem, and panicle masses were estimated by a destructive harvest and weighing protocol requiring 150 person-hours. Five clusters of QTL for biomass traits were identified. All clusters included QTL for PAI and leaf mass. Fewer clusters included QTL for vegetative mass (4 QTL), total above-ground mass (3 QTL) and stem mass (2 QTL). There was not overlap between QTL clusters for PAI and those detected for traits associated with bushiness and partitioning to seeds. This demonstrates the potential for a non-destructive, high-throughput phenotyping technique that is cheap and scalable to replace destructive and slow harvest methods for discovery of genotype-to-phenotype relationships for grass crop biomass production.

This research was funded through Subaward No. 23009-UI, CFDA # 81.049 between University of Illinois and Donald Danforth Plant Science Center Under Prime Agreement No. DE-SC0008769 from Department of Energy.

Comparative metabolic network analysis of *Setaria italica* and *Setaria viridis*

Arvind K. Chavali, Peifen Zhang, Chuan Wang, Pascal Schläpfer, and **Seung Y. Rhee*** (srhee@carnegiescience.edu)

Carnegie Institution for Science, Department of Plant Biology, Stanford, CA

Project Goals: Genetically tractable model systems closely related to bioenergy grasses need to be developed to drive the crop improvement required for large scale, ecologically sustainable bioenergy production. *Setaria viridis* is an ideal candidate C_4 panacoid grass. The objectives of this project are to utilize genomic, computational, and engineering tools to begin the genetic dissection of drought and density response in *S. viridis*. This will be achieved through: 1) Quantitative trait and association genetics; 2) novel controlled environment and field phenotyping combined with molecular and chemical profiling; 3) development of metabolic and gene networks; 4) development of transformation technologies; and 5) reverse genetic testing of candidate genes.

To achieve these goals, we created genome-scale metabolic network reconstructions of *Setaria italica*, one of the oldest crops in the world, and *Setaria viridis*, a wild ancestor. Comparative network analysis will shed light on how these grasses are able to allocate metabolic resources for biomass production and C_4 photosynthesis, often limited during abiotically stressed conditions. We were motivated by the question of whether the process of domestication of *S. italica* over millennia might have had an effect on the overall metabolic network function. We analyzed both network reconstructions using Flux Balance Analysis (FBA), a computational technique that generates steady-state flux profiles of an organism's reaction network while optimizing for the production of growth-related metabolites. The use of FBA-based computational modeling allowed us to compare key aspects of primary and secondary metabolism between *S. italica* and *S. viridis*. We evaluated overall metabolic network function in both species under drought and nutrient-limiting conditions in the following specific ways: (a) we probed for differences in metabolic traits relevant to biomass such as production of carbohydrates, synthesis of amino acids, and ATP requirements in water or nutrient replete and limiting conditions; (b) we assessed network robustness by systematically removing individual genes and setting any associated reaction fluxes to zero, and subsequently evaluating for the production of biomass; (c) we searched for dependencies between reactions and identified the presence of correlated reaction sets (reaction modules that tend to function together). Future work will involve integrating gene expression datasets with our FBA models in order to generate condition-specific metabolic networks.

This project is funded by the United States Department of Energy's Biological and Environmental Research Program Grant number DE-SC0008769.

Kinetic modeling of the phenylpropanoid pathway in Arabidopsis

Longyun Guo[1]* (guo165@purdue.edu), Peng Wang[1], Rohit Jaini[2], Natalia Dudareva[1], Clint Chapple[1] and **John Morgan**[1, 2]

[1]Department of Biochemistry, Purdue University, West Lafayette, IN; [2]School of Chemical Engineering, Purdue University, West Lafayette, IN-47907

Project Goals: This project aims to rationally manipulate lignin metabolism resulting in reduced costs for biofuel production. There are 11 enzyme families involved in lignin monomer biosynthesis from phenylalanine, which makes impractical a comprehensive experimental approach to search for the optimal combinations of genetic targets for pathway manipulation. We are developing a kinetic model for lignin biosynthesis in wild type Arabidopsis, and validate with time course intracellular metabolite measurements in various mutant lines. Such a kinetic model will serve as the basis for reliable and rigorous *in silico* analysis of genetic targets to obtain desired pathway features. Meanwhile a competing pathway towards the biofuel 2-phenylethanol has also been engineered into Arabidopsis to redirect carbon flux away from lignin, and we have included this route into our kinetic model.

Lignin is the second most abundant polymer in the plant cell wall, which is essential for normal growth because of its cross-linking property and hydrophobicity. On the other hand, lignin impedes the efficient breakdown of lignocellulosic biomass for industrial applications. Precursors of lignin are synthesized from phenylpropanoid metabolism in plants, which makes the genetic engineering of this pathway a promising strategy to manipulate lignin content and composition for improved biofuel yield. Although individual enzymatic steps in phenylpropanoid metabolism have been well characterized, a systematic scheme connecting all the steps into a single model is lacking. Kinetic modeling combined with *in vivo* time-course metabolite profiling provides a mechanistic and biologically relevant way to understand the plant metabolism, from which reliable predictions can be made to guide metabolic engineering design. We selected Arabidopsis primary stem as the experimental system for modeling lignin formation. In order to obtain *in vivo* measurements for model training, excised 5-week-old stems were fed with different concentrations of $^{13}C_6$-ring labeled phenylalanine, and both the amount and isotopic enrichment of downstream intermediates were quantified with LC-MS/MS at multiple time points after feeding. Maximal activities of phenylalanine ammonia lyase (PAL) and 4-coumarate: CoA ligase (4CL) were determined along the same feeding period, and averaged lignin deposition rate was estimated from total lignin content over development. A kinetic model for general phenylpropanoid metabolism in Arabidopsis was then constructed, with the parameters identified through fitting the model's outputs with training datasets, and validated with another dataset from an independent experiment. The current model is able to capture pathway dynamics over a wide range of feeding treatments, and can be used to explore *in vivo* metabolic behaviors under different conditions.

This research is supported by the award DE-SC0008628 from the Office of Biological and Environmental Research in the US Department of Energy.

Model-Guided Metabolic Engineering of Increased 2-Phenylethanol Production in Plants

Joseph Lynch[1]* (lynch51@purdue.edu), Shaunak Ray[2], John Morgan[1,2], **Natalia Dudareva**[1], Clint Chapple[1]

[1]Department of Biochemistry, Purdue University, West Lafayette, IN; [2]School of Chemical Engineering, Purdue University, West Lafayette, IN

Project Goals: We seek to employ metabolic modeling for an improved engineering strategy for the overproduction of 2-phenylethanol, a potential biofuel. Our approach is to use a kinetic model to analyze the changes observed in transgenic Arabidopsis plants overexpressing the enzymes catalyzing the multiple steps biosynthesis of 2-phenylethanol from phenylalanine in Arabidopsis. Output from the model identify targets for further metabolic engineering strategies for optimized biofuel production.

2-Phenylethanol (2-PE) is a naturally occurring organic volatile with a characteristic rose scent. Currently, 2-phenylethanol from both natural and artificial sources is utilized for flavoring and fragrance. The physicochemical propterties of 2-PE's make it a potential biofuel, which can be used as a substitute for ethanol in petroleum-derived gasoline. However, its use for this purpose is limited by a lack of economically viable large scale production. Over the last decade the biosynthetic pathway of 2-PE formation was established in plants, and its production competes for substrate with lignin biosynthesis, which prevents efficient extraction of cellulose in ethanol production. 2-PE derives from phenylalanine, which is first deaminated and decarboxylated by a single enzyme, phenylacetaldehyde synthase,to form phenylacetaldehyde. Subsequent reduction by phenylacetaldehyde reductase forms 2-PE. While overexpression of these two enzymes, both by ourselves and others, has been successful in obtaining increased production of 2-PE, the *in planta* accumulation observed remains far lower than desired. Here we describe a metabolic model-guided approach to further enhance 2-PE production. Transgenic *Arabidopsis thaliana* were generated that overexpress the phenylacetaldehyde synthase from Arabidopsis (AAS) in tandem with the phenylacetaldehyde reductase (PAR) from tomato. Metabolite analysis revealed an apparent maximum accumulation beyond which higher gene expression was ineffective. Metabolic profiling of emitted volatiles confirmed that this was not due to release of either the 2-PE or its intermediate phenylacetaldehyde to the atmosphere. Moreover, measurements of glycosylated 2-PE indicated that limited 2-PE accumulation in transgenic plants was not due to further metabolic sequestration. Analysis of the results via our dynamic metabolic model of the phenylpropanoid network predicted that 2-PE production in these transgenic plants was substrate limited. This model prediction was further validated by combining overexpression of PAR and PAAS with overexpression a feedback-insensitive DAHP synthase, the latter of which has previously been shown to elevate intracellular phenylalanine content. This model-guided strategy successfully increased 2-PE accumulation by more than an order of magnitude compared to the overexpression of the 2-PE pathway alone. We continue to utilize the kinetic model in combination with metabolic profiling by LC/MS/MS to better refine our strategy and rationalize additional improvements.

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

Next-Generation Protein Interactomes for Plant Systems Biology and Biomass Feedstocks Research

Shelly A. Trigg[1,2,3*] (swanamaker@salk.edu), Renee Garza[1], Andrew MacWilliams[1], Joseph Nery[1], Joaquin Reina[1,4], Anna Bartlett[1], Rosa Castanon[1], Adeline Goubil[1], Joseph Feeney[1], Ronan O'Malley[1,2], Mary Galli[1,2] and **Joseph R. Ecker**[1,2,5]

[1]Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla CA; [2]Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla CA; [3]Biological Sciences Department, University of California, San Diego; Bioinformatics Program, University of California, San Diego; and [5]Howard Hughes Medical Institute

Project Goals: Elucidate interactome characteristics giving rise to desirable phenotypic traits in biofuel feedstock crops by rapidly identifying all their protein-protein interactions using the *en masse* next-gen yeast two-hybrid screening system, ProCREate.

Biofuel crop cultivation is a necessary step in heading towards a sustainable future, making their genomic studies a priority. While technology platforms that currently exist for studying non-model crop species, like switch-grass or sorghum, have yielded large quantities of genomic and expression data, still a large gap exists between molecular mechanism and phenotype. The aspect of molecular activity at the level of protein-protein interactions has recently begun to bridge this gap, providing a more global perspective. Interactome analysis has defined more specific functional roles of proteins based on their interaction partners, neighborhoods, and other network features, making it possible to distinguish unique modules of immune response to different plant pathogens[1]. As we work towards cultivating heartier biofuel crops, interactome data will lead to uncovering crop-specific defense and development networks. However, the collection of protein interaction data has been limited to expensive, time-consuming, hard-to-scale assays that mostly require cloned ORF collections. For these reasons, we have successfully developed a highly scalable, economical, sensitive, and potentially quantitative yeast two-hybrid assay, ProCREate, that can be universally applied to generate proteome-wide primary interactome data. ProCREate enables *en masse* pooling and massively paralleled sequencing for the identification of interacting proteins by exploiting Cre-lox recombination. As a proof of principle, we thoroughly screened a well-validated collection of ~2000 *Arabidopsis* transcription factors in 10 replicate experiments, generating a comprehensive *Arabidopsis thaliana* transcription factor interaction network (AtTFIN1). After observing a high recall rate, a high overlap with literature, and rational novel interactions, we then tested ProCREate's scalability by generating and screening *Arabidopsis* cDNA libraries. Once these recently collected datasets are further validated, ProCREate will be used to screen ORF/cDNA libraries from feedstock plant tissues. The interactome data generated will yield deeper insight into many molecular processes and pathways that can be used to guide improvement of feedstock productivity and sustainability.

References

1. Jiang, Z., Dong, X., and Zhang, Z. *Nature Scientific Reports* **6**, 19149 (2016)

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.

Moving Recombineering Beyond *Escherichia Coli*

Christopher Gregg,[1] Xavier Rios,[1] Marc Lajoie,[2] Timothy Wannier,[1]* (timothy_wannier@hms.harvard.edu), Gabriel Filsinger,[1] and **George M. Church**[1]

[1]Harvard Medical School, Department of Genetics, Boston, MA; and [2]University of Washington, Seattle, WA

http://arep.med.harvard.edu (Church lab)

Project Goals: We aim to develop the ability to make multiplexed, high-throughput gene edits in a wide array of microbial species. By improving our mechanistic knowledge of recombineering and developing the ability to screen large libraries of protein variants, we endeavor to greatly expand the reach of Multiplex Automated Genome Engineering (MAGE). Finally, we hope to demonstrate the utility of this platform technology in applications ranging from improving biomolecule yield in Streptomyces to studying photosynthesis in cyanobacteria.

It has been demonstrated in our laboratory that Multiplex Automated Genome Engineering (MAGE) is an easy and efficient mechanism for allowing multiplexed genomic mutations in Escherichia coli.[1] The MAGE process has enabled rapid improvement of metabolic synthesis pathways, genome-wide recoding, and bio-containment applications.[2-4] The protein at the core of MAGE is the λ Red recombinase, λ β (Bet), a phage recombinase, that when expressed improves recombination efficiency of single-stranded DNA oligonucleotides into the bacterial genome ~1E5-fold.[5] Here, we present a platform technology, Serial Evolutionary Enrichment for Recombinases (SEER), that enables the rapid discovery of Bet variants that work well in intractable prokaryotic strains. We queried metagenomic space for Bet homologues, and synthesized a library of putative recombinases with representatives from all families of prokaryotic single stranded annealing proteins. This library (Bet included) was then subjected to successive rounds of enrichment in E. coli for functional recombinases. Surprisingly, we found library members that recombined at or above the frequencies seen for Bet in Escherichia coli, including homologues from Lactobacillus reuteri and Corynebacterium glutamicum.

We further investigated the molecular mechanism of Bet-mediated recombination, and demonstrate an interaction with single-stranded binding protein (SSB). Based on in vitro annealing assays, Bet acts to specifically unload SSB from SSB-coated single-stranded DNA (ssDNA). Removing SSB from ssDNA (like the lagging strand of replication) then enables strand-strand annealing, which is the mechanism by which mutagenic ssDNA is incorporated into the replication fork in Bet-mediated recombineering.[5] Furthermore, we show that an SSB taken from the same host organism as the Bet recombinase homolog can improve the functioning of said Bet homolog in other organisms. The SEER platform will be improved by this knowledge, as Bet fusion proteins will be introduced to take advantage of high recombineering efficiency and organism-specific SSB interaction. High-throughput evolutionary screening will

allow us to test thousands of protein variants in parallel to quickly and robustly identify high-efficiency recombinase mutants for any prokaryotic microbe with the requisite biological toolkit, which includes competency, known plasmids, and inducible promoters.

References

1. Wang, H.H. et al. 2009. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*. 460, 7257 (Aug. 2009), 894–8.

2. Wang, H.H. et al. 2012. Genome-scale promoter engineering by co-selection MAGE. *Nature methods*. 9, 6 (Jun. 2012), 591–3.

3. Lajoie, M.J. et al. 2013. Probing the limits of genetic recoding in essential genes. *Science (New York, N.Y.)*. 342, 6156 (Oct. 2013), 361–3.

4. Lajoie, M.J. et al. 2013. Genomically recoded organisms expand biological functions. *Science (New York, N.Y.)*. 342, 6156 (Oct. 2013), 357–60.

5. Lajoie, M.J. et al. 2012. Manipulating replisome dynamics to enhance lambda Red-mediated multiplex genome engineering. *Nucleic acids research*. 40, 22 (Dec. 2012), e170.

This project was funded by DOE grant DE-FG02-02ER63445.

Engineering bacterial translational machinery for incorporation of D-amino acids

Po-Yi Huang[1] (pyhuang@hms.harvard.edu), Narasimhan Kamesh[1], and **George M. Church**[1]

[1]Harvard Medical School, Genetics Department, Massachusetts, Boston

http://arep.med.harvard.edu (Church lab)

Project Goals: Life is an anti-entropic phenomenon with two mutually reinforcing characteristics—homochirality and stereospecific catalysis—that have resulted in the exclusive presence of L-amino acids in proteins in the living world. Molecules from the mirror kingdom, such as proteins with D-amino acids (D-AAs), will be more resistant to existing biodegradation systems, and could provide new tools for biotechnology, classes of pharmaceuticals, and pathways for producing recalcitrant carbon. By gradually rewiring existing bio-machineries we aim to build a bridge to the space of mirror-imaged biomolecules. Here we investigated translational mechanisms relevant to protein synthesis with D-AAs, including tRNA amino acylation, EF-Tu binding of amino acyl-tRNAs, and ribosome catalysis of peptidyl transfer. Our data demonstrate that affinity between EF-Tu and amino acyl-tRNAs is critical to D-AA incorporation, and suggest that ribosome stalling on D-AAs is due failure of a substrate-assisted mechanism.

D-amino acid (D-AA) containing peptides (DAACP) are widely present in bioactivities within and secretions from microbes, fungi and amphibians[1]. In nature, these molecules are made through non-ribosomal pathways, such as non-ribosomal peptide synthesis or post-translational modifications such as epimerization. D-AA-containing proteins and peptides, mimicking the concept of DAACP, have been shown to have prolonged half lives in serum and resistance toward proteases without immunogenicity, desirable properties for pharmaceuticals. However, current protein expression methods cannot be applied to explore this immense space of potential compounds due to evolved barriers in translational machinery against incorporating D-AAs into proteins.

We used a purified *E. coli* protein synthesis system[2] as a model system in which to study how to best engineer translation machinery to tolerate D-AAs without interference of D-AA oxidase and D-aminoacyl-tRNA deacylase. While exceptions exist, it is believed that mechanisms have evolved to block the incorporation of D-amino acid at almost every step of protein synthesis.[3] There are three major steps at which D-AAs are discriminated from L-AAs in core translation machinery: aminoacylation of tRNAs by aminoacyl-tRNA synthetases (aaRSes), formation of ternary complexes with EF-Tu-GTP and their delivery to the ribosome, and the ribosome's own catalysis of peptide bond formation.

We began by surveying the amino acylation specificity of all 20 aaRSes toward D-AA, and then compared this against both chemical and ribozyme-catalyzed acylation methods. Next, we tested the overall effect of EF-Tu's binding to D- or L-AA-tRNAs, based on different tRNA backbones

and anticodons.[4] Our data suggested that tRNA backbones with high affinity to EF-Tu could improve the delivery of D-AA-tRNA to ribosome and hence D-AA incorporation.

It has long been argued if *E. coli* ribosome can catalyze peptide bond formation with D-AAs. Recent studies indicate that D-AA containing peptides can be produced, but with greatly reduced kinetics.[5–8] Our data agree with reports of single incorporations of D-AA in elongating peptide chains but we have not observed measurable consecutive D-AA incorporations. However, in our initial ribosome engineering experiments, we have found a few mutants in which D-AAs can be incorporated with increased efficiency relative to L-AAs, although with lower overall peptide synthesis activity. This leads us to hypothesize that L-AA incorporation involves a substrate-assisted mechanism that is not available with D-AA[9].

References

1. Ollivaux, C., Soyez, D. & Toullec, J.-Y. Biogenesis of d-amino acid containing peptides/proteins: where, when and how? *J. Pept. Sci.* **20,** 595–612 (2014).
2. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19,** 751–5 (2001).
3. Ahmad, S. *et al.* Mechanism of chiral proofreading during translation of the genetic code. *Elife* **2,** e01519 (2013).
4. Dale, T., Sanderson, L. E. & Uhlenbeck, O. C. The Affinity of Elongation Factor Tu for an Aminoacyl-tRNA Is Modulated by the Esterified Amino Acid†. *Biochemistry* **43,** 6159–6166 (2004).
5. Dedkova, L. M., Fahmi, N. E., Golovine, S. Y. & Hecht, S. M. Enhanced d-Amino Acid Incorporation into Protein by Modified Ribosomes. *J. Am. Chem. Soc.* **125,** 6616–6617 (2003).
6. Fujino, T., Goto, Y., Suga, H. & Murakami, H. Reevaluation of the d-Amino Acid Compatibility with the Elongation Event in Translation. *J. Am. Chem. Soc.* **135,** 1830–1837 (2013).
7. Englander, M. T. *et al.* The ribosome can discriminate the chirality of amino acids within its peptidyl-transferase center. *Proc. Natl. Acad. Sci.* **112 ,** 6038–6043 (2015).
8. Achenbach, J. *et al.* Outwitting EF-Tu and the ribosome: translation with d-amino acids. *Nucleic Acids Res.* **43 ,** 5687–5698 (2015).
9. Wallin, G. & Åqvist, J. The transition state for peptide bond formation reveals the ribosome as a water trap. *Proc. Natl. Acad. Sci.* **107,** 1888–1893 (2010).

This project has been funded by DOE grant DE-FG02-02ER63445, the Harvard Origin of Life Initiative, and EMD Millipore Inc.

A General Strategy for Sensing Small Molecules in Eukaryotes

Justin Feng[1,2,9]*(justinfeng@fas.harvard.edu), Benjamin W. Jester[3,4,9], Christine E. Tinberg[5,9], Daniel J. Mandell[2,6,9 5], Mauricio S. Antunes[7], Raj Chari[2], Kevin J. Morey[7], Xavier Rios[2], June I. Medford[7 6] , **George M. Church[2,6]**, Stanley Fields[3,4,8], and David Baker[4,5]

[1]Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA; [2]Department of Genetics, Harvard Medical School, Boston, MA; [3]Department of Genome Sciences, University of Washington, Seattle, WA; [4]Howard Hughes Medical Institute, University of Washington, Seattle, WA; [5]Department of Biochemistry, University of Washington, Seattle, WA; [6]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA; [7]Department of Biology, Colorado State University, Fort Collins, CO; [8]Department of Medicine, University of Washington, Seattle, WA; [9]authors contributed equally

http://arep.med.harvard.edu (Church lab)

Project Goals: Microbes engineered to express synthesis pathways for useful molecules usually require extensive pathway and metabolic optimization to become efficient producers. Selections from large libraries of engineered or random variants can be a highly effective optimization method, but this requires the microbe to be able to sense target molecules and transduce information about their presence into growth or death signals[1]. However, such biosensors do not exist for many microbially producible useful molecules because they have been absent from the microbes' evolutionary histories. To address this issue, our lab previously developed methods for engineering the specificity of natural allosteric transcription factors like LacI to efficiently and specifically respond to novel ligands[2]. However, because the structural requirements for preserving allostery constrains the flexibility of ligand retargeting, we sought more modular methods for generating biosensors that operated by other mechanisms.

We describe a method to create biosensors starting with a computationally designed ligand-binding domain (LBD). The LBD is fused to a reporter and is destabilized by mutation such that the fusion accumulates only in cells containing the target ligand. We illustrate the power of this method by developing biosensors for digoxin and progesterone. Addition of ligand to cells expressing a biosensor activates transcription in yeast, mammalian cells and plants, with a dynamic range of up to ~100-fold. We use the biosensors to improve the biotransformation of pregnenolone to progesterone and to regulate CRISPR activity. In concert with computational LBD design approaches, this method should enable the generation of biosensors for a broad range of molecules.

References

1. Raman, S., Rogers, J. K., Taylor, N. D. & Church, G. M. Evolution-guided optimization of biosynthetic pathways. *Proc. Natl. Acad. Sci.* 201409523 (2014).

doi:10.1073/pnas.1409523111

2. Taylor, N. D. *et al.* Engineering an allosteric transcription factor to respond to new ligands. *Nat. Methods* **In press,** 1–11 (2016).

Funding for this project was through the Defense Threat Reduction Agency to D.B. and J.I.M; NIH grant P41 GM103533 to D.B. and S.F; DOE grant DE-FG02-02ER63445, NSF SynBERC awards MCB-134189 and EEC-0540879 to G.M.C.; NSF DGE1144152 to J.F.; HHMI to D.J.M, D.B. and S.F.

Design, Synthesis and Testing of a 57-Codon Genome

Matthieu Landon[1,2,3]*†(matthieulandon@fas.harvard.edu), Nili Ostrov[1]†, Marc Guell[1,9]†, Gleb Kuznetsov[1,5]†, Jun Teramoto[1,4], Julie E. Norville[1], Natalie Cervantes[1], Minerva Zhou[7], Michael G. Napolitano[6] , Mark Moosburner[1], Ben Pruitt[9], Nick Conway[9], Dan Goodman[1], Kerry Singh[7], Cameron L. Gardner[1], Barry L. Wanner[1,4,8], Marc Lajoie[1], **George M. Church[1]**

[1]Department of Genetics, Harvard Medical School, Boston, MA; [2]Program in Systems Biology, Harvard University, Cambridge, MA; [3]Ecole des Mines de Paris, Mines Paristech, 75272 Paris, France; [4]Department of Biological Sciences, Purdue University, West Lafayette, IN; [5]Program in Biophysics, Harvard University, Cambridge, MA; [6]Program in Biological and Biomedical Sciences, Harvard University, Cambridge, MA; [7]Department of Biological Engineering, MIT, Cambridge, MA; [8]Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA; [9]Wyss Institute for Biologically Inspired Engineering, Boston, MA; †Equal contribution

http://arep.med.harvard.edu (Church lab)

Project Goals: Recoding, replacing codons genome-wide in an organism, is a unique and transformative way to explore currently unknown biological landscapes. By changing all instances of a target codon to synonymous codons and disabling the target codon function, invading DNA from natural sources cannot be translated, making the GRO resistant to infection and unreceptive to horizontal transfer. Unused codons can also be assigned to nonstandard amino acids (NSAAs) to generate proteins with novel functions[2], or to make the GRO biocontainable by making provision of NSAAs essential to growth[1]. We previously created a GRO in which all instances of the very rare UAG codon in *E. coli* were changed to UAAs (321 changes)[3]. To create GROs with larger scales of change, we have been developing radical recoding methods by which the genome is resynthesized with all target codons reassigned to other values. Here, our goal has been to develop these methods and generate an *E. coli* that lacks 7 codons. This work underscores the feasibility of extensively rewriting genomes, and establishes a framework for large-scale design, assembly, troubleshooting and phenotypic analysis of synthetic organisms.

We report the computational design, synthesis and partial testing of segments of a recoded *Escherichia coli* genome in which all 62,214 instances of seven codons were replaced with synonymous alternatives across all protein coding genes. Replacement values for target codons AGA (Arg), AGG (Arg), AGC (Ser), AGU (Ser), UUG (Leu), UUA (Leu) and UAG (Stop) were chosen by in-house designed software that attempts to retain key mRNA features such as secondary structure and ribosome binding sites, refactors overlapping genes, and avoids sequences that are difficult to synthesize. The entire computationally redesigned 3.98Mb genome was divided into 1254 fragments and synthesized *de novo* by a variety of vendors. As even a single lethal design error can lead to an inviable strain, we developed a pipeline for assembling the fragments into ~50 kb segments, testing these segments as plasmids in separate *E. coli* strains, deleting the corresponding non-recoded chromosomal sequence to demonstrate that the

recoded sequence complements the original, and integrating the recoded sequence directly into the chromosome. To debug failures at the various steps of this process, we also developed a troubleshooting pipeline to diagnose the causative gene, codon, or other feature, and to determine corrections. At this time, we have assembled the entire recoded genome in 87 segments and validated 54% of genes, including 50% of all essential genes, by individually testing 47 segments up to the chromosomal integration step. We found that 99.3% of 1918 recoded genes retained functionality with only a modest effect on strain fitness, and investigated 13 design flaws using our troubleshooting pipeline. As we push towards generating viable, chromosomally integrated strains for all 87 recoded segments, we are simultaneously developing and optimizing segment assembly strategies for creating a strain with a fully recoded genome. Once complete, a 57-codon organism could broaden synthetic functionality in living cells.

References

1. D. J. Mandell *et al.*, Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518**, 55 (2015).
2. C. C. Liu, P. G. Schultz, Adding new chemistries to the genetic code. *Annu Rev Biochem* **79**, 413 (2010).
3. AM. J. Lajoie *et al.*, Genomically recoded organisms expand biological functions. *Science* **342**, 357 (2013).

This project has been funded by DOE grant DE-FG02-02ER63445 and DARPA grant BAA-12-64.

Risk and escape policies, procedures, and practices: Issues and implications for biosystems design (synthetic biology) R&D

Amy K. Wolfe[1]* (wolfeak@ornl.gov), Rachael A. Bergmann[2], Joel P. Hewett[2], Savannah C. Stelling[3]

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee, [2]Oak Ridge National Laboratory, Oak Ridge, Tennessee via Oak Ridge Associated Universities, [3]Columbia University Graduate School of Public Health, New York, NY (formerly Oak Ridge Associate Universities)

Project Goals: This project investigates risk and containment issues associated with biosystems design (synthetic biology) research and development (R&D) from social and institutional perspectives. It aims to identify circumstances that affect human health and environmental risks stemming from biosystems design R&D, thereby identifying opportunities for avoiding, minimizing, or mitigating those risks. By focusing on research practices in a variety of research settings and associated with different target organisms, this project seeks to identify possible blind spots that inadvertently could create or increase human health or environmental risks.

To achieve our project goals, we investigate risk and containment issues associated with biosystems design R&D from three perspectives:

- Public sources—secondary data collected and analyzed from formal (e.g., journal articles) and informal (e.g., news articles, reports) publications related to biosystems design;
- Scientists conducting biosystems design R&D—primary data we gather through interviews with people engaged in energy- or environment-related biosystems design R&D; and
- Biosafety professionals— primary data we gather through interviews with people who play key biosafety roles in biosystems design R&D (especially institutional biosafety committee members).

Our analyses center on risk- and containment-related research practices and on issues associated with risks to human health and the environment. We divide “research practices” into two broad categories—practices used in the day-to-day conduct of research and practices associated with the organism or system being designed. Our analyses emphasize key elements that shape the social and institutional context within which biosystems design R&D takes place, such as research setting, research goal, organism studied, formal and informal rules, and disciplinary training. We study “research practice” because that is where the set of elements (listed above and others) that influence research context translate into behavior.

To date, we have conducted interviews with dozens of scientists and 10 biosafety professionals. These interviews reveal notable variability in risk- and containment-related research practices

and institutional approaches, even in seemingly similar circumstances. As examples, scientists report differences in containment and disposal practices when conducting research on the same category of target organism, and only some institutional biosafety committees members report using a 'plus' approach in assigning Biosafety Levels to labs (e.g., BSL 1+). Interviews also reveal some remarkably consistent responses. For example, virtually all of the scientists interviewed report that they use routine research practices in their laboratory or greenhouse settings, and that their day-to-day research practices are not biosystems design- or synthetic biology-specific. We are continuing to gather interview data, to analyze those data, and to consider the implications of those analyses for risk- and containment-related research practices.

References

Wolfe, A. K., M. F. Campa, R. A. Bergmann, S. C. Stelling, D. J. Bjornstad, B. L. Shumpert (accepted). Synthetic biology R&D risks: Social-institutional contexts matter!" *Trends in Biotechnology.*

Wolfe, A. K. (2015) "Societal aspects of synthetic biology: Organisms and applications matter!" *Journal of Responsible Innovation,* 2:1, 121–123.

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

Experimental and Computational Tools for Sequence to Activity Mapping

Gur Pines* (gur.pines@colorado.edu), Andrew Garst, Marcelo Bassalo, Andrea L. Halweg-Edwards and **Ryan T Gill**

University of Colorado, Boulder

http://www.dynamcc.com

Project Goals: A comprehensive sequence to activity mapping is central to a successful protein, pathway and genome level engineering. Even at the single protein level, the mutational space is too vast for complete exploration. We develop a codon compression approach that helps in reducing library size when performing saturation mutagenesis using degenerate oligonucleotides. We also develop a method to systematically edit genomic loci in the single codon resolution. This method, termed CRISPR EnAbled Trackable genome Engineering (CREATE) combines traditional lambda red recombineering with the high efficiency editing of the CRISPR-CAS9 system. Using CREATE, we successfully saturated every site in the essential E. coli *folA* gene and mapped both the essential sites for activity and specific mutations conferring resistance to its inhibitor, trimethoprim.

The common approach for saturation mutagenesis is using NNK (N = A/C/G/T, K = G/T) which covers all amino acids in a single codon. However, since NNK codes for 32 codons, some of the 20 amino acids are represented more than once, leading to amino acid bias. Moreover, this approach will always include the wild type (wt) amino acid and a stop codon. These factors needlessly increase library size and downstream screening efforts. Also, depending on the target organism, some of the NNK codons might suffer from extremely low usage hindering protein expression. Our computational approach allows to define the exact amino acid collection that is desired, along with the target organism. The output is a pool of degenerate codons that covers the exact query without any off target amino acids, and with high usage codons, as defined by the user[1]. For example, when completely saturating a site, the compressed codons will cover 19 amino acids, excluding the wt and stop codons. Another option is to define an exact set of amino acids to replace the wt. This can be knowledge-based or in order to keep the wt residue properties such as hydrophobicity or size. Another feature allows to define the level of redundancy within the codon pool. We added this function since it is known that in some cases synonymous mutations may have a significant effect on mRNA stability, protein folding rates and more[2,3]. We now have the codon compression tools in a dedicated website to allow easy access. The website also allows to upload custom genetic and usage tables enabling users that are interested in non standard or synthetic genetic codes to use these tools. We hope that this

improved access will increase the audience to our approach and will help in speeding up protein engineering.

CREATE is a method for rapid and efficient genome editing in the single codon resolution over tens to hundreds of thousands of loci in parallel. This method combines phage-based recombineering[4] with CRISPR genome editing[5]: site-specific gRNAs and editing templates are synthesized in parallel on a microarray and later pool-cloned into a plasmid. The gRNA, together with an inducible Cas9 digest the genome in a sequence-specific manner, while the editing template is integrated genomically by the recombineering machinery. The editing template harbors two mutations: the first is the desired modification and the second silently mutates a PAM sequence, which is necessary for CAS9 cleavage. This approach allows to build genome-wide libraries and complete saturation of whole genes. Since the plasmid editing templates highly correlates with the genomic edits, simple plasmid sequencing replaces whole genome interrogation, making this system easily trackable. We demonstrate CREATE by completely saturating the essential E. coli *folA* gene, generating its full sequence to activity map. Moreover, we also challenged these library cells with trimethoprim, a FolA specific inhibitor, and found two mutations within the same site that confer resistance to this drug.

References

1. Pines, G. *et al.* Codon compression algorithms for saturation mutagenesis. *ACS Synth. Biol.* **4,** 604–614 (2015).
2. Plotkin, J. B. & Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* **12,** 32–42 (2011).
3. Firnberg, E., Labonte, J. W., Gray, J. J. & Ostermeier, M. A Comprehensive, High-Resolution Map of a Gene's Fitness Landscape. *Mol. Biol. Evol.* **31,** 1581–1592 (2014).
4. Pines, G., Freed, E. F., Winkler, J. D. & Gill, R. T. Bacterial Recombineering: Genome Engineering via Phage-Based Homologous Recombination. *ACS Synth. Biol.* **4,** 1176–1185 (2015).
5. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31,** 233–239 (2013).

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

Biochemical Production from Lignocellulose by CRISPR EnAbled Trackable genome Engineering (CREATE)

Rongming Liu* （Rongming.liu@colorado.edu）, Andrew Garst, Liya Liang,

Andrea L. Halweg-Edwards，**Ryan T. Gill (PI)**

Renewable and Sustainable Energy Institute, University of Colorado Boulder, Boulder, CO 80309

Project Goals: We are developing tools for rationally-based protein engineering which will allow multiple modifications from the single protein to the whole pathway levels. CRISPR EnAbled Trackable genome Engineering (CREATE) is a new genome-scale engineering strategy that couples the high efficiency of CRISPR editing with the massively multiplexed rational design offered by parallel oligomer synthesis. Here we used CREATE to construct a 3HP producing strain which can utilize hemicellulose-based hydrolysate as carbon source.

Metabolic engineering has expanded from a focus on designs requiring a small number of genetic modifications to increasingly complex designs driven by advances in genome-scale engineering technologies. CRISPR EnAbled Trackable genome Engineering (CREATE) is a new genome-scale engineering strategy that couples the high efficiency of CRISPR editing with the massively multiplexed rational design offered by parallel oligomer synthesis (1). Here we used CREATE to construct a 3HP producing strain which can utilize hemicellulose-based hydrolysate as carbon source.

After pretreatment of corn stover or corn stalk, glucose and xylose are the main carbon sources in these lignocellulose hydroysate. However, *Escherichia coli* can not co-utilize glucose and xylose at the same time. To solve the problem, we used CRIPSR editing method to delete the *ptsHI* genes. As a result, the strain *E. coli* BG_{3HP}, Δ*ptsHI* can utilize glucose and xylose simultaneously, but the glucose consumption rate was also decreased by the deletion of *ptsHI.* To enhance the glucose consumption rate, we used RBS calculator to design a RBS library for *glk* gene that is important gene in another glucose transport system (2). Furthermore, we used CREATE strategy to construct this library and did the selection. As a result, one of the library strains, BG_{3HP}, Δ*ptsHI*, glkup-7, can grow faster than others.

During the pretreatment of corn stover or corn stalk, side-reaction products (furfural, 5-hydroxymethylfurfural, formate, acetate, and soluble lignin products) are formed. Furfural (dehydration product of pentose sugars) is widely regarded as one of the most important inhibitors. To enhance the tolerance of furfural, we used CREATE strategy to construct the library targeting on 7 transcription and translation genes and 10 high level regulators, and did the selection for furfural tolerance. After sequencing of positive mutant, several mutant of rpoD, crp, and nusA can grow faster than the control strain with 2 g/L furfural. We used one of the positive mutant of rpoD strain as host strain, and introduced 3HP pathway into this positive mutant. As a result, the strain BG_{3HP}, Δ*ptsHI*, glkup-7, rpoD-2, psmart-J23119-Camcr-ydfG, can generate 5 g/L 3HP in the flask fermentation with sugar mixture media containing 2 g/L furfural, while the control strain can only synthesis 0.5 g/L 3HP under same conditions.

Reference

1 Andrew Garst et al., Nature Biotechnology (2016), under review.
2 Iman Farasat et al., Molecular Systems Biology (2014) 10: 731.

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

Title:

CRISPR EnAbled Trackable Genome Engineering: A Technology for High Throughput Genetic Prototyping of Large DNA constructs.

Authors:

Andrew Garst[1*], Marcelo Bassalo[1,2], Gur Pines[1], Sean A. Lynch[1], Andrea Halweg-Edwards[1], Rongming Liu[1], Zhiwen W. Wang[1,3,4,5], **Ryan T. Gill**[1]

Affiliations:

[1] Department of Chemical and Biological Engineering, University of Colorado Boulder, Boulder, Colorado 80303, USA.

[2] Department of Molecular, Cellular and Developmental Biology, University of Colorado Boulder, Boulder, Colorado 80303, USA.

[3] Department of Biochemical Engineering, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, People's Republic of China.

[4] SynBio Research Platform, Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), Tianjin University, Tianjin 300072, People's Republic of China.

[5] Key Laboratory of Systems Bioengineering (Ministry of Education), Tianjin University, Tianjin 300072, People's Republic of China.

* Presenting Author

Project Goals:

The goal of this project was to develop a generalizable genome editing technology with the following capabilities:

1. editing anywhere in the genome
2. at efficiencies >90% at many loci
3. using a broad range of mutations (ranging from single nucleotide changes to indels of varying size)
4. allow tracking of each mutation in the population to enable relative fitness comparisons
5. perform such editing at tens of thousands of edits in parallel.

As a result of this work we have developed a simple and unique solution that can simultaneously address all of these engineering objectives.

Abstract text.

The declining cost and increasing throughput of DNA sequencing has revolutionized modern biological workflows and enabled increasingly sophisticated approaches to understanding the effects of genotypic variation on molecular processes involved in metabolic function, regulatory control and disease. Although similar improvements in DNA synthesis have occurred in parallel, technologies that can effectively implement synthetic DNA to systematically manipulate and study living systems have lagged behind. Strategies that span the design-build-test forward-engineering cycle therefore offer the promise of closing this gap by enabling simplified and systematic end-to-end read/write workflows. Here we describe CRISPR

EnAbled Trackable genome Engineering (CREATE), a strategy that couples the high efficiency of CRISPR editing with massively parallel oligomer synthesis to enable trackable precision editing on a genome wide scale. In this work we employed CREATE it to a variety of application areas that are of general importance to the biological engineering community. For example we applied CREATE to survey > 4000 mutations in the AcrB one of the central efflux pumps in *E. coli* that is implicated in tolerance to a broad range of solvents and a key component of multiple antibiotic resistance phenotypes. We identified multiple mutations to a loop-helix motif adjacent the central funnel that appear to improve efflux activity against multiple substrates, suggesting that this region may serve as an ideal engineering target. Furthermore we show that CREATE can scale to the whole genome level and serve as a tool to reconstruct mutations identified from long term adaptive laboratory evolution (ALE) studies or perform genome scale searches with up to single nucleotide resolution that improve industrially relevant antibiotic resistance and solvent tolerance phenotypes.

Funding statement.
This work was funded under DOE grant DESC008812 awarded to Ryan Gill at the University of Colorado.

Bioverse: a comprehensive platform for designing biological materials

Andrea L. Halweg-Edwards*[1], Andrew Garst[1], Alaksh Choudhury[1], Cierra Walker[2], Adam Robbins-Pianka[1], Akash Gaonkar[3], Gur Pines[1], Ryan Hockstad[1], Katia Tarasava[4], Deanne Sammond[5], Adam Arkin[6], **Ryan T. Gill (PI)[1]**

*Presenter: andrea.edwards@colorado.edu

[1]Renewable and Sustainable Energy Institute, University of Colorado Boulder, Boulder, CO 80309
[3]Department of Computer Science, University of Colorado Boulder, Boulder, CO 80309
[4]Materials Science and Engineering Program, University of Colorado Boulder, Boulder, CO 80309
[5]Biosciences Center, National Renewable Energy Laboratory, Golden, Colorado, 80401
[6]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Project Goals: We aim to build a user-friendly platform for designing biological materials and exploring molecular adaptation. Our goal is to consolidate biological databases with structural, functional, and phylogenetics information to enhance the search speed and retrieval of relevant design features for a wide range of query types. As a result, the bioverse platform will allow advanced, google-esque query capabilities allowing engineers to design experiments that probe important processes including biophysical properties that confer disease phenotypes as well as design products ranging from secondary metabolites to biological packaging materials.

Abstract

The discovery of CRISPR-enabled genomic and episomal editing technology has sparked a surge in development of synthetic biology tools for editing DNA. Coupled with array-based DNA synthesis technology, biologists now have the opportunity to edit targeted sequences at unprecedented scales and throughputs. In response to this recent progress, our BioDesign Group is developing *bioverse*, a freely-available and open-sourced computational design platform. This web-based design tool will provide users with the capability to generate tens of thousands of rationally designed editing cassettes based on our patented CRISPR EnAbled Trackable genome Engineering (CREATE) technology, which is compatible with array-based DNA synthesis platforms. Bioverse will incorporate phylogenetic informatics as well as structure-guided design principles powered by the Rosetta software suite for computational modeling and analysis of biological macromolecules. Furthermore, our design platform will be built upon active machine learning approaches that will allow users to input their test data back into the pipeline to improve mutant-phenotype predictive power and inform the next round of library synthesis. Additionally, we are working to integrate bioverse with the Department of Energy Systems Biology Knowlegebase (Kbase) to allow crowd-sourced feedback on macromolecular design principles and best practices as well as providing the community with a wide range of metabolic engineering tools.

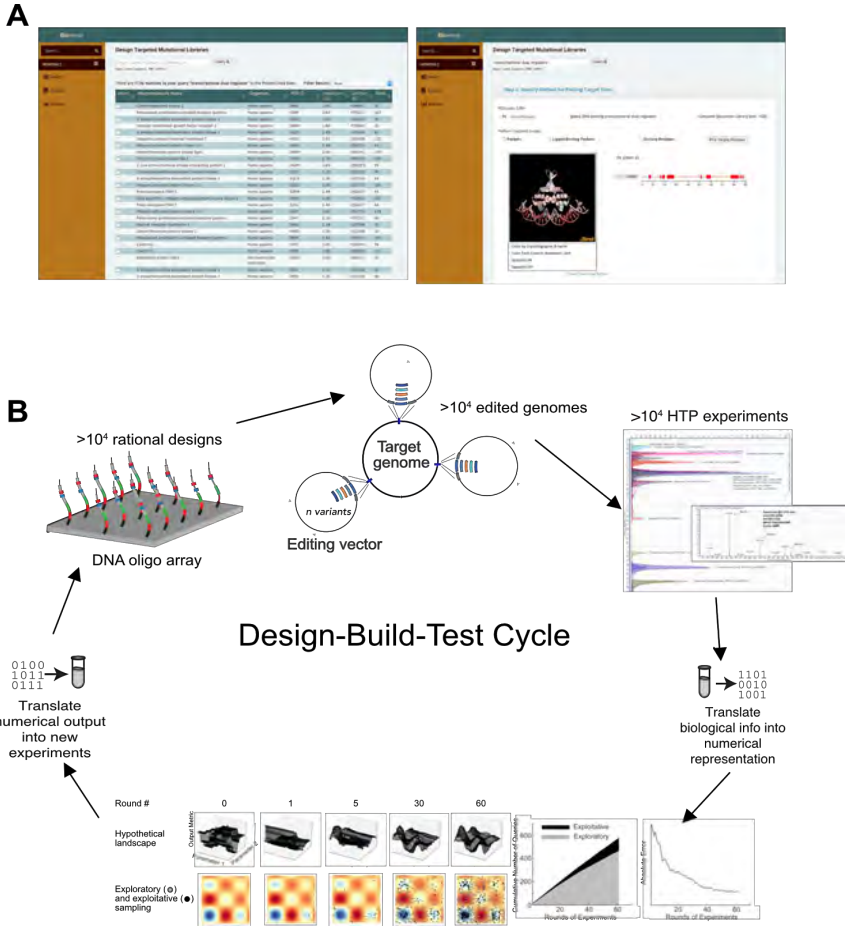


Figure 1. Synthesis Aided Design. Affordable synthetic DNA oligomers combined with high throughput genome editing and testing technologies allow engineers to rapidly generate informative datasets regarding genotype-phenotype relationships. **A)** The bioverse platform will allow researchers to rapidly design biological materials for testing a wide range of phenotypes including small molecule production and molecular adaptation. **B)** Bioverse aims to design libraries that include a sufficient range of hypotheses and controls, allowing application of active machine learning to improve the predictive power of genotype-phenotype models during iterative rounds of Design-Build-Test-Learn. In this example, mass spectrometry based methods are used to collect data on a library of >10,000 genetic variants, and classification and regression trees are used to inform the next round of experiments. Exploration and exploitation of experimental datasets allows efficient optimization of the mathematical model describing the relationship between genotype and phenotype.

Funding Statement

Grant title: A Platform for Genome-scale Design, Redesign, and Optimization of Bacterial Systems; Project grant number (DE-SC008812) and FWP number (ERWER44).

A Strategy for Genome-scale Design, Redesign, and Optimization for Ethylene Production in *E. coli*

Sean Lynch[1,2,3], , Carrie Eckert[1], Bradley Prythero[1,2], Jianping Yu[1], Pin-Ching Maness[1], and **Ryan Gill**[2,3]* (rtg@colorado.edu)

[1]Biosciences Center, National Renewable Energy Laboratory, Golden, CO; [2]Department of Chemical and Biological Engineering, University of Colorado, Boulder; [3]Renewable and Sustainable Energy Institute, NREL/University of Colorado, Boulder

Project Goal: This project aims to apply rational design and state-of-the-art synthetic and systems biology tools to design and optimize *E. coli* for sustained production of biofuels. Chassis biofuel strains, optimized for production based on predictive design and systems biology knowledge, will serve as the framework for high throughput genome re-design. Using targeted genome-scale and multiplex genome-engineering technologies, strains with improved production will be selected for, and gene-to-trait mapping will identify key factors for further optimization. Herein, we will focus on the construction of an *E. coli* prototype chassis strain for the production of ethylene and a selection strategy aimed to improve thermal stability of the key enzyme involved in ethylene production. Moreover, we are developing high-throughput strategies for selection of strains with increased biofuel/precursor levels.

Abstract

Ethylene is the most highly utilized organic compound for the production of plastics and chemicals, and its catalytic polymerization to alkane fuels has been demonstrated. At present, global ethylene production involves steam cracking of a fossil-based feedstock, representing the largest CO_2-emitting process in the chemical industry. Biological ethylene production has the potential to provide a sustainable alternative while mitigating CO_2 emission. The expression of a single gene found in some bacteria and fungi, ethylene-forming enzyme (*efe*), can catalyze ethylene formation (1, 2). Construction of the first generation chassis strain is based on *E. coli* MG1655 as the host and the *efe* gene from *Pseudomonas syringae* (*Ps*). EFE has been postulated to catalyze ethylene production according to the equation (3):

3 α-ketoglutarate + Arginine + $3O_2 \rightarrow 2C_2H_4$ + Succinate + $7CO_2 + 3H_2O$ + guanidine + P5C*

The two key substrates α-ketoglutarate (AKG) and arginine are tightly controlled by an intricate regulatory network that coordinates carbon and nitrogen metabolism (Figure 1). We conducted genetic modifications to rewire central carbon metabolic flux and improved ethylene production by 2.3-fold (4). This chassis strain will serve as the framework to guide genome-scale redesign and optimization to further boost ethylene production. Succinate is a byproduct of the EFE reaction. We generated a succinate auxotroph in *E. coli* and showed that it must rely on an active heterologous EFE pathway yielding succinate to afford growth. EFE is not stable above 30 °C. We thereby screen for thermal stable EFE by expressing an *efe* mutant library in the succinate auxotroph and select for growth at 37 °C. We identified key mutations of *efe* mapped to semi-conserved residues in EFE homologues with its outcomes unraveling the catalytic mechanism of EFE. Work is also ongoing to construct high-throughput sensors to screen for AKG and ethylene,

in situ. As such, current work from our groups at the National Renewable Energy Laboratory and the University of Colorado at Boulder seeks to improve ethylene production by combining traditional metabolic engineering strategies with synthetic biology-enabled evolutionary approaches involving the high-throughput construction of genome-scale libraries. Coupled with novel screens and selections, these methods will identify strains with increased production of key intermediates and/or ethylene.

*P5C: L-Δ^1-pyrroline-5-carboxylate

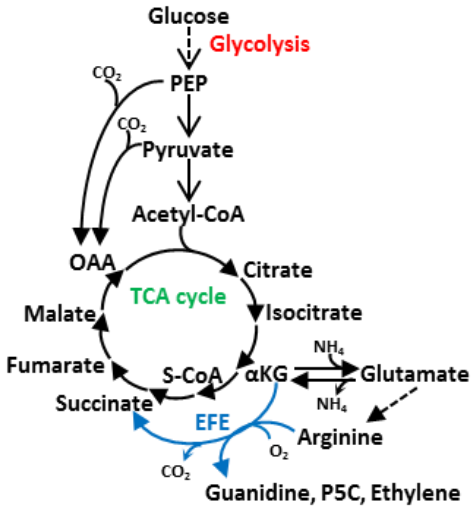


Figure 1. Putative metabolic scheme for ethylene production in *E. coli*. EFE: ethylene-forming enzyme. P5C: L-Δ^1-pyrroline-5-carboxylate.

References

1. Fukuda, H., T. Ogawa, M. Tazaki, K. Nagahama, T. Fujii, S. Tanase, and Y. Morino. 1992. Two reactions are simultaneously catalyzed by a single enzyme, the arginine-dependent simultaneous formation of two products, ethylene and succinate, from 2-oxoglutarate by and enzyme from *Pseudomonas syringae*. Biochem. Biophys, Res. Comm. 188: 483-489.
2. Eckert, C., W. Xu, W. Xiong, S. Lynch, J. Ungerer, L. Tao, R. Gill, P. C. Maness, and J. Yu. 2014. Ethylene-forming enzyme and bioethylene production. Biotech. Biofuels 7:33-43.
3. Fukuda, H., T. Ogawa, M. Tazaki, K. Nagahama, T. Fujii, S. Tanase, and Y. Morino.. 1992. Molecular cloning in Escherichia coli, expression and nucleotide sequence of the gene for the ethylene-forming enzyme of *Pseudomonas syringae* pv. *phaseolicola* PK2. Biochem. Biophys. Res. Comm. 826-832.
4. Lynch, S., C. Eckert, J. Yu, R. Gill, and P. C. Maness. 2016. Overcoming substrate limitations for improved production of ethylene in *E. coli*. Biotech. Biofuels 9:3, DOI 10.1186/s13068-015-0413-x

Funding Statement

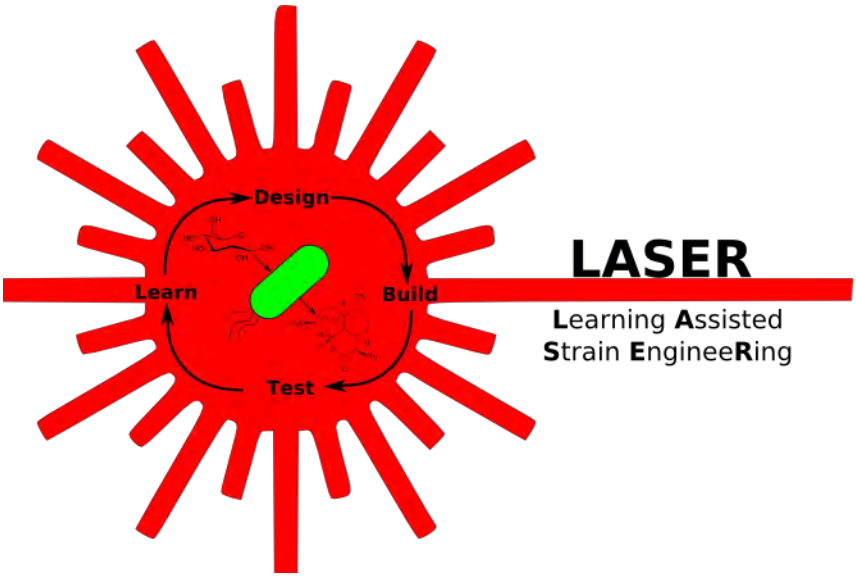
Grant title: A Platform for Genome-scale Design, Redesign, and Optimization of Bacterial Systems; Project grant number (DE-SC008812) and FWP number (ERWER44) for NREL.

The LASER Database: Building a Learning Platform for Metabolic Engineers

James D. Winkler* (james.winkler@colorado.edu), Andrea L. Halweg-Edwards* (andrea.edwards@colorado.edu), **Ryan T. Gill**

Department of Chemical and Biological Engineering, University of Colorado-Boulder

Renewable and Sustainable Energy Institute, University of Colorado-Boulder



Project Goals: Metabolic engineers have successfully engineered microorganisms to produce compounds ranging from biofuels to pharmaceuticals. However, this engineering process is largely *ad hoc*, with the majority of biocatalysts still being developed through researcher intuition instead of formalized modeling. We first developed the 2014 LASER (Learning Assisted Strain EngineeRing) database to address this issue by providing a standardized platform and software tools for analyzing engineered strains that could be used for quantitative field analysis. After a year of additional data curation and software development, we now introduce the expanded LASER 2015 (http://laser.colorado.edu/) database, containing over 600 curated designs from more than 400 papers. Key areas of expansion include the establishment of the first complexity metrics for assessing the engineering effort required for developing a given biocatalyst, development of machine learning-based approaches for developing convenient chassis strains, and a comprehensive "state of the field" overview of metabolic engineering. We also introduce a unique approach towards estimating the complexity of metabolic engineering designs, in order to reduce the experimental effort required for implementing potential design libraries. We anticipate that LASER will remain a powerful basis for understanding and improving metabolic engineering practices for years to come.

Project Website: laser.colorado.edu

HT-CRISPRi studies of gene regulation and function in *E. Coli*

Harneet S. Rishi[1,2] (HSRishi@lbl.gov), Esteban Toro,[2,3] David Chen,[3,4] Guillaume Cambray,[5] Honglei Liu,[6] Xiaowo Wang,[6,7] Lei S. Qi,[8,9] and **Adam P. Arkin**[2,3]

[1]Biophysics Graduate Program, University of California Berkeley, Berkeley, CA; [2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [3]Department of Bioengineering, University of California Berkeley, Berkeley, CA; [4]UC Berkeley-UCSF Graduate Program in Bioengineering, Berkeley, CA; [5]California Institute for Quantitative Biosciences, University of California Berkeley, Berkeley, CA; [6]Bioinformatics Division, Center for Synthetic and Systems Biology, Tsinghua National Laboratory for Information Science and Technology, Beijing, China; [7]Department of Automation, Tsinghua University, Beijing, China; [8]Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, CA; and [9]California Institute for Quantitative Biomedical Research, San Francisco, CA

http://genomicscience.energy.gov/biosystemsdesign/2012awards.shtml

Project Goals: Synthetic biology aims to leverage the engineering principles of modularity, standardization, and reliability with the design-build-test-learn cycle to rapidly engineer novel biological functions. One of the key hurdles in adopting this strategy is that the innate complexity of biological systems (ex: unmapped interaction networks, context dependence, temporal & spatial population variations) makes it difficult to understand first principles, which in turn makes it difficult to predictably build scalable systems. High-throughput technologies to quantitatively characterize function-phenotype landscapes can help overcome such barriers. As such, we aim to (1) develop CRISPRi into a scalable, high-throughput platform for rapid profiling of genomic features in *E. coli* across a number of conditions and (2) extend the functionality of the HT-CRISPRi platform to perform double transcriptional knockdowns for investigating the epistatic landscape of *E. coli* and revealing the basis of complex traits.

The CRISPR-associated protein Cas9 has been adapted as a versatile tool for transcriptional regulation, genome editing, and imaging in a number of organisms. Here, we apply the catalytically inactive dCas9 to conduct high-throughput transcriptional and regulatory studies in *E. coli*. Using an Agilent OLS library of 32992 unique sgRNAs, we targeted 4500 genes, 5400 promoters, 640 transcription factor binding sites, and 106 sRNAs in the *E. coli* genome. These genomic targets cover a wide range of functionalities such as metabolism, stress response, transport, and cell division. By leveraging CRISPRi with next-generation sequencing, we were able to interrogate the fitness effect of transcriptional knockdown for each of the aforementioned genomic features in a single-pot experiment both aerobically and anaerobically. Our fitness results agreed well with current knockout databases, and our ability to induce transcriptional knockdown at any point during an experiment has allowed us to explore target essentiality under different conditions (aerobic, anaerobic, minimal media, etc.) with great ease. We demonstrated this by showing that although *nrdA* and *nrdB* are essential under aerobic conditions – and are annotated as such in databases – they are dispensable anaerobically. We also show that CRISPRi can recapitulate genomic features with redundant functions. Furthermore, our genome-wide

library has allowed us to explore the robustness of CRISPR interference within bacterial operons. We are currently extending the HT-CRISPRi platform to scalably perform double transcriptional knockdowns in *E.* coli in order to interrogate libraries of knockdown pairs for epistatic genetic interactions. Overall, HT-CRISPRi enables single-pot, precise measurements of fitness for a large set of genomic features and will prove useful in genomic studies of model and non-model organisms.

This work is supported by the Genome Science program within the Office of Biological and Environmental Research (Project grant number DE-SC008812, Funding Opportunity Announcement DE-FOA-0000640*).*

RNase III as a Tool to Manipulate Transcript Stability: Identifying *in vivo* Targets in *Escherichia coli*

Gina Gordon,[1,*] (gcgordon@wisc.edu), Jeffrey C. Cameron,[2] and **Brian F. Pfleger**[1]

[1]University of Wisconsin, Madison; [2]University of Colorado, Boulder

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

Project Goals:

For this project we will use RNA-sequencing techniques to study mRNA stability in *Escherichia coli* and *Synechococcus* sp. PCC7002, a sequenced cyanobacteria with promising industrial traits. We will use this method to study how various RNA processing enzymes affect decay of mRNA from a global perspective. We will use the knowledge gained from our studies of mRNA turnover to develop design rules for (de)stabilizing transcripts and apply these rules to ongoing metabolic engineering projects in PCC 7002 for use in photosynthetic biorefineries.

Our specific objectives are:

(1) Quantify decay rates for each nucleotide in the *E. coli* and PCC 7002 transcriptome with RNA sequencing
(2) Design and test strategies for (de)stabilizing transcripts in PCC 7002
(3) Apply design rules to improving biofuel production in PCC 7002

Messenger RNA (mRNA) is a labile intermediate that affects protein expression levels. For metabolic engineering purposes we would ideally be able to predict and precisely control protein expression based on a given DNA sequence. Unfortunately there is insufficient foundational knowledge about RNA stability to be able to predict the half-life of a given transcript. This arises due to the numerous and complex ribonucleases (RNases) that facilitate mRNA degradation. We are particularly interested in RNase III, a ribonuclease that cleaves long double-stranded regions of RNA, because of its potential use to alter transcript stability and process structure RNAs (e.g. CRISPR guide RNAs). RNase III recognizes mRNA secondary structure and not a conserved nucleotide sequence, but it is still more amenable to engineering than other RNases that have less selectivity. Using RNA-seq we have identified novel targets of RNase III in *Escherichia coli*, and we have shown that RNase III regulates protein expression of important metabolic enzymes, processes read-through transcripts, and is involved in the turnover of mRNA of leader peptides. With our greatly expanded list of RNase III sites and sequences, we will test if these sites can be placed next to genes of interest to alter their stability and subsequent protein expression.

This work is supported by US-Department of Energy (DE-SC0010329) and a NIH Biotechnology Training Fellowship to Gina Gordon (NIGMS- 5 T32 GM08349).

RDP: Data and Tools for Microbial Community Analysis

Benli Chai[1]* (chaibenl@msu.edu), Yanni Sun,[1] C. Titus Brown,[2] James M. Tiedje,[1] and **James R. Cole**[1]

[1]Center for Microbial Ecology, Michigan State University, East Lansing. MI; [2]University of California-Davis, Davis, CA

http://rdp.cme.msu.edu
http://fungene.cme.msu.edu
https://github.com/rdpstaff

Project Goals: RDP offers aligned and annotated rRNA and important ecofunctional gene sequences with related analysis services to the research community. These services help researchers with the discovery and characterization of microbes important to bioenergy production, biogeochemical cycles, climate change, greenhouse gas production, and environmental bioremediation.

RDP's data collections include 3,224,600 16S rRNA and 108,901 fungal 28S rRNA sequences as of February 2016. Over the past year, RDP websites (Cole et al., 2014) were visited, on average, by 7805 researchers (unique IPs) in 15,464 analysis sessions each month.

During 2015, we updated the the RDP Classifier (Wang et al., 2007) and RDP Taxonomy three times to reflect recently discovered bacterial, archaeal, and fungal lineages and latest taxonomic emendations. The RDP Taxonomy now models over 2500 genera (including about 100 unofficial genera), an increase of over 800 genera, with over 13,000 training sequences. Most RDP tools are now available as open source command-line versions through RDP's GitHub repository (https://github.com/rdpstaff). This includes our recently published Xander software package (Wang et al., 2015). Xander incorporates our novel method for assembling protein-coding sequences for genes of interest from a metagenomic dataset. In addition to the software packages, the repository includes additional resources including examples, documentation and tutorials. These command-line tools provide researchers with an independent method to analyze their own data, including high-throughput data and many of these tools are already used in third-party pipelines. These stand-alone versions of our tools have been created for easy porting to KBase in the future.

RDP's FunGene Pipeline & Repository (Fish et al., 2013) provides databases for 264 protein coding genes useful as phylogenetic markers and for following important ecological functions. In addition to the aligned and annotated gene and protein sequences, FunGene provides online analysis functions and tools for selecting subsets of sequences for download and further analysis. Use of the FunGene web, on average, was 1069 researchers per month in 1753 analysis sessions. During the past year, we updated FunGene data releases nine times from searches of the primary sequence databases.

We have made several performance improvements to the website to make access to sequence data much faster, especially for genes with a large number of data. In addition to optimizing existing gene models in N and C cycles, we have added more genes of environmental importance, such as the *acdS* gene which promotes microbe-mediated plant growth and tolerance to drought and salinity stresses, a group of key genes in metabolic responses to environmental toxicants such as polychlorinated dibenzo-*p*-dioxins (PCDDs), as well as antibiotic resistance gene data using Resfams reference sequences. In addition, we are leveraging FunGene to provide training data for our Xander gene-targeted assembler and for high-throughput qPCR gene-targeted primer and probe development.

References

Cole, J. R., Q. Wang, J. A. Fish, B. Chai, D. M. McGarrell, Y. Sun, C. T. Brown, A. Porras-Alfaro, C. R. Kuske, and J. M. Tiedje. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42(Database issue): D633-D642. doi: 10.1093/nar/gkt1244

Fish, J. A., B. Chai, Q. Wang, Y. Sun, C. T. Brown, J. M. Tiedje, and J. R. Cole. (2013). FunGene: the functional gene pipeline and repository. *Front. Microbiol.* 4: 291. doi: 10.3389/fmicb.2013.00291

Wang, Q., J. A. Fish, M. Gilman, Y. Sun, C. T. Brown, J. M. Tiedje, and J. R. Cole. (2015). Xander: employing a novel method for efficient gene-targeted metagenomic assembly. *Microbiome* 3: 32 doi: 10.1186/s40168-015-0093-6

Wang, Q, G. M. Garrity, J. M. Tiedje, and J. R. Cole. (2007). Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl Environ Microbiol.* 73: 5261-5267. doi: 10.1128/AEM.00062-07

This research was supported by the Office of Science (BER), U.S. Department of Energy Grant No. DE-FG02-99ER62848, with contributions from Office of Science (BER), U.S. Department of Energy Grant Nos. DE-SC0014108, DE-SC0010715, DE-FC02-07ER64494, NIEHS Superfund Research Program Award 5P42ES004911-23 and National Science Foundation Award No. DBI-1356380.

Computational Analysis of Microbial Metabolism for Fuel and Chemical Production

Xiaolin Zhang[1], Christopher J. Tervo[1], Trang T. Vu[1], Hugh Purdy[1]* (hmpurdy@wisc.edu), and **Jennifer L. Reed**[1]

[1]Dept. of Chemical and Biological Engineering, Univ. of Wisconsin-Madison, Madison, WI

Project Goals: Cyanobacteria offer a promising route for directly converting solar energy and CO_2 into biofuels. The objectives of this research are to integrate modeling and experimental approaches to guide development of a butanol producing cyanobacterium, *Synechococcus* sp. PCC 7002. New computational approaches will be developed to facilitate these efforts which will (1) design experiments and analyze their results, and (2) identify genetic engineering strategies for improving butanol production in *S.* 7002. Experiments will subsequently be performed to construct and analyze *Synechococcus* 7002 strains engineered for butanol production. The developed approaches will be systematically applied to suggest genetic engineering strategies for improving production of a variety of biofuels in five other microorganisms. This research will support the U.S. Department of Energy's mission for developing renewable ways of producing advanced biofuels.

Renewable sources of transportation fuels and chemicals are needed to reduce the amount of oil used to satisfy transportation energy needs in the U.S. and to alleviate our dependence on foreign sources of oil. Microbes can be used to produce a wide variety of liquid biofuels including: ethanol, butanol, isobutanol, isoprene, hydrogen, and alkanes. Cyanobacteria offer an alternative route for converting solar energy and CO_2 into biofuels, without the need for using lignocellulosic biomass as an intermediate. The biofuel production capabilities of microbes can be improved through metabolic engineering, where metabolic and regulatory processes are adjusted using targeted genetic manipulations. Traditionally, metabolic engineering strategies are found through manual inspection of metabolic pathways, where enzymes involved in biosynthesis are overexpressed or added, competing pathways are eliminated, and the performance of resulting strains are evaluated. However, such approaches cannot predict the effects that these changes will have on other parts of metabolism, and generally will not suggest alterations to more distant pathways. Computational models of cellular metabolic and regulatory networks can be used instead to guide and accelerate these metabolic engineering efforts by integrating and analyzing experimental data, and identifying genetic manipulations that would increase product yields.

We systematically evaluated what potential products could be produced by *Escherichia coli* (using both native and heterologous pathways) and then evaluated their distance (in terms of metabolic reactions steps) from central metabolic precursors [1]. Using a genome-scale metabolic model of *E. coli* and a set of potential heterologous reactions (from the KEGG database), ~1,800 non-native products could potentially be produced in *E. coli* using heterologous enzymes, with ~300 having commercial applications. Subsequent analysis found

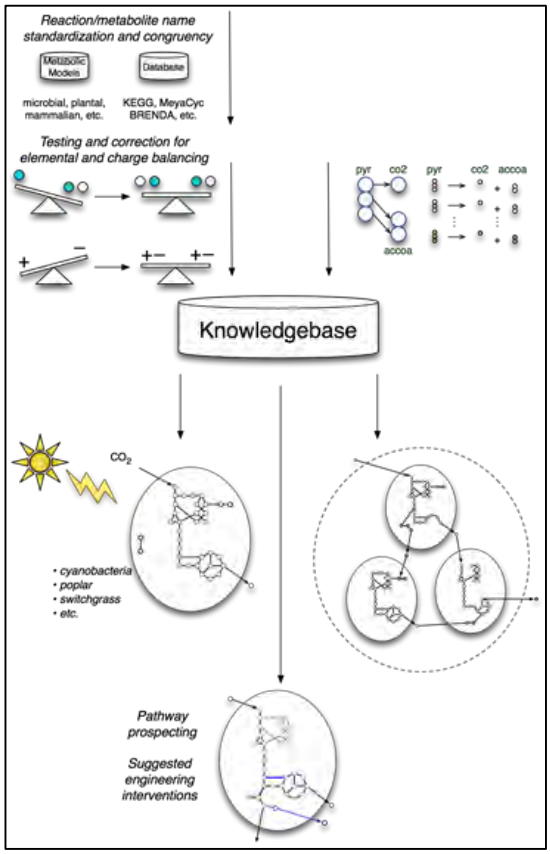
Development of a Knowledgebase (MetRxn) of Metabolites, Reactions and Atom Mappings to Accelerate Discovery and Redesign

Akhil Kumar[2]*(azk172@psu.edu), Saratram Gopalakrishnan[1]*(sxg375@psu.edu), Margaret Simons[1], Siu Hung Joshua Chan[1], and **Costas D. Maranas[1](costas@psu.edu)**

[1]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA; [2]The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA

http://www.maranasgroup.com/

Project Goals: The project aims to expand genome-scale metabolic (GSM) models to include complex interactions through multi-tissues and multi-organisms or complex constraints based on atom-mapping information and Metabolic Flux Analysis (MFA). The whole-organism GSM model of maize will be used to determine bottlenecks in nitrogen metabolism and suggest genetic manipulations to improve nitrogen use efficiency. The multi-organism models will be used to explain experimental observations, analyze the physiological responses, and predict the interactions within microbial communities. We also developed tools and algorithms to organize and disseminate standardized metabolite and reaction information, map atom transitions from reactants to products, identify gene to protein to reaction relations, and design novel biotransformations to xenometabolites. This will facilitate the construction of a mapping model for flux elucidation using 13-C MFA at the genome-scale to provide insights into the role of model scale-up and biomass composition on metabolic flux resolution.



Genome-scale metabolic (GSM) models are a platform used to investigate the metabolic behaviors of complex organisms, microbial communities, and model organisms. Flux balance analysis (FBA) of reconstructed multi-tissue and multi-organism models can be used to determine interactions between different cell/tissue-types or organisms, resolve bottlenecks in limiting pathways, and study the metabolic trade-offs between species-level and community-level fitness functions. FBA often reports large solution spaces which can be further constrained using genome-scale MFA facilitated by isotope-labeling data and incorporating atom-mapping.

A whole plant model of maize was developed by reconstructing the root, stalk, leaf, kernel and tassel tissues using the phloem as the main form of metabolite transport among the tissues using transcriptomic data to determine the set of reactions included in each tissue. This model was simulated at three different growth conditions: vegetative leaf growth, tassel development, and kernel filling. Growth rate proportions were estimated for each growth stage and dry weight proportions were used to normalize the transport reaction flux between tissues. As the plant developed from the vegetative growth stage to the tassel development stage, 138 reactions had differing flux ranges. However, from the tassel growth stage to the kernel filling stage, only 77 reaction flux ranges differed.

Multi-organism models also facilitate the investigation of synergy among microbiota, the interactions between the community and the host, and the effect of the diet on human health. A gut microbial model was developed to represent the metabolism in two major clades of bacteria (i.e. Firmicutes and Bacteroidetes), along with *Lactobacillus*. A reduced abundance of *Lactobacillus* in the gut microbiota and its bile salt hydrolase activity leads to the accumulation of the conjugated bile acid tauro-β-muricholic acid

and inhibits the intestinal farnesoid X receptor. By contrasting the minimum nutrients required in the presence and absence of bile salt, we found that the availability of a sulfur source is the single nutrient essentiality that differed between the two cases. In another study, we used a core microbial model of the representative Firmicutes and Bacteroidetes, as well as *Lactococcus garvieae* to model equol production in the gut community. It has been reported by Magee et al. that a higher fraction of the Asian population is able to produce equol compared to the Western population, however increased soy consumption did not convert a western non-producer into an equol producer. By modeling equol production at various growth ratios, we observed that equol production was indeed maximal at high *L. garvieae* abundance and low sugar availability to *L. garvieae*. In a low sugar environment, it appears that *L. garvieae* coverts diadzein (a soy ingredient) into equol as a means of replenishing the NAD^+ pool.

Reaction and metabolite data for construction of both maize and gut microbial metabolic model were primarily derived from the curated MetRxn database. The MetRxn knowledgebase is a unified repository of metabolite and reaction information from various metabolic models and databases. Overlapping information from 8 databases and 112 metabolic models was curated and standardized into 44,784 unique reactions and a million plus unique metabolites. During curation, incompatibilities related content representation, stoichiometric errors such as elemental or charge imbalances, and incomplete atomistic details were resolved and corrected for using a host of cheminformatics, lexicographic and phonetic algorithms. Users can access the standardized repository on a searchable web interface at www.metrxn.che.psu.edu. In addition, all charge and mass balanced reactions within the database are processed by our novel algorithm; Canonical Labelling for Clique Approximation (CLCA). CLCA leverages prime factorization to quickly generate unique molecular graphs, detect symmetries for all metabolites, and resolve bond transformation and atom transition information in each reaction. Bond transformation information was further leveraged to recognize the recurring schemes in substrate to product conversions across all reactions. The common substrate-product transformation schemes identified in 44,784 reactions have been encoded as 6,211 reaction rules within MetRxn. These reaction rules will enable users to annotate the otherwise regarded specialist enzyme with putative secondary activity; and design novel biotransformation schemes to various xenometabolites. Enzyme promiscuity annotations are expected to expand the metabolic potential and fill the many biochemical gaps between phenotype and model predictions.

Metabolic models used in 13-C MFA generally include a limited number of reactions from the central metabolic network. CLCA atom transition information is utilized for the construction of genome-scale mapping models (GSMM) to demonstrate the feasibility of genome-scale MFA and address the impact of model scale-up on prediction fidelity of metabolic fluxes using 13-C MFA. We have compared and contrasted fluxes and ranges estimated by minimizing the sum of square of differences between predicted and experimentally measured labeling patterns using a core model (75 reactions and 65 metabolites) based on central metabolism and a GSMM (697 reactions and 595 metabolites) obtained upon eliminating inactive reactions from *i*AF1260. While both the topology and estimated values of the metabolic fluxes remain largely consistent between the base and GSMM, 20 key reactions in central metabolism and transhydrogenase have wider flux ranges in the GSMM due to the possible activity of alternate routes and futile cycles. Inferred ranges for 81% of the reactions in the GSM model varied less than one-tenth of the basis glucose uptake rate because as many as 411 reactions in the GSM are growth coupled and determined by measured growth rate. Finally, the loss of information associated with mapping fluxes from MFA on a core model to a GSM model is quantified to reveal that the flux range of 295 reactions was narrower than supported by data, demonstrating that assumptions made during the core model construction propagate onto the GSM model leading to possibly erroneous conclusions about reaction flux identifiability.

Supported by funding from the U.S. Department of Energy by grant DE-FG02-05ER25684.

that pyruvate was the closest central metabolic precursor to the most non-native commercial products. Since pyruvate has industrial applications and can be converted into valuable products, we sought to develop a strain of *E. coli* that could produce pyruvate at high yields. Guided by a genome-scale metabolic model of *E. coli*, we then identified different strategies for enhancing production of pyruvate from glucose. We constructed a number of strains which achieved yields up to 0.92 g pyruvate per g substrate (~95% theoretical yield) and which can be used to produce other biofuels and biochemicals (Zhang and Reed, in preparation). These results illustrate how computational models can be used to prioritize precursor-based strategies and identify genetic modifications to enhance precursor production.

Current efforts are extending these analyses to evaluate chemical production in the cyanobacerium, *Synechococcus sp*. PCC 7002. Strain designs for enhancing butanol (and other chemicals) production in *Synechococcus sp*. PCC 7002 have been made using computational models and are currently being constructed [2].

References

1. Zhang X, Tervo CJ, and JL Reed. Metabolic Assessment of E. coli as a Biofactory for Commercial Products. *Metabolic Engineering*. Accepted (2016).
2. Vu TT, Hill EA, Kucek LA, Konopka AE, Beliaev AS, and JL Reed. Computational evaluation of *Synechococcus* sp. PCC 7002 metabolism for chemical production. *Biotechnology Journal*, 8(5):619-30 (2013).

This work was supported by the Office of Science (BER), U.S. Department of Energy (DE-SC0008103), NSF (1053712), and W.M. Keck Foundation.

Semantic Index of Phenotypic and Genotypic Data

Charles T. Parker,[1]* (chuck.t.parker@namesforlife.com), Nenad Krdzavac,[2] Kevin Petersen,[2] Grace Rodriguez,[1] Oishi Bagchi,[2] and **George M. Garrity**[1,2]

[1]NamesforLife, LLC; East Lansing, Michigan; [2]Michigan State University; East Lansing, Michigan

http://prokaryote.info

Project Goals: The goal of this project is to develop a semantic data resource that can serve as a basis for predictive modeling of microbial phenotype. The core technical objectives are twofold: (1) to build a database of normalized phenotypic descriptions (observational data) using the primary taxonomic literature of bacterial and archaeal type strains, and (2) to construct an ontology with reasoning capabilities to make accurate phenotypic and environmental inferences based on that data. This project is tightly coupled with ongoing DOE projects (the Genomic Encyclopedia of Bacteria and Archaea, the Microbial Earth Project, the Community Science Project) and with two key publications, Standards in Genomic Sciences (SIGS) and the International Journal of Systematic and Evolutionary Microbiology (IJSEM).

Predictive modeling of cells and metacommunities requires high quality input data, but not all data are of similar quality nor are they amenable to computational analysis without extensive cleaning, interpretation and normalization. Key among those are phenotypic data, which are more complex than sequence data, occur in a wide variety of forms, often use complex and non-uniform descriptors and are scattered about specialized databases and scientific and technical literature. Incorporating phenotypic information into metagenomic analyses requires expertise in harvesting, modeling and interpreting these data.

The Semantic Index of Phenotypic and Genotypic Data will address this problem by providing a resource of reference phenotypic data for all validly published type strains of Bacteria and Archaea, based on concepts and observational data drawn from the primary taxonomic literature. In the Phase I project we developed software to construct and analyze a corpus of this literature and to extract putative feature domain vocabularies comprising approximately 40,000 candidate phenotypic terms used in new and emended descriptions of the 13,213 distinct type strains of Bacteria and Archaea. In Phase II, these vocabularies have served as the basis for developing a phenotypic ontology, a repository of phenotypic data that is undergoing normalization of phenotypic descriptions for each species. We have found that many of the phenotypes applied to microbes describe a combination of quantitative environmental conditions and qualitative growth and metabolic capabilities. Such terms have proven difficult to implement in query systems because of their context-based interpretations and conceptual overlap across multiple feature domains. We have furthered our work on novel design patterns for ontology development [1] that address these problems and remove barriers to machine reasoning and complex query

answering over these complex terms, while preserving the bi-directional mapping back to human interpretation at multiple levels of abstraction. An international patent application on this method published in 2015 and is expected to proceed into the National Phase in 2016. Critical to implementing this system has been the adoption of a SPARQL Inferencing Notation (SPIN) reasoner coupled with a triple store rule-based reasoner. This approach resolves ambiguity attributed to the semantic equivalence and imprecision of phenotypic terms arising in literature and in databases. In order to better facilitate access to knowledge extracted from the literature and encoded in the ontology, we are implementing a special-purpose web portal to accommodate query and retrieval of biological resources by term or concept, with a multi-tier query platform conforming to current search standards and backed by Semantic Web and Linked Data query standards. In addition to linking to the primary literature, related ontologies and source data, we are also incorporating public data from NCBI, USDA and the Joint Genome Institute in order to provide researchers with immediate access to the appropriate resources for a set of strains, along with consistent, accurate interpretations of available knowledge about those strains that are usable for predictive modeling and in other research and commercial applications. As part of a participation in a broader ontology ecosystem, we are leveraging existing ontologies, including CHEBI, to provide mapping between representations of concepts in other knowledge resources.

As part of our commercialization activities, we continue to develop several software components resulting from this project into a commercial semantic search and document analysis platform with end products being used in Standards in Genomic Sciences and the International Journal of Systematic and Evolutionary Microbiology.

Publications

1. Parker, CT, Garrity, GM and Krdzavac, NB. Systems and Methods for Establishing Semantic Equivalence Between Concepts. U.S. Patent Application No. PCT/US2014/056808. Filed September 22, 2014. Washington, DC: U.S. Patent and Trademark Office. WIPO Application WO/2015/042536 A1. Published March 26, 2015.
2. Parker, CT, Garrity, GM, Ussery, D, Wanchai, V and Nookaew, I. Method for Identification and Characterization of Prokaryotes Using Whole Genome DNA Sequence Data. US Provisional Patent Application No. 62/232,925. Filed September 25, 2015. Washington, DC: U.S. Patent and Trademark Office.
3. Parker CT, Tindall BJ, Garrity GM. International Code of Nomenclature of Prokaryotes. *Int J Syst Evol Microbiol*. 2015. doi:10.1099/ijsem.0.000778 PMID:26596770

Funding for this project was provided through the DOE SBIR/STTR program (DE-SC0006191). Public funding for development of the NamesforLife infrastructure was received from the DOE SBIR/STTR program (DE-FG02-07ER86321), the Michigan Small Business Technology Development Corporation, the Michigan Strategic Fund, and the Michigan Universities Commercialization Initiative.

High-quality Genome-scale Metabolic Reconstructions for Multi-scale Microbial Communities

Laura Kutt* (laura.kutt@ext.uni.lu), Stefania Magnusdottir, Almut Heinken, Eugen Bauer, Dimitry A. Ravcheev, Ronan M.T. Flemingand **Ines Thiele**

Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg.

http://stanford.edu/group/SOL/multiscale/

Project Goals: There are an increasing number of automatically reconstructed genome scale metabolic models available for various organisms from databases, such as KBase; however, these models require further microbe specific refinement. The refinement process is time-consuming and laborious. This project introduces a pipeline that reduces many errors that are associated with automatic reconstructions. The reconstructions are modified to include correct reaction directionalities and to allow feasible ATP production. Further refinement focuses on manual curation of the models against current available knowledge. It involves extensive literature search for carbon sources and fermentation products, and comparative genomic approach for central metabolism, respiration, and amino acid biosynthesis pathways. The inclusion of these pathways increases the predictive power of the models. High quality metabolic models can be then applied to genetic engineering of specific microbes to increase the yield or degradation of compound of interest, or investigate microbe-microbe relationships in complex microbial communities.

In silico genome scale metabolic models are often used in genetic engineering to optimize the microbe metabolism of for a specific task, for example for degradation of environmental pollutants in the case of oil spills. This type of optimization is usually done by flux balance analysis, which allows evaluation of gene deletions in the context of metabolic system. Often the elimination of a specific metabolic enzyme does not lead to the increased yield of the compound of interest due to compensatory effects in metabolic systems. In silico models enable the production of possible solutions, which can be tested in laboratory. However, the predictions depend on the quality of the genome-scale metabolic models used. The Department of Energy Systems Biology Knowledgebase (KBase) provides tools for automatic generation of these genome scale metabolic models. However, these models need substantial improvement to provide reliable microbe specific predictions. Our pipeline identifies the main errors associated with the draft models, and undertakes manual curation of species-specific pathways. The pipeline first checks the metabolic networks for reaction directionalities, internal flux loops and futile cycles, which are common problems in the automatic draft models, for example due to high number of transport reactions. The second stage involves extensive manual curation of various metabolic pathways such as fermentation products, carbon sources, amino acid biosynthesis, central metabolism, and respiratory pathways. This refinement is necessary due to high variability between different genera of bacteria, for example, uptake of carbon sources is highly species specific. This pipeline also checks for anaerobic growth, and the increased biomass production in the presence of oxygen. Anaerobic growth should be enabled for microbes that thrive in anaerobic environments. This pipeline fulfils a need for faster reconstruction of high quality models, especially in the light of increasing availability of (meta-) genomic data. We also developed a computational toolbox for modeling of multi-scale microbial communities as well as the tailoring of the microbial community models based on meta-omics data. The reconstruction and modeling software is implemented in Matlab (Mathworks, Inc., Natic, MA, USA) and shall form an extension to The COnstraint-Based Reconstruction and Analysis (COBRA) Toolbox [1].

References

[1] Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols 6: 1290–1307.

This work was in part supported by the U.S. Department of Energy, Offices of Advanced Scientific Computing Research and the Biological and Environmental Research as part of the Scientific Discovery Through Advanced Computing program, grant #DE-SC0010429.

The COBRA Toolbox: a comprehensive and powerful modelling resource

Hulda S. Haraldsdóttir[1], Hoai Minh Le[1], Longfei Mao[1], Ines Thiele[1] and **Ronan M. T. Fleming**[3]*
(ronan.mt.fleming@gmail.com).

[1]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

https://opencobra.github.io/cobratoolbox/

Project Goals: This project aims to develop COBRA Toolbox: a package of methods to simulate, analyse, predict and visualise a variety of metabolic phenotypes using genome-scale models.

Over the past decade, COnstraint-Based Reconstruction and Analysis (COBRA) methods have widely been used in several fields including microbial metabolic engineering, modelling of gene expression and metabolism at genome-scale on modelling transcriptional and signaling networks. The COBRA Toolbox, a MATLAB package of COBRA methods, provides systems biology researchers with a high-level interface to a variety of methods for constraint-based modelling of genome-scale stoichiometric models of cellular biochemistry. The first version of COBRA Toolbox, released in 2007, contains a variety of methods, such as flux variability analysis, growth-rate optimisation, robustness analysis, gene essentiality analysis and model curation tools. Version 2.0 [3] expanded the first version by adding new functions including network gap filling, C13 analysis, metabolic engineering, omics-guided analysis and visualisation. The COBRA Toolbox has now gained widespread use and has become a leading software package for genome-scale analysis of metabolism. The success of COBRA Toolbox has been used in more than 500 studies. The COBRA Toolbox has rapidly developed in recent years based on contributions from a growing number of researchers. Our recently contributed functions to the COBRA Toolbox include: (i) consistent estimation of Gibbs energy using component contributions [2], (ii) identification of conserved moieties in metabolic networks [1], (iii) reconstruction context-specific metabolic network models from generic reconstructions [5], (iv) robust flux balance analysis [4], (v) Cell Designer - Matlab interface for parsing and editing files, and (vi) tools for conversion of a reconstruction into a COBRA model in a quality controlled manner. The latest version of The COBRA toolbox is available at https://github.com/opencobra/cobratoolbox and includes contributions from a community of >25 developers: https://github.com/opencobra/cobratoolbox/graphs/contributors.

References

[1] Hulda S. Haraldsdóttir and Ronan M. T. Fleming. Identification of conserved moieties in metabolic networks by graph theoretical analysis of atom transition networks. *Submitted*, 2015.

[2] Elad Noor, Hulda S. Haraldsdóttir, Ron Milo, and Ronan M. T. Fleming. Consistent Estimation of Gibbs Energy Using Component Contributions. *PLoS Comput Biol*, 9(7):e1003098, July 2013.

[3] Jan Schellenberger, Richard Que, Ronan M. T. Fleming, Ines Thiele, Jeffrey D. Orth, Adam M. Feist, Daniel C. Zielinski, Aarash Bordbar, Nathan E. Lewis, Sorena Rahmanian, Joseph Kang, Daniel R. Hyduke, and Bernhard Ø Palsson. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols*, 6(9):1290–1307, September 2011. 00182.

[4] Yuekai Sun, Ronan MT Fleming, Ines Thiele, and Michael A. Saunders. Robust flux balance analysis of multiscale biochemical reaction networks. *BMC Bioinformatics*, 14(1):240, July 2013. 00000 PMID: 23899245.

[5] Nikos Vlassis, Maria Pires Pacheco, and Thomas Sauter. Fast reconstruction of compact context-specific metabolic network models. *PLoS Comput Biol*, 10(1):e1003424, January 2014. 00001.

This work was supported by the U.S. Department of Energy, Offices of Advanced Scientific Computing Research and the Biological and Environmental Research as part of the Scientific Discovery Through Advanced Computing program, grant #DE-SC0010429.

Quadruple-precision solution of genome-scale models of Metabolism and macromolecular Expression

Ding Ma[1], Laurence Yang[2], Ines Thiele[3], Bernhard Palsson[2], **Ronan M. T. Fleming**[3]*(ronan.mt.fleming@gmail.com), Michael Saunders[1]

[1]Dept of Management Science and Engineering, Stanford University, Stanford, CA, USA; [2]Dept of Bioengineering, University of California, San Diego, CA, USA; [3]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg.

http://stanford.edu/group/SOL/multiscale/

Project Goals: Development and biological applications of industrial quality linear and nonlinear optimization solvers for Constraint-Based Reconstruction and Analysis of multiscale mechanistic models of biochemical networks.

Constraint-Based Reconstruction and Analysis (COBRA) is currently the only methodology that permits integrated modeling of Metabolism and macromolecular Expression (ME) at genome-scale. Linear optimization can be used to compute steady-state flux solutions to ME models. Special care must be taken when computing such solutions as they contain fluxes that are spread over many orders of magnitude. Standard floating point solvers may return inaccurate solutions, or may have difficulty determining whether a feasible solution exists. Genome-scale ME models contain upwards of 70,000 variables and continue to increase in size, so accurate but inefficient exact simplex solvers are an impractical option. We developed qMINOS, a quadruple-precision version of MINOS, our industrial quality linear and nonlinear optimization solver. We tested the performance of qMINOS on a range of multiscale linear optimization problems and established a squential solution procedure involving double- and quadruple- precision simplex solvers that together achieve a balance between efficient and reliable solution to ME models. The prior existence of efficient numerical optimization algorithms enabled the exponential growth in biological applications of metabolic models. Likewise, qMINOS now guaruntees the accurate solution to linear, nonlinear, genome-scale and multi-scale ME models in reasonable time. Platform specific compiled qMINOS code has been disseminated to the community via the latest version of The COBRA Toolbox [1], available at https://github.com/opencobra/cobratoolbox.

References

[1] Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols 6: 1290–1307.

This work was supported by the U.S. Department of Energy, Offices of Advanced Scientific Computing Research and the Biological and Environmental Research as part of the Scientific Discovery Through Advanced Computing program, grant #DE-SC0010429.

Finding steady states in genome-scale biochemical reaction networks

Francisco J. Aragón Artacho[1], Phan T. Vuong[2]* (vuong.phan@uni.lu), and **Ronan M. T. Fleming**[2]

[1]Department of Statistics and Operations Research, University of Alicante, Spain; [2]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

http://stanford.edu/group/SOL/multiscale/

Project Goals: The problem of finding non-equilibrium steady states in genome-scale biochemical reaction networks is reformulated as an optimisation problem involving difference of convex functions (DC function). The resulted optimisation problem is then tackled by Boosted Difference of Convex function Algorithms (BDCA). Numerical tests on various biochemical models show the effiency of the implemented algorithms.

Many problems arising in science and engineering applications require the development of algorithms to minimise a nonconvex function. If a nonconvex function admits a decomposition, this may be exploited to tailor specialised optimisation algorithms. Our main focus is the following optimisation problem

$$\underset{x\in\mathbb{R}^m}{\text{minimise}}\ \phi(x) := f_1(x) - f_2(x),$$

where $f_1, f_2 : \mathbb{R}^m \to \mathbb{R}$ are continuously differentiable convex functions and

$$\inf_{x\in\mathbb{R}^m} \phi(x) > -\infty.$$

In biochemistry, this problem arises in the study of non-equilibrium steady states of biochemical reaction networks. We introduce two new algorithms to find stationary points of DC programs, called *Boosted Difference of Convex function Algorithms* (BDCA), which accelerate DCA [4] with a line search using an Armijo type rule. The first algorithm directly uses a backtracking technique, while the second uses a quadratic interpolation of the objective function together with backtracking. Our algorithms are based on both DCA and the proximal point algorithm approach of Fukushima–Mine [3]. We analyse the rate of convergence under the Łojasiewicz property [2] of the objective function. We discovered that the objective function arising in these biochemical reaction networks is real analytic, a class of functions which is known to satisfy the Łojasiewicz property [2]. Numerical tests on various biochemical models clearly show that our algorithm outperforms DCA, being on average more than four times faster in both computational time and the number of iterations [1]. This algorithm is guarunteed to find a non-equilibrium steady state concentration for any genome-scale biochemical network that admits one such steady state.

References

[1] Francisco J Aragón Artacho, Ronan MT Fleming, and Phan T Vuong. Accelerating the dc algorithm for smooth functions. *arXiv preprint arXiv:1507.07375*, 2015.

[2] Stanislaw Lojasiewicz. *Ensembles semi-analytiques*. Université de Gracovie, 1965.

[3] H. Mine and M. Fukushima. A minimization method for the sum of a convex function and a continuously differentiable function. *Journal of Optimization Theory and Applications*, 33(1):9–23, January 1981.

[4] Pham Dinh Tao and El Bernoussi Souad. Algorithms for solving a class of nonconvex optimization problems. methods of subgradients. In J.-B. Hiriart-Urruty, editor, *FERMAT Days 85: Mathematics for Optimization*, volume 129 of *North-Holland Mathematics Studies*, pages 249–271. Elsevier, 1986.

This work was supported by the U.S. Department of Energy, Offices of Advanced Scientific Computing Research and the Biological and Environmental Research as part of the Scientific Discovery Through Advanced Computing program, grant #DE-SC0010429.

Comparison of atom mapping algorithms for metabolic reactions

German Preciat[1], Hulda S. Haraldsdóttir[1]* (hulda.haraldsdottir@uni.lu), and **Ronan M. T. Fleming**[1]

[1]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

http://stanford.edu/group/SOL/multiscale/

Project Goals: We aim to represent metabolic reactions at atomic resolution by saturating metabolic databases with structural formulas for metabolites and atom mappings for reactions. Atom level representations of metabolites and reactions extend the range of applications for metabolic network reconstructions to include, for example, estimation of thermodynamic parameters [1], identification of conserved moieties [2], and stable isotope assisted metabolic flux analysis [3].

Metabolic reactions conserve mass and elements. Each instance of a reaction must therefore map every substrate atom to a specific product atom of the same element. Realisable atom mappings are determined by organic chemistry and reaction mechanisms. Atom mapping data for metabolic reactions open the door to a growing list of applications [3, 4, 5, 2] that are not available with data at the level of reaction stoichiometry. Until recently, acquiring atom mapping data for genome-scale metabolic network reconstructions was a labour intensive prospect. However, a number of algorithms to predict atom mappings have now become available. Here, we compare four recently published algorithms on criteria including accuracy, speed, and availability. The algorithms are DREAM [6], Pathway Tools [7], ICMAP [8] and CLCA [9]. Accuracy was determined by comparison to a set of manually curated atom mappings. We discuss common issues including hydrogen atom mapping and molecular symmetry. We conclude with an effective strategy to increase the coverage of high quality atom mapping data in the metabolic databases.

References

[1] Noor E, Haraldsdóttir HS, Milo R, Fleming RMT (2013) Consistent Estimation of Gibbs Energy Using Component Contributions. PLoS Comput Biol 9: e1003098.

[2] Haraldsdóttir HS, Fleming RMT (2015) Identification of conserved moieties in metabolic networks by graph theoretical analysis of atom transition networks. Submitted .

[3] Wiechert W (2001) 13c Metabolic Flux Analysis. Metabolic Engineering 3: 195–206.

[4] Kotera M, Okuno Y, Hattori M, Goto S, Kanehisa M (2004) Computational Assignment of the EC Numbers for Genomic-Scale Analysis of Enzymatic Reactions. Journal of the American Chemical Society 126: 16487–16498.

[5] Pey J, Planes FJ, Beasley JE (2014) Refining carbon flux paths using atomic trace data. Bioinformatics 30: 975–980.

[6] First EL, Gounaris CE, Floudas CA (2012) Stereochemically Consistent Reaction Mapping and Identification of Multiple Reaction Mechanisms through Integer Linear Optimization. Journal of Chemical Information and Modeling 52: 84–92.

[7] Latendresse M, Malerich JP, Travers M, Karp PD (2012) Accurate Atom-Mapping Computation for Biochemical Reactions. Journal of Chemical Information and Modeling 52: 2970–2982.

[8] Kraut H, Eiblmaier J, Grethe G, Löw P, Matuszczyk H, et al. (2013) Algorithm for reaction classification. Journal of Chemical Information and Modeling 53: 2884–2895.

[9] Kumar A, Maranas CD (2014) CLCA: Maximum Common Molecular Substructure Queries within the MetRxn Database. Journal of Chemical Information and Modeling 54: 3417–3438.

This work was supported by the U.S. Department of Energy, Offices of Advanced Scientific Computing Research and the Biological and Environmental Research as part of the Scientific Discovery Through Advanced Computing program, grant #DE-SC0010429.

Uniform sampling of metabolic networks

Ben Cousins[1], Hulda S. Haraldsdóttir[2]* (hulda.haraldsdottir@uni.lu), Santosh Vempala[1]and **Ronan M. T. Fleming**[2]

[1]School of Computer Science, Algorithms and Randomness Center, Georgia Institute of Technology, Atlanta, Georgia; [2]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

http://stanford.edu/group/SOL/multiscale/

Project Goals: Development of a framework for mass conserved elementary kinetic modelling of metabolic networks [1, 2, 3, 4, 5]. This collaborative project was motivated by a need for data to constrain and validate metabolic models. Sampling algorithms have demonstrated applications in measurement and estimation of kinetic parameters, steady state fluxes and metabolite concentrations for biochemical systems [6, 7, 8].

Constraint-based metabolic modelling provides a framework to explore feasible steady state fluxes in metabolic networks. Physicochemical constraints imposed, e.g., by network topology, mass conservation and substrate availability are formulated as linear equalities and inequalities that define a high-dimensional convex set. Uniform sampling of this set provides an unbiased characterisation of the metabolic capabilities of a cell or organism [9]. However, uniform sampling of steady state metabolic flux sets has proven algorithmically challenging due to their high dimensionality and inherent anisotropy. Here, we evaluate the performance of a recently published sampling algorithm [10] on metabolic networks of increasing size. The algorithm is based upon the provably efficient hit-and-run random walk [11] and crucially uses a rounding preprocessing step to place the set of feasible metabolic fluxes in near-isotropic position. This algorithm converges to a uniform sampling distribution up to 25 times faster than a popular artificial centering hit-and-run algorithm [12]. We demonstrate the effects of improved convergence on predictions of the metabolic capabilities of *E. coli*.

References

[1] Sun Y, Fleming RM, Thiele I, Saunders MA (2013) Robust flux balance analysis of multiscale biochemical reaction networks. BMC Bioinformatics 14: 240.

[2] Noor E, Haraldsdóttir HS, Milo R, Fleming RMT (2013) Consistent Estimation of Gibbs Energy Using Component Contributions. PLoS Comput Biol 9: e1003098.

[3] Artacho FJA, Fleming RMT (2014) Globally convergent algorithms for finding zeros of duplomonotone mappings. Optimization Letters : 1–16.

[4] Artacho FJA, Fleming RMT, Vuong PT (2015) Accelerating the DC algorithm for smooth functions. arXiv:150707375 [math, q-bio] .

[5] Haraldsdóttir HS, Fleming RMT (2015) Identification of conserved moieties in metabolic networks by graph theoretical analysis of atom transition networks. Submitted .

[6] Yang L, Mahadevan R, Cluett W (2010) Designing experiments from noisy metabolomics data to refine constraint-based models. In: American Control Conference (ACC), 2010. pp. 5143–5148. doi:10.1109/ACC.2010.5530678.

[7] Schellenberger J, Zielinski DC, Choi W, Madireddi S, Portnoy V, et al. (2012) Predicting outcomes of steady-state 13c isotope tracing experiments using Monte Carlo sampling. BMC systems biology 6: 9.

[8] Khodayari A, Zomorrodi AR, Liao JC, Maranas CD (2014) A kinetic model of Escherichia coli core metabolism satisfying multiple sets of mutant flux data. Metabolic Engineering 25: 50–62.

[9] Schellenberger J, Palsson BØ (2009) Use of Randomized Sampling for Analysis of Metabolic Networks. Journal of Biological Chemistry 284: 5457–5461.

[10] Cousins B, Vempala S (2015) A practical volume algorithm. Mathematical Programming Computation : 1–28.

[11] Lovasz L, Vempala S (2006) Hit-and-run from a corner. In: SIAM journal on computing. Society for Industrial and Applied Mathematics, volume 35, pp. 985–1005.

[12] Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols 6: 1290–1307.

This work was supported by the U.S. Department of Energy, Offices of Advanced Scientific Computing Research and the Biological and Environmental Research as part of the Scientific Discovery Through Advanced Computing program, grant #DE-SC0010429.

KBase: An Integrated Systems Biology Knowledgebase for Predictive Biological and Environmental Research

Adam P. Arkin[1] (aparkin@lbl.gov), Rick Stevens[2], Bob Cottingham[3], Sergei Maslov[4,5], Fernando Perez[1], and the KBase Team at the following institutions

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Argonne National Laboratory, Argonne, IL; [3]Oak Ridge National Laboratory, Oak Ridge, TN; [4] University of Illinois at Urbana-Champaign; [5] Brookhaven National Laboratory, Upton, NY; [6] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; [7] Hope College, Holland, MI; [8] University of Tennessee, Knoxville, TN.

kbase.us

Project Goals: KBase is a computational platform designed to meet the key challenges of systems biology: predicting and ultimately designing biological function. The project's long-term goal is to help researchers understand how plants and microbes interact with and affect each other and environmental processes, and how these interactions could be harnessed for sustainable bioenergy and environmental solutions. These objectives will be advanced by KBase capabilities for data analysis; model-inspired experimental design; and dissemination of data, results, conclusions, and tools. Within this environment, KBase seeks to encourage biologists to collaboratively integrate, analyze, and interpret heterogeneous data to develop predictive, multiscale models of biological function. This integration of new information and tools from both external reference sources and users will enable researchers to greatly amplify the results of their own work by more effectively leveraging that of others.

The DOE Systems Biology Knowledgebase (KBase) enables secure sharing of data, tools, methods, and conclusions in a unified, extensible system that allows researchers to collaboratively generate and test hypotheses about biological functions; perform large-scale analyses on scalable computing infrastructure; and combine multiple lines of evidence to accurately model plant and microbial physiology and community dynamics.

KBase enables users to upload their own data or access public data in KBase to execute and share customized analyses that target their specific systems biology hypotheses. These computational experiments or analyses are captured in dynamic, interactive documents called Narratives that promote collaboration and reproducibility of scientific results. In addition to data and analysis steps, Narratives can include images, notes, and links, and can be kept private, shared with colleagues and collaborators, or made public for the benefit of the wider research community.

Major classes of KBase analysis tools include microbial genome assembly and annotation, phylogenetics and comparative genomics, expression analysis, and metabolic modeling.

Capabilities for metabolic model reconstruction and flux balance analysis simulation can provide insight into the metabolic pathways and interactions between plants and microbes, enabling researchers to identify, for example, the biochemical reactions active in biomass production. KBase's recently developed RNA-Seq pipeline and downstream analysis tools enable users to quantify expression from RNA-Seq reads, thereby identifying differential expression between different tissues, developmental stages, environmental conditions, and genetic backgrounds.

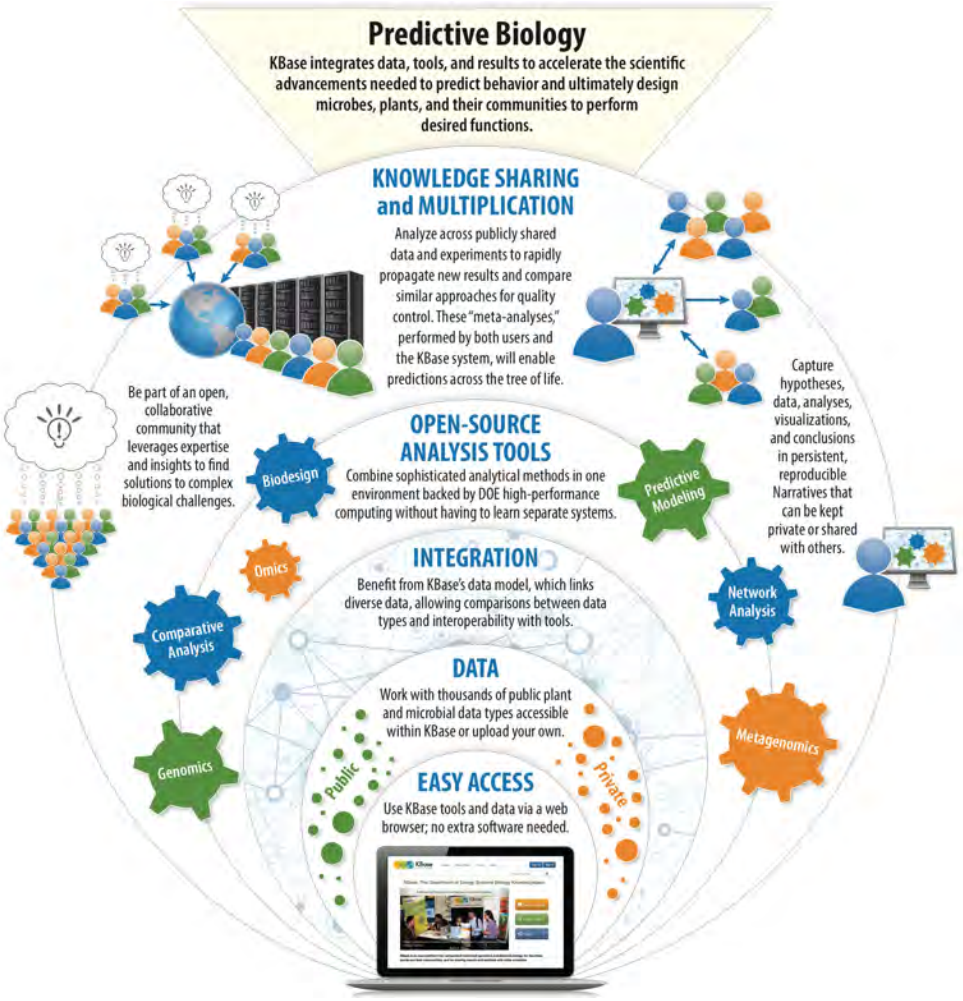


Fig. 1. Schematic overview of the KBase platform.

In addition to these resources, KBase has launched the Software Development Kit (SDK) that enables users to integrate other open source command-line tools and applications into the KBase platform. This new capability will accelerate the incorporation of new science functionality by allowing developers to more easily add—in a standardized and well-tested way—a wide variety of tools that then can be discovered and leveraged by other KBase users.

KBase is funded by the Genomic Science program within the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under award numbers DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886.

Enabling ENIGMA collaborative research in the DOE Systems Biology Knowledgebase

Pavel S. Novichkov[1]*(psnovichkov@lbl.gov), Alexey E. Kazakov[1], Adam P. Arkin[1] and **Paul D. Adams**[1]

[1]*Lawrence Berkeley National Laboratory, Berkeley CA*

http://enigma.lbl.gov

Project Goals: Sharing the data, computational analysis, and hypotheses between members of a large-scale collaborative project like ENIGMA is critical to facilitate collaboration and accelerate the pace of scientific discovery. The goal of this project was to support the ENGIMA collaborative research by enabling necessary data models and tools to start working in the DOE Systems Biology Knowldegebase.

The Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) is a large-scale, multi-institutional collaborative project that involves many teams with different expertise working together to address challenging biological problems. A project like this generates a wide range of experimental and computational types of data that needs to be properly stored and shared between members of the project to facilitate collaboration and accelerate the pace of scientific discovery. It is critical to enable the sharing of not only the raw data, but also the major steps in the data analysis, and ultimately, the generated hypotheses. The DOE Systems Biology Knowledgebase (KBase) is a powerful computational platform that was originally designed to specifically address these needs of the scientific community. Collaboration and sharing, data provenance and data integration, extension by new data types and new data analysis are the key features of the KBase computational infrastructure.

The goal of this project was to enable collaborative research of the ENIGMA Metal Metabolism campaign in the DOE Systems Biology Knowledgebase by implementing necessary data models and computational tools in the KBase infrastructure. The overarching goal of the ENIGMA Metal Metabolism campaign is to elucidate the fundamental mechanisms that drive metal assimilation and investigate metallobiochemistry of microbial communities in the ORNL wells. The primary data types generated by this campaign toward the goal include: (i) survey of the ORNL wells with analysis of different geochemical parameters such as concentration of various metals; (ii) growth assays of the selected isolates under different metal conditions; (iii) HPLC-ICPMS screens to identify metalloproteins involved in metal toxicity and metal reduction. To support these primary data types, we designed and implemented KBase data models representing environmental sample sets, growth data, and chromatography data. The deposition of the data into KBase Narrative is facilitated by the well-documented file formats and developed KBase uploaders for all three data types. A collection of KBase widgets and methods was developed to enable visualization and analysis of the uploaded data. “View Well Samples 2D Plot” and “View Well Sample Histogram” allows for studying the correlation of a particular geochemical parameter across all wells and comparison of different geochemical parameters across a selected number of wells. “View Growth Curves” and “View Growth Parameters Plot” generates plots visualizing growth curves for either all or selected conditions, calculates various growth parameters, such as growth rate or maximum OD, and allows to study its dependence on the growth conditions. “View Growth Parameters 2D Plot” was designed specifically to support

large-scale growth assays to investigate several media parameters, e.g. concentrations of several metals or a combination of concentration of a particular metal and knockout strains. The “View Chromatograms” method allows for basic visualization of chromatography data with zoom in/out features to study particular peaks.

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

Extrachromosomal Plasmid DNA Project

Ankita Kothari[1*], Yu-Wei Wu[1], Marimikel Charrier[1], Lara Rajeev[1], Andrea M. Rocha [2], Terry C. Hazen[2], Paramvir S. Dehal[1], Dylan Chivian[1], Sarah J. Spencer[3], Eric J. Alm[3], Steven Singer[1], Aindrila Mukhopadhyay[1], Adam P. Arkin[1] and **Paul D. Adams**

[1]*Lawrence Berkeley National Laboratory, Berkeley, CA;* [2]University of Tennessee, Knoxville, TN; [3] Massachusetts Institute of Technology, Cambridge, MA.

http://enigma.lbl.gov

Project Goals: The extrachromosomal plasmid DNA project is one of the discovery projects funded by ENIGMA. It is a short term, high impact, investigatory effort to study the plasmid populations of the ENIGMA wells. The goal of this project is to exclusively explore the prevalence of plasmid DNA in these communities. Additionally, the project aims at studying the relevant functional genes that are typically encoded on the plasmids, conferring advantageous traits to the host, in these communities.

Plasmids are autonomously replicating extra-chromosomal genetic elements that often act as mediators of horizontal gene transfer in the environment. Plasmids host and distribute non-essential genes, independent of the host's chromosome, thereby benefitting the host bacteria in certain specific environmental conditions. Native plasmids typically range from 2 kb to 250 kb in size and have been shown to be present in 10-30% of the cultivated isolates from varied environments. The best-characterized wells at the Oakridge site are now documented to contain several hundred bacterial strains, many of which are likely to contain plasmids.

This is the first study to selectively isolate and analyze the plasmid population from these sites. To optimize a robust method that isolates a range of plasmid sizes, we developed a model system comprising of three strains containing plasmids of sizes – 5kb, 48kb and 202kb in equal proportions, and tested the potential of various alkaline hydrolysis based methods to isolate plasmids from the serial dilutions of the mixed population. The presence of each plasmid was determined by targeting a unique plasmid borne gene via qPCR. In order to get rid of genomic DNA contamination, the isolated DNA was subjected to Plasmid-Safe-ATP-Dependent DNase and the lack of genomic DNA contamination was confirmed using degenerate primers targeting the 16s rRNA coding sequence. The total plasmid DNA thus obtained was amplified using Phi29 DNA polymerase to generate high-quality template for use in DNA sequencing. To increase the sensitivity of the plasmid isolation procedure, the extraction procedures and Phi29 amplification conditions were optimized. Subsequently, plasmid DNA was isolated from the wells GW460 and GW456 and subjected to deep sequencing using the Illumina paired-end protocol. The reads obtained were trimmed using Trimmomatics, assembled using IDBA-UD (Iterative De Bruijn graph Assembler for reads with Highly Uneven Sequencing Depth) and subjected to MG-RAST to produce gene calls, functional annotations and taxonomic classification.

A total of 42543 (including 130 circular contigs) and 32313 contigs (including 760 circular contigs) above 2kb size were detected from the wells GW456 and GW460, respectively. These encode several known plasmid associated genes along with genes involved in secondary metabolism, antibiotic resistance, metal resistance, and nitrogen metabolism, to name a few.

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

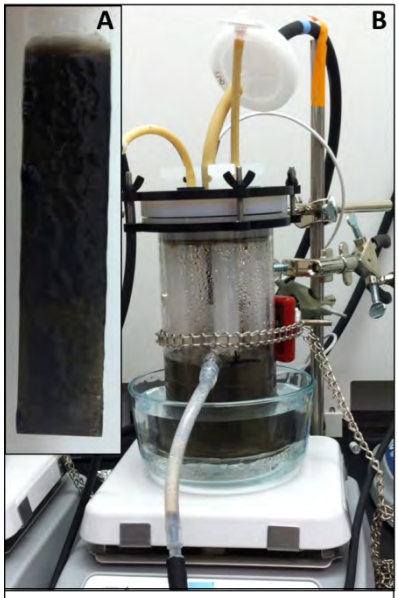
Single-Base Change by Laboratory-Driven Evolution Eliminates Biofilm Formation in *Desulfovibrio vulgaris* Hildenborough

Kara B. De León[1]* (deleonkb@missouri.edu), Grant M. Zane[1], Gregory P. Krantz[2], Matthew W. Fields[2], Judy D. Wall[1], Adam P. Arkin[3,4] and **Paul D. Adams**[3,4]

[1]University of Missouri, Columbia; [2]Montana State University, Bozeman; [3]Lawrence Berkeley National Laboratory, Berkeley, CA; [4]University of California, Berkeley

http://enigma.lbl.gov

Project Goals: The goal of Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) is to advance fundamental knowledge on the impact of microbial communities on ecosystems. Subsurface microbial communities, including those in heavy-metal and radionuclide sites of interest to the DOE, are commonly attached to sediment particles as a biofilm. To understand these communities, we must consider biofilm growth strategy from a genetic context. *Desulfovibrio* are important heavy-metal reducers that contribute to the overall redox of the environment. Our goal is to determine the genetic requirements for biofilm formation in *Desulfovibrio vulgaris* Hildenborough (DvH). In pursuing this goal, an inter-laboratory collaboration through ENIGMA has led to the discovery that two wild-type DvH strains have diverged in biofilm formation ability likely due to laboratory-driven evolution and has revealed a possible mechanism for biofilm formation in DvH.



(A) DvH on glass slide from
(B) CDC Biofilm Reactor

Desulfovibrio vulgaris Hildenborough (DvH) is a sulfate-reducing bacterium present in heavy- metal and radionuclide contaminated sites that is capable of heavy-metal reduction and contributes to the overall redox state of the environment. Though predominantly found attached to sediment particles as a biofilm, the genetic requirements of DvH biofilm formation have not been determined. Our goal is to determine the mechanisms and genetic requirements of biofilm formation in DvH. In pursuing this goal, inter-laboratory collaboration as part of ENIGMA has led to the discovery that two wild-type DvH strains, both originally from ATCC 29579, have diverged in biofilm formation capacity due to laboratory-driven evolution. The wild-type DvH used at the University of Missouri (DvH-MO) is partially deficient in biofilm formation as compared to data published for what should have been the same strain used in Matthew Fields' lab at Montana State University (DvH-MT). The genomes were re-sequenced from

planktonic cultures of DvH-MT and DvH-MO, and DvH-MO steady-state biofilm. In DvH-MO, a single nucleotide polymorphism (SNP) in the ABC transporter (DVU1017) of a type I secretion system (T1SS) has resulted in an Ala635 to Pro change in a conserved α helix near the ATP-binding site. However, after DvH-MO forms a biofilm, a secondary SNP predominates and results in Leu635. We hypothesized that the Ala to Pro change inhibited protein transport by the T1SS and the secondary SNP resulting in a Leu restored transport. These predictions were confirmed by introduction of the SNPs into DvH-MT by site-directed mutagenesis. Therefore, protein transport via the T1SS is required for biofilm formation in DvH and a single nucleotide change due to laboratory-driven evolution is sufficient to stop biofilm formation. Proteins encoded in DVU1012 and DVU1545 both contain T1SS export motifs and are abundant in the DvH biofilm matrix. Inhibited transport of these proteins may have caused biofilm deficiency in DvH-MO. The double deletion mutant is deficient in biofilm formation similar to ΔDVU1017. As either DVU1012 or DVU1545 is sufficient to form biofilm, these biofilm structure proteins likely require DVU1017 for export. These findings have led to a proposed mechanism for biofilm formation in DvH and emphasize the importance of monitoring laboratory-driven evolution, especially between collaborating laboratories.

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

Microbial responses to toxic metals in the Oak Ridge Reservation environment

M.P. Thorgersen[1*](mthorger@uga.edu), B.J. Vaccaro[1], W.A. Lancaster[1], F.L. Poole[1], A.E. Kazakov[2], L. Rajeev[2], M. Garber[2], G.M. Zane[3], M.N. Price[2], K.M. Wetmore[2], A.M. Rocha[4], T. Mehlhorn[4], P.S. Novichkov[2], A.M. Deutschbauer[2], A. Mukhopadhyay[2], J.D. Wall[3], R. Chakraborty[2], T.C. Hazen[4], M.W.W. Adams[1], A.P. Arkin[2], and **P.D. Adams**[2]

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

http://enigma.lbl.gov

Project Goals: The environmentally relevant isolation of several metal-resistant microorganisms is described along with the characterization of a novel uranium-binding complex from a strain isolated from the ORR site.

Abstract: The metal resistance campaign is focused on investigating molecular mechanisms to microbial metal resistance. One of the defining characteristics of the Oak Ridge Reservation (ORR) environment is the presence of mixed industrial waste and the effect this waste has on the groundwater microbial community. Metals are a key component of the mixed waste with concentrations of uranium, aluminum, manganese, cadmium and cobalt in contaminated groundwater wells over 1,000 times greater than those in pristine background wells. The metal resistance campaign is exploring the effects of metal toxicity on the ORR groundwater community in several ways. Strains will be described that were isolated from ORR groundwater in media that contain metal concentrations similar to those present in the contaminated environments. The physiology of metal toxicity was studied through measurement of genome-wide gene fitness under copper, zinc, chromium and uranium toxicity using the model organism *Pseudomonas stutzeri* RCH2. In addition, the properties of a uranium-binding protein complex from the ORR isolate *Pelosinus fermentans* UFO1 will be presented.

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

The Metal Resistome Of An Environmental Bacterium

H. Liu[1]* (hualanliu@lbl.gov), H.K. Carlson[1], M. N. Price[1], A. Deutschbauer[1], A. P. Arkin[1], **P. D. Adams**[2]

[1]Environmental Genomics and Systems Biology, [2]Molecular Biophysics and Integrated Bioimaging; *Lawrence Berkeley National Laboratory, Berkeley, CA, USA.*

http://enigma.lbl.gov

Project Goals: We aim to develop versatile high throughput methods for microbial functional genomics studies. In this project, we will map the metal resistome of an ENIGMA isolate across various metals, and expand this approach to other environmental microorganisms, which will identify common and unique mechanisms bacteria have evolved to deal with elevated metal stress.

Understanding how microbes survive and grow under metal-stressed conditions is a fundamental challenge in environmental microbiology. Though many individual studies have been done in microorganisms with well-known metals including zinc, copper, and chromium, there is little information about global metal-resistance genetic determinants across many different kinds of metals. To address this, we developed a high-throughput culturing system for systematic determination of inhibitory concentrations of metals under aerobic and anaerobic conditions. Using this system, we determined the half maximal inhibitory concentration (IC50) for 24 metals against *Pseudomonas fluorescens* N2E2, an environmental isolate from the Oak Ridge field research site. To globally identify the gene(s) responsible for resistance to these 24 metals, we are using randomly barcoded transposon site sequencing (RB-TnSeq) to identify genes with resistant and hypersensitive phenotypes to elevated metal stress. Here, we present our experimental system and preliminary analysis of metal-resistance determinants in *P. fluorescens* N2E2 with a focus on the similarities and differences in the response of this bacterium to different metals. To facilitate follow-up studies of the functional genomic data, we developed an efficient system to construct in-frame deletion mutants. This genetic system does not just facilitate single gene function studies, but can also be used as a platform for generating double mutants for genetic interaction studies. Taken together, our experimental system will push forward our understanding about the importance of genes as single determinants to metal resistance but also how they interact with each other as part of more complex cellular networks.

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

Atypical Iron Sulfur Cluster Biosynthesis in Sulfate Reducing Bacteria

Valentine Trotter[1], Avneesh Saini[1], Morgan N. Price[1], Jennifer He[1], Jennifer V. Kuehl[1], Kelly M. Wetmore[1], Nancy L. Liu[1], Grant M. Zane[2], Samuel R. Fels[2], Thomas R. Juba[2], Maxim Shatsky[1], John-Marc Chandonia[1], Judy D. Wall[2], Adam M. Deutschbauer[1], Gareth P. Butland[1*](GPButland@lbl.gov), Adam P. Arkin[1] and **Paul D. Adams[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2] University of Missouri, Columbia, MO

http://enigma.lbl.gov

Project Goals: One of the key aims of Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) is to obtain insight into the functionally important activities of key microbes and their interactions in the environment. ENIGMA investigators are specifically interested in how microbes respond to environmental changes and have developed techniques to enable evidence-based annotation of gene function using high throughput mutagenesis and phenotyping of mutants under a range of stress conditions. Here, using the model sulfate-reducing bacterium *Desulfovibrio vulgaris,* our goal was to utilize a combination of traditional biochemistry and molecular biology together with a high throughput functional genomics approach to screen for genetic interactions. Together these data are permitting us to gain detailed insight into the key essential process of iron sulfur cluster biosynthesis, which is known to be impacted by environmentally encountered stresses.

Iron sulfur (FeS) cluster containing proteins make essential contributions to many key cellular processes but can readily be damaged by environmentally encountered oxidative stress, exposure to toxic metals or reactive N-oxyanions. Damage of key FeS enzymes such as pyruvate-ferredoxin oxidoreductase (PFOR) has been proposed as the source of inviability of sulfate reducing bacteria (SRBs) and other anaerobes when encountering oxygen. Using *Desulfovibrio vulgaris* Hildenborough (DvH) as a model SRB, we have characterized DvH strains harboring mutations in known FeS cluster biosynthesis factors and assayed the relative contributions of these factors to FeS cluster biosynthesis by monitoring activity of FeS dependent enzymes including PFOR. To date, enzymatic assays suggests that mutations in DVU1021(*sufB*), DVU1382 (*sufA*) and DVU0664 (*nifS*) have a detrimental effect on multiple FeS enzymes, although functional redundancy between biosynthesis factors appears to be significant. Only DvH strains lacking DVU1021 or DVU1382 displayed severe growth defects in rich growth media, consistent with the hypothesis that the SufBC predicted FeS scaffold complex is the likely primary site of *de novo* FeS cluster assembly in DvH and may work with DVU1382 to mature multiple FeS enzymes, deficiencies in which can impact the growth rate of the cell. In order to fully uncover the relationships between DvH FeS cluster biosynthesis factors, we have utilized a TnSeq-based procedure (RB-TnSeq; [1]) to conduct high throughput genetic interaction screening. We have observed synthetic lethality between SufBC and NifSU systems (no *sufB*/*C* transposon insertions in *nifS*/*U* background or *v.v.*) which is consistent with partial

functional redundancy. Strikingly, the observed viability of a *nifS* mutant in DvH where it is predicted to be the sole cysteine desulfurase, contradicts current FeS cluster biosynthesis dogma regarding the essentiality of cysteine desulfurases as a sulfur source for *de novo* FeS cluster formation. Further work is ongoing to confirm *nifS* genetic interactions.

References

1. Wetmore, K.M., et al., *Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons.* MBio, 2015. **6**(3): p. e00306-15.

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

The Properties of and Microbial Interactions with Natural Organic Matter Extracted from Oak Ridge FRC

Romy Chakraborty[1]* (rchakraborty@lbl.gov), Xiaoqin Wu[1], Sindhu Jagadamma[2], Terry C. Hazen[2], Nicholas Justice[1], Stefan Jenkins[1], Trent R. Northen[1], Matthew W. Fields[3], Patricia Fox[1], Peter Nico[1], Adam P. Arkin[1] and **Paul D. Adams**[1]

[1]*Lawrence Berkeley National Laboratory, Berkeley CA;* [2]Oak Ridge National Laboratory, Oak Ridge, TN & University of Tennessee, Knoxville, TN; [3]Montana State University, Bozeman, MT

Project Goal: Natural organic matter (NOM) availability and transformations determines much of the carbon (C) flux in subsurface environments. However, the molecular signature of this pool of C is largely unknown, and the microbial activities that regulate NOM turnover are still poorly resolved. The goal of this discovery project was to ascertain physical and chemical characteristics of NOM, using minimally destructive techniques that retains the originality of the material, and to use it as a C source in enrichments to study it's turnover by microbial communities at Oak Ridge FRC. Discovery projects are short term, high impact, investigatory efforts to drive changes in science or technological capability that deeply impact the program in some way.

NOM was obtained by extracting FW305 and FW306 sediment samples from different depths using MilliQ-water via shaking and sonication. Using this extraction, total organic C in NOM was 0.4% in the surficial layers and dropped to 0.15% in the deeper layers. The amount of inorganic C in extracted NOM decreased significantly with depth. Results from UV and FTIR analyses showed that extracted NOM mainly contained aromatic and unsaturated compounds in shallower depths and mostly polysaccharides in deeper sediments. HPSEC was used to study the molecular weight distribution of the NOM and a stirred ultrafiltration cell was used to fractionate the extracted NOM based on molecular weight. Extracted NOM was provided as the sole C source to microbes present in background well waters of Oak Ridge FRC. Analysis of the enriched microbial community, and transformed NOM metabolites was carried out. Several fine-scale chemical techniques including FTIR, LC-TOF-MS, and Orbitrap were used to characterize the metabolites, and 16S rRNA sequencing and metatranscriptomics were used to identify the changes in microbial communities in these enrichments.

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

Predictions of Microbial Coexistence on Limited Resources

Onur Erbilgin[1,2], Benjamin P. Bowen[1,2], Trent R. Northen[1,2], Adam P. Arkin[1,3], **Paul D. Adams[1,3]**

[1]*Lawrence Berkeley National Laboratory, Berkeley, CA;* [2]Joint Genome Institute, Walnut Creek CA; [3]University of California, Berkeley, CA.

http://enigma.lbl.gov

Project Goals: To develop a set of ecological and evolutionary principles that explains and can predict the distribution of metabolic networks of microbial communities. In a 'bottom-up' strategy, Metabolomics will be used to determine resource usage patterns of diverse environmental microbial isolates, and these patterns will be used to predict and test mixed community compositions to develop simple rules for biochemical ecology. These principles can then be used to design or modify microbial assemblages.

Microorganisms were the first living beings on this planet, rising from an incredibly dynamic environment. Over the past 3.5 billion years, the environments in which these microbes live have changed dramatically and multicellular life has evolved, creating countless new environments that microbes have colonized, forming interdependent microbial communities. Throughout this time, microbes have developed complex and highly regulated metabolic systems to efficiently use resources. However, until recently, microbes have generally been studied as individuals and little is known what role(s) their metabolic systems have in microbial consortia. In order to fill this knowledge gap, we are determining the resource usage kinetics of microbial isolates of every energy source in defined media.

We use this information to predict how the isolates might behave in mixed communities and how to alter the environment to achieve a desired outcome (e.g. causing one species to dominate the culture). We show quantitative differences between usage of resources in three environmental isolates grown on a defined medium of amino acids and glucose, and discuss how these differences affect the outcomes of mixed community experiments. In addition to targeted metabolomics, we have the capability of conducting high-throughput untargeted metabolomics experiments. The targeted mass spectrometry used to analyze the defined media and spent media informs parameters in the Metabolite Atlas platform, including retention time, relative ionization efficiencies, and fragmentation patterns. Metabolite Atlas users can search their raw experimental data (mass spectra) in the context of this information to more confidentially identify and characterize the compounds detected.

This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (http://enigma.lbl.gov), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S.

Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231

Population filtering in sediment biofilms from dynamic, source planktonic communities

A. Zelaya*[1,2](anna.zelaya@montana.edu), K.L. Bailey[3], P. Zhang[4], S.P. Preheim[5], J. Van Nostrand[4], D.A. Elias[3,7], E.J. Alm[5,7], J. Zhou[4,7], M.W. Fields[1,2,7], A.P. Arkin[6,7] and **P.D. Adams**[6,7]

[1]*Center for Biofilm Engineering, Bozeman, MT;* [2]*Department of Microbiology & Immunology, Montana State University, Bozeman, MT;* [3]*Division of Environmental Sciences, Oak Ridge National Laboratory, Oak Ridge, TN;* [4]*Institute for Environmental Genomics, University of Oklahoma, Norman, OK;*
[5]*Department of Biological Engineering, MIT, Boston, MA,*
[6]*Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley,* CA*;*
[7]*ENIGMA (http://enigma.lbl.gov/)*

Project Goals: The goal of Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) is to push the frontier of knowledge on the impact of microbial communities on ecosystems and to elucidate microbial assembly rules. As part of this, we must link genetic capacities to ecological function. In these sites, the microbial communities are often predominantly attached to sediment particles as a biofilm. We have characterized temporal dynamics of bacterial groundwater communities and compared to the establishment of particle-associated bacterial communities on native sediments incubated down-well. In pursuing this goal, inter-laboratory collaboration facilitated by ENIGMA has led to the identification of aquifer population distributions between the liquid/solid boundaries *in situ*.

Understanding the factors that determine microbial assembly, composition, and function in subsurface environments are critical to assessing contributions to biogeochemical processes such as carbon cycling and bioremediation. However, these factors are still not fully understood. In this study, surrogate sediment samples were incubated for 3 months in 3 wells (FW-301, FW-303, FW-305) within the background site of the Oak Ridge Field Research Center in Oak Ridge, TN. Local sediment biofilm communities were compared to those of the groundwater (source diversity). Groundwater samples from each well were collected approximately 3 times a week in order to identify members of planktonic communities. Field well geochemistry was also measured. Multiple sediment samples (n=12) were used per well to determine inter- and intra-well variation. Geochemical measurements reveal that organic acids and anions remained relatively stable in all wells over the studied time period. Wells had similar median values for oxidation-reduction potential but not for conductivity, pH, and dissolved oxygen. Additionally, the temporal stability of conductivity, pH, and dissolved oxygen was different for each well, with FW-303 being the most stable and FW-301 being the most dynamic. Community analysis of local and source diversity via ss-rRNA paired-end sequencing and distribution-based clustering revealed higher richness, diversity, and variability in source groundwater communities compared to sediment-associated communities. Groundwater sequences displayed transitory predominance unique to each

well over time, and were characterized by periodic blooms of different populations (*e.g. Curvibacter*, *Dechloromonas* in FW-301, *Anoxybacillus*, *Methylophilus* in FW-305). In sediment samples, 20-40% of the communities consisted of populations that were abundant at less than 5% of the total sampled diversity. Inter-well sediment biofilms (across wells) were also distinct from each other. Intra-well sediment biofilms showed much less variability, although sediment from FW-305 was the most variable. For all wells, some populations observed at low abundances in planktonic communities became predominant in the biofilm (*e.g. Aquabacterium*, *Pseudomonas*). However, these populations were represented by unique sets of OTUs in each well. Ordination analysis revealed sediment biofilm communities were distinct from corresponding groundwater communities, with the exception of FW-305 (the youngest well), which showed greater similarity between within-well planktonic and sediment communities. Intra-well population networks (SPARCC) with strong associations (correlation cutoff $\geq \pm 0.8$) revealed that FW301 had the greatest number of positive associations between sediment populations, as well as the most negative associations between groundwater and sediment populations. Many more correlations were observed in biofilms than corresponding groundwater populations, and may have been a consequence of temporal changes. Conversely, FW-305 had fewer associations overall, and none that were negative. These results indicate a shift in local community structure that is influenced by the available source community as well as hydrology.

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

Assays for spatial structure and transdomain dynamics in environmental communities

S.J. Spencer[1]* (sjspence@mit.edu), M.V. Tamminen[2], A.M. Rocha[3], T.C. Hazen[3], E.J. Alm[1]; A.P. Arkin[4,5]; and **P.D. Adams[4,5]**

[1]Massachusetts Institute of Technology, Cambridge; [2]ETH Zürich, Switzerland; [3]Oak Ridge National Laboratory, TN; [4]University of California, Berkeley; [5]*Lawrence Berkeley National Laboratory, Berkeley CA*

Project Goals: Our group focuses on biotechnology development that moves the field of microbial ecology toward complete genomic awareness. Specifically, we use bulk water-in-oil emulsion droplets combined with tailored molecular biology to provide a more comprehensive microscale view of transdomain ecological players and their functional capacity within complex environments.

The bulk activity of microbial communities is composed of the additive effect of microscale interactions between bacteria, viruses, and eukaryotes coexisting within a dynamic environment. These microscale competitive or mutualistic exchanges bridge between the foundational principles of ecology and the global activity of microbial communities that we observe in bulk assays. Before we are able to understand, model, or perturb systems at the macroscopic scale, we need improved methods at the resolution of individual cells. Our group recently developed an emulsion-based droplet assay termed epicPCR (emulsion, paired isolation, and concatenation PCR) to physically link functional genes with phylogenetic indicators within single cells. Here we expand upon this platform to map the physical associations of bacteria with each other and with eukaryotic hosts.

We are beginning to assay biofilm and host-prey structures by capturing small aggregates of cells in nanoliter droplets, then physically linking segments of the 16S rRNA gene between cells. In biofilms a preliminary untargeted assay, trying to link every cell with every other adjacent cell, highly favored only the most abundant strains present. We've now redesigned primer sets that anneal to specific phyla of interest and their physical partners, a semi-targeted version of the assay. With this approach we've recovered library contructs enriched for cells as rare as 1 in 10,000 within complex biofilms. In parallel to assays of bacterial proximity, we've refined the same droplet methodology to capture eukaryotes with their bacterial symbionts and prey. Sequenced co-aggregations between eukaryotes and bacteria in both wastewater and lake water are enriched for predators, heterotrophs, and known symbionts. We plan to apply this approach to Oak Ridge FRC samples, and in preparation we've completed standard eukaryotic and bacterial sequencing from control wells. These structural assays in combination will provide novel microscale data about the complex interchanges connecting bacteria with each other and their broader ecological context.

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

Temporal Variation in Groundwater Microbial Community Structure: Implications for Groundwater Monitoring

Andrea M. Rocha[1, 2*] (rochaam@ornl.gov), B. Adams[1,2], C. Paradis[1,2], Tonia L. Mehlhorn[1], Jennifer E. Earles[1], Kenneth A. Lowe[1], Dawn M. Klingeman[1], David B. Watson[1], Dominique C. Joyner[1, 2], Sindhu Jagadamma[1,2], Julian L. Fortney[1, 2], Jizhong J. Zhou[3], Joy D. Van Nostrand[3], Michael W. W. Adams[4], Romy Chakraborty[5], Dwayne Elias[1], Eric J. Alm[6], and Terry C. Hazen[1, 2], Adam P. Arkin[5], and **Paul D. Adams[5]**.

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

http://enigma.lbl.gov

Project goals: The overarching goal of the Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) field microbiology component is to understand the interactions between environmentally relevant microbial communities and their environment. As part of this, we must understand the mechanisms that may potentially impact microbial community structure, function, and activity. At the DOE Oak Ridge field site, groundwater wells are subject to temporal and seasonal changes in groundwater hydrology and geochemistry, which may result in temporal bias in microbial community monitoring data. Here, we characterize the microbial community response to temporal variability to determine to what extent variation in groundwater geochemistry and hydrology impacts microbial community structure and function.

Large-scale groundwater sampling events, such as those associated with biomonitoring, may span the course of several weeks or months. Based on the sampling methods and time-scales involved, microbial community and geochemical data may contain significant temporal bias, as well as, biases attributed to external factors. In this study, we characterized the temporal dynamics of microbial groundwater communities at the background site of the Oak Ridge Field Research Center in Oak Ridge, TN to determine (1) if, and to what extent, temporal and/or seasonal variation of the groundwater geochemistry affects the microbial community structure and (2) to assess the impact of groundwater flow and transport of geochemical constituents on the microbial structure. To determine how resilient microbial communities are to daily and weekly changes in groundwater chemistry, the temporal dynamics of microbial communities from six groundwater wells were monitored and compared to geochemical and hydrological measurements. Of the six wells, we physically and chemically cleaned four wells to remove biofilm and attached particulates from the well casing. The remaining two wells that weren't cleaned served as controls. All wells were sampled prior to cleaning to establish a baseline microbial community profile. Post-cleaning, each well was sampled a total of twelve times from December 1, 2014 – January 12, 2015. For each well and time-point, groundwater samples were collected for geochemical and microbial community analyses using low flow sampling methods. Nucleic acids were collected by sequentially filtering water through a 10μm pre-filter and 0.2μm-membrane filter and then extracted using a Modified Miller method. A total of 6,959

OTUs were identified across all six wells and size fractions. Of the OTUs Proteobacteria represented a significant portion of the taxa. Analyses of microbial community data indicate overall diversity of the taxa did not vary significantly during time-course sampling. However, significant shifts in the population were observed between cleaning treatments and during sampling time point 01/5/15 for select wells. Additionally, daily and weekly variation in the relative OTU abundance within each well was detected. Throughout the study, groundwater geochemical measurements were relatively stable. However, shallow wells varied in concentrations of Na^{+}, K^{+}, Ca^{2+}, HCO_3^{-}, and CO_3^{-2} following rain events. The geochemical values for these ions are consistent with distinct differences in water types between deep and shallow wells. The stability of the geochemical measurements may indicate that groundwater chemistry is not a dominant factor in the observed daily and weekly variances, but rather contributes to taxonomic differences observed between well depths. However, further analysis of geochemical shifts at higher resolutions is necessary to understand the full impact of geochemistry on microbial response. Furthermore, analyses of groundwater results indicate that the pumping/sampling of wells did not contribute to sampling bias. Overall, results demonstrate that groundwater microbial community data contain temporal biases. As such caution must be used when designing large-scale sampling events.

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

The Phenotypic Landscape Of Bacteria

A.M. Deutschbauer[1]* (AMDeutschbauer@lbl.gov), M.N. Price[1], K.M. Wetmore[1], R.J. Waters[2], Y. Suh[1], M. Callaghan[1], J. Ray[1], R.A. Melnyk[1], J.V. Kuehl[1], J.S. Lamson[1], Z. Esquivel[1], M.J. Blow[2], J. Bristow[2], A.P. Arkin[1], **P.D. Adams**[3]

[1]Environmental Genomics and Systems Biology, [2]Joint Genome Institute, [3]Molecular Biophysics and Integrated Bioimaging,; *Lawrence Berkeley National Laboratory, Berkeley, CA, USA.*

http://enigma.lbl.gov

Project Goals: The Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) program broadly seeks to understand the interactions between environmentally relevant microorganisms and their environment. One aim of this large interdisciplinary project is to rapidly bring environmental bacteria to model-organism by systematically annotating the functions of poorly characterized genes. Here, we describe a high-throughput genetic platform for generating millions of gene-phenotype measurements and how these data can be used to fill the sequence-to-function gap.

Current gene function annotation pipelines fail to identify a role for 40% genes in the typical bacterial genome. To explore the functions of these uncharacterized genes, we generated genome-wide transposon mutant populations from 25 bacteria across four bacterial classes, including 8 strains isolated from the Oak Ridge field research site. Growth of mutant populations across ~4,000 experiments, followed by random barcode transposon sequencing (RB-TnSeq), enabled over 13 million individual measurements of gene fitness, and revealed significant phenotypes for 7,375 previously uncharacterized genes. Among these genes, 4,690 had phenotypes under a specific experimental condition, or shared phenotypic patterns with a gene of known function, thus enabling specific predictions of gene function. Finally, since 11% of hypothetical genes across all sequenced genomes have at least one ortholog with a phenotype in our dataset, we demonstrate the ability of high-throughput genetics to identify roles for many of the uncharacterized proteins in bacterial genomes.

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

Engineering Robust Hosts for Microbial Biofuel Production

Tim Tomko[1], Yik Siu, Jessica Lindle[1], Will Turner[1], **Mary J. Dunlop** (mjdunlop@uvm.edu)[1,*]

[1] University of Vermont, School of Engineering, Burlington, VT 05405

Project Goals: **The overall goal of this project is to enhance microbial synthesis of next-generation biofuels by developing tools for improving microbial tolerance of biofuel production conditions. Research is organized around three objectives: (1) Identify novel biofuel tolerance mechanisms from microorganisms that naturally thrive in hydrocarbon-rich environments. (2) Engineer a synthetic feedback loop that responds to biofuel production. To optimize biofuel production yields, cells must balance several competing sources of stress. We are designing and constructing a novel feedback loop that senses biofuel production and turns on export pumps in response. (3) Integrate multiple tolerance strategies in a biofuel production strain. In addition to having the potential to greatly enhance biofuel yields, this work advances understanding of how multiple tolerance mechanisms interact within a cell.**

Abstract:
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. Studies have shown that strains engineered to increase tolerance can improve biofuel production yields.

Microorganisms that survive in oil-rich environments are a valuable source of tolerance mechanisms. Using genomic DNA from the hydrocarbon-degrading microbe *Marinobacter aquaeolei*, we constructed a transgenic library that we expressed in *Escherichia coli*. We exposed cells to inhibitory levels of pinene, a monoterpene that can serve as a jet fuel precursor with chemical properties similar to existing tactical fuels. Using a sequential strategy with a fosmid library followed by a plasmid library, we were able to isolate a region of DNA from the *M. aquaeolei* genome that conferred pinene tolerance when expressed in *E. coli*. We determined that a single gene, *yceI*, was responsible for the tolerance improvements. Overexpression of this gene placed no additional burden on the host. We also tested tolerance to other monoterpenes and showed that *yceI* selectively improves tolerance. The genomes of hydrocarbon-tolerant microbes represent a rich resource for tolerance engineering. Using a transgenic library, we were able to identify a single gene that improves *E. coli*'s tolerance to the bio-jet fuel precursor pinene.

In addition to identifying novel tolerance mechanics, we are designing control systems for efflux pumps known to export biofuel. Pump overexpression inhibits cell growth, suggesting a trade-off between biofuel and pump toxicity. To counter this, we are using the protein MexR, native to *Pseudomonas aeruginosa*, as a biosensor because it detects oxidative stress such as that caused by the introduction of biofuels. In the feedback loop design, MexR represses the expression of an efflux pump derived from *M. aquaeolei* by binding to a synthetic promoter

region. We developed a library of synthetic promoters, which vary the number and location of MexR binding sites, and screened these for tolerance to pinene, a known pump substrate. The screen tests both constant and dynamic biofuel environments. The dynamic environment is important for selecting a sensor that performs well in both the presence and absence of biofuel. Our experimental findings are further supported by a mathematical model describing the dynamic sensor selection. By applying dynamic inputs to the selection, we show that it is possible to select for traits that satisfy multiple goals (such as performing well in the presence and absence of biofuel). Furthermore, we demonstrate that the underlying diversity in a library is heavily influenced by the initial circuit design. Overall, our findings argue that rational synthetic circuit design, coupled with diversity generation and dynamic selection are powerful tools for many synthetic biology applications, including biofuel production.

We have also studied whether expressing multiple pumps in combination could further increase biofuel tolerance. With multiple pumps, the combined impact of pump toxicity and benefits from increased tolerance are unclear. To address this, we measured tolerance of *E. coli* to pinene in one-pump and two-pump strains. To support our experiments, we developed a mathematical model describing toxicity due to biofuel and overexpression of pumps. We found that data from one-pump strains can accurately predict the performance of two-pump strains. This result suggests that it may be possible to dramatically reduce the number of experiments required for characterizing the effects of combined biofuel tolerance mechanisms.

Publications

1. J. M. Lindle, M. J. Dunlop. "Performing Selections Under Dynamic Conditions for Synthetic Biology Applications." *Integrative Biology*, DOI: 10.1039/c5ib00286a, 2016.
2. T. Tomko, M. J. Dunlop. "Engineering Improved Bio-jet Fuel Tolerance in Escherichia coli using a Transgenic Library from the Hydrocarbon-degrader Marinobacter aquaeolei." *Biotechnology for Biofuels*. 8:165, 2015.
3. W. J. Turner, M. J. Dunlop. "Trade-offs in Improving Biofuel Tolerance Using Combinations of Efflux Pumps." *ACS Synthetic Biology*, DOI: 10.1021/sb500307w, 2014.
4. M. Frederix, K. Hutter, J. Leu, T. S. Batth, W. J. Turner, T. L. Ruegg, H. Blanch, B. A. Simmons, P. D. Adams, J. D. Keasling, M. P. Thelen, M. J. Dunlop, C. J. Petzold, A. Mukhopadhyay. "Development of a native Escherichia coli induction system for ionic liquid tolerance." *PLoS One*, 2014.
5. M. E. Harrison, M. J. Dunlop. "Synthetic Feedback Loop Model for Increasing Microbial Biofuel Production Using a Biosensor." *Frontiers in Microbiology* 3:360, 2012.
6. M. J. Dunlop. "Engineering Microbes for Tolerance to Next-Generation Biofuels." *Biotechnology for Biofuels* 4:32, 2011.

This research was supported by the University of Vermont and the Office of Science (Biological and Environmental Research), US Department of Energy.

Systems biology towards a continuous platform for biofuels production

Adam J. E. Freedman[1], Jason T. Boock[2], Geoffrey Tompsett[3], **Kristala L. J. Prather[2], Michael T. Timko[3]** and **Janelle R. Thompson*[1] (jthompson@mit.edu)**

[1]Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge MA; [2]Chemical Engineering, Massachusetts Institute of Technology, Cambridge MA; [3]Chemical Engineering, Worcester Polytechnic Institute, Worcester, MA

Project Goals: Environmental strain isolation provides an opportunity to discover organisms with unique growth characteristics and physiological traits to overcome bioprocess challenges including end product toxicity, culture contamination and low temperature growth. We hypothesize that an environmental isolate of *B. megaterium* SR7, which we recovered from a deep subsurface supercritical carbon dioxide ($scCO_2$) well, will provide bioprocess advantages when cultured in a dual-phase reactor of growth media and $scCO_2$. We are working towards the following goals (1) Develop *B. megaterium* into a bioproduction host for biofuels (2) Engineer de novo pathways for biosynthesis of longer chain fuels in *B. megaterium* and (3) Develop and model a two-phase stripping chemostat for continuous biosynthesis and in situ extraction of biofuels.

Abstract: To develop strain SR7 for use in dual-phase bioreactors we have carried out genome sequencing and functional annotation, physiological growth characterization, and established a toolset for genetic modification. After sequencing *B. megaterium* SR7's 5.45 Mbp genome, natural metabolite profiles affirmed the strain's use of glycolysis and the TCA cycle for anaerobic energy generation, as expected by functional genomic annotations. Process improvements, including optimized minimal media formulation and altered mixing regimes, established consistent growth at 1 atm CO_2 as a higher throughput model system for $scCO_2$ conditions. Based on findings from 1 atm CO_2 cultures, we amended $scCO_2$ media obtaining improved growth under $scCO_2$.

We developed a genetic system for strain SR7 using a protoplast-based transformation protocol employing a *B. megaterium* compatible plasmid that is maintained in SR7 for at least 100 hours in both aerobic and anaerobic cultures. We also identified xylose- and IPTG-inducible promoters capable of 10- to 250-fold inducible heterologous protein expression under aerobic and 1 atm CO_2 conditions; the specific IPTG promoter has not been demonstrated in *B. megaterium* previously. We engineered SR7 to produce isobutanol by introducing a two-enzyme (2-ketoisovalerate decarboxylase (KivD) and alcohol dehydrogenase (Adh)) pathway and feeding 2-ketoisovalerate. The two-step conversion occurs at approximately 70-80% substrate conversion in aerobically and 1 atm CO_2 grown cells. However, the intermediate aldehyde was found to accumulate at short culture times. Due to the high partition coefficient for the aldehyde to the $scCO_2$ phase, five alternative alcohol dehydrogenases were tested. A variant was found that lowered the build-up of isobutyraldehyde and resulted in conversion of 2-ketoisovalerate to isobutanol above 85%.

In addition to systems work on SR7, we designed and constructed a bench-scale, continuous flow, pressurized fermentation system centered on a 300-mL high-pressure

chemostat that is equipped with specialized features for monitoring cell growth by optical density and measuring pH. Recent abiotic experimental modeling of butanol extraction by $scCO_2$ has enabled us to obtain mass transfer coefficients that will inform our chemostat design in order to minimize build-up of toxic products. We have begun process level energy balance calculations, aiming to identify conditions that optimize the energy return of fuel products relative to input energy costs.

This work was supported by the Office of Biological and Environmental Research in the DOE Office of Science (DE-SC0012555).

NanoSIMS Isotope Imaging and Genome-enabled Metabolic Modeling to Investigate Algal-Bacterial Interactions in Biofuel-Producing Communities

Xavier Mayali[1] (mayali1@llnl.gov), Ty Samo[1], Daniel Nilson[1], Anthony Siccardi[2], Todd Lane[3], Jeffrey Kimbrel[1], Patrik D'haeseleer[1], Marc Griesemer[1], **Ali Navid**[1], Carol Zhou[1], **Peter K. Weber**[1], **Jennifer Pett-Ridge**[1]

[1]Lawrence Livermore National Laboratory, Livermore CA; [2]Texas A&M Corpus Christi, [3]Sandia National Laboratory, Livemore CA

Project Goals: The LLNL Biofuels SFA seeks to support robust and sustainable microalgae fuel production through a systems biology understanding of algal-bacterial interactions. We hypothesize that by understanding the factors that control cellular physiology and biogeochemical fluxes in and out of algal cells, particularly through the phycosphere, we can advance the efficiency and reliability of algal biofuel production. Our research includes studies of beneficial traits of phycosphere-associated bacteria, systems biology studies of model algae, and genome-enabled metabolic modeling to predict the interspecies exchanges that promote algal growth, lipid production and healthy co-cultures. Our overall goal is to develop a comprehensive understanding of complex microbial communities needed to advance the use of biological properties for practical energy production.

To better understand how bacteria promote the growth of microalgae, we are investigating algal-bacterial interactions in batch laboratory cultures of two algal species (*Nannochloropsis salina* CCMP 1776 and *Phaeodactylum tricornutum* CCMP 2561). Outdoor algal pond samples from Corpus Christi, Texas, were obtained and the free-living bacterial fraction (< 1 micron) was added to previously bacteria-free (axenic) cultures to create primary enrichments. These samples were further enriched for algal cell surface (phycosphere) associated bacteria by collecting the algal cells, washing and creating secondary enrichments via dilution cultures. Multiple enrichments were investigated for the ability of attached bacteria to influence the cell-specific carbon fixation rates of the microalgal cells using NanoSIMS and isotope probing (NanoSIP[1]). Cultures were incubated with ^{13}C bicarbonate to track inorganic C fixation and ^{15}N-leucine to track bacterial production and then analyzed with LLNL's NanoSIMS 50 to quantify individual cell's isotope incorporation. Most primary bacterial enrichments did not have significant effects on microalgal cell specific C fixation rates, and attached bacteria generally exhibited faster growth than free-living bacteria. However, the majority of secondary enrichments exhibited the opposite pattern: algal cells with attached bacteria had increased cell specific C fixation rates, and free-living bacteria had increased growth rates compared to phycosphere-attached bacteria. Current and future efforts will include the isolation of single phycosphere-associated bacterial species with probiotic effects, and whole genome sequencing to enable metabolic modeling and an understanding of the genetic basis for observed probiotic effects.

Algal-bacterial mutualism involves the secretion and exchange of nutrients and vitamins, as well as signaling molecules, into the phycosphere. The ability of an individual microalgal cell or bacterium to utilize these extracellular chemicals can be computationally modeled to predict optimal conditions for incorporation of these

compounds into increased cellular biomass or bioproducts. High-quality modeling, however, requires comprehensive and accurate genome annotation, which depends on high-quality sequencing and genome assemblies, in addition to highly curated and validated functional annotation databases. Our analysis pipeline combines high-quality draft genomes from JGI, with our in house functional annotation pipeline to map enzymes to metabolic pathways. The functional annotation includes JGI IMG and RAST annotations supplemented with the LLNL PSAT tools[2]. In our current work, we are comparing and integrating multiple functional annotation and metabolic reconstruction tools to increase the number of reactions and metabolites predicted from a genome. In future work, curated models will be used to validate isotope exchange determined by NanoSIMS, as well as to rapidly identify potential new routes of nutrient exchanges for algal-bacterial interactions to be experimentally tested.

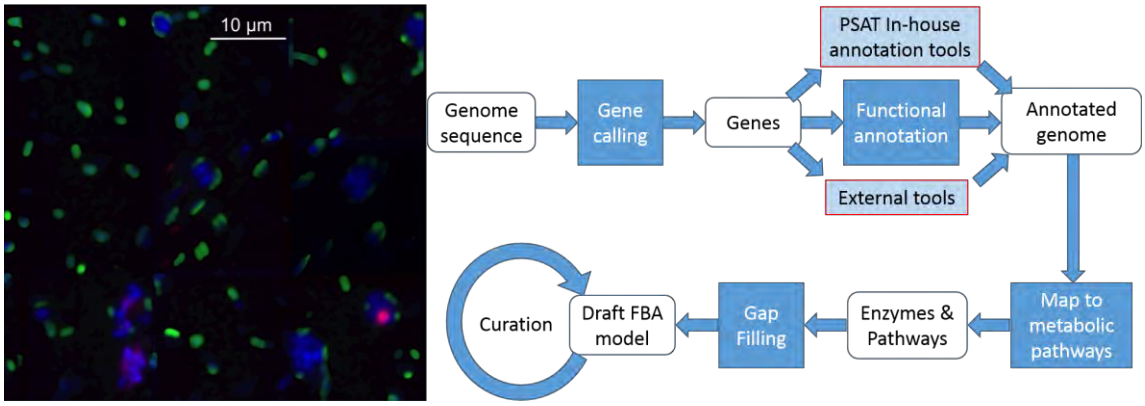


Figure 1. Left: False color stitched NanoSIMS images of *N. salina* secondary enrichment culture: green = heterotrophic ^{15}N leucine incorporation, red = autotrophic ^{13}C carbon fixation, blue = biomass; Right: current pipeline for functional annotation to obtain flux balance analysis models from draft genomes.

References

1 Pett-Ridge, J. & Weber, P. K. in *Microbial Systems Biology: Methods and Protocols* (ed Ali Navid) 375-408 (Humana Press, 2012).

2 Leung E, Huang A, Cadag E, Montana A, Soliman JL, Zhou CLE (2016) Protein Sequence Annotation Tool (PSAT): a centralized web-based meta-server for high-throughput sequence annotations. BMC Bioinformatics. **DOI:** 10.1186/s12859-016-0887-y

This research was supported by the LLNL Biofuels Scientific Focus Area, funded by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under FWP SCW1039. Work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344.

Linking microbial identity and function in phototrophic mats and biofilms

Rhona K. Stuart[1*] (stuart25@llnl.gov), Mary Lipton[2], Amy Boaro[2], Craig Everroad[3], Leslie Bebout[3], Angela Detweiler[3], **Peter Weber**[1], **Brad Bebout**[3], **Michael Thelen**[1], **Jennifer Pett-Ridge**[1]

[1]Lawrence Livermore National Laboratory, Livermore, CA, [2]Pacific Northwest National Laboratory, Richland, WA, [3]NASA Ames Research Center, Mountain View, CA

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

Cyanobacterial communities that form laminated mats are highly diverse microbial assemblages, and have been studied for decades as analogs for early earth life. Yet the partitioning of light and geochemical energy into biomass within these complex microbial systems is not clear. To gain a comprehensive understanding of these communities, we cultivate and compare members of hypersaline microbial mats for phenotypic characteristics, and use both MS proteomics and NanoSIMS stable isotope imaging to investigate mechanisms of energy flow.

To approach the relationship between microbial community composition and function in these mats, we have isolated a large number of the organisms and measured biogeochemical rates in both pure cultures and defined isolate mixtures. Isolates were tested for growth on two dozen different substrates, and whole genome metabolic reconstruction used to analyze the functional roles these isolates may play in the community. Experiments with isolate mixtures with presumably similar functional roles are being used to test the degree to which functional redundancy is important for resisting environmental stress. We have documented differences in substrate preference by heterotrophic bacteria, and salinity tolerance and photosynthesis versus irradiance relationships in cyanobacteria. We have found a positive growth-photosynthesis relationship between cyanobacteria and all heterotrophic bacteria tested to date. We are investigating these effects with a suite of 'omics techniques and a culturing approach where the three dimensional structure of natural mats is recreated using artificial substrates.

It has become increasingly clear that many microbial primary producers can also play roles as organic consumers, but there are few studies that assess metabolic regulation of photoautotroph organic matter consumption. This is especially relevant in mats because cyanobacteria produce an extensive organic extracellular matrix, providing the community with a physical buffer and a rich source of nutrients. We examined a single cyanobacterium and associated heterotrophs isolated from a microbial mat. By applying stable isotope tracing at the single cell level, we can quantify cyanobacterial assimilation of complex extracellular organic C and N under different metabolic conditions (Figure 1) (Stuart et al 2015; Stuart et al in review). We investigated the metabolic foundations of organic matter reuse by comparing exoproteome composition and incorporation of ^{13}C-^{15}N labeled extracellular organic matter in a unicyanobacterial biofilm incubated under different light regimes. Reuse by cyanobacteria accounted for almost half of all uptake in the community, indicating they are successful competitors for organic C and N. Under

lighted conditions, we measured increased excretion of extracellular polymeric substances (EPS) and proteins involved in micronutrient transport, suggesting requirements for micronutrients may drive substrate uptake during daylight hours. When photosynthesis was chemically inhibited, cyanobacteria incorporated extracellular organic matter with a low C/N ratio. By contrast, in the dark, cyanobacteria incorporated high C/N extracellular material, decreased their excretion of EPS, and increased expression of degradative exoproteins, implying use of the extracellular domain for C storage. Simultaneously, in prolonged dark incubations, associated heterotrophic bacteria increased in abundance and upregulated their transport proteins, suggesting that cyanobacterial reuse may incidentally control heterotroph resource availability under normal day-night diel regimes. Light availability and resulting metabolic status of these primary producers may dictate both composition and turnover rates of extracellular organic matter.

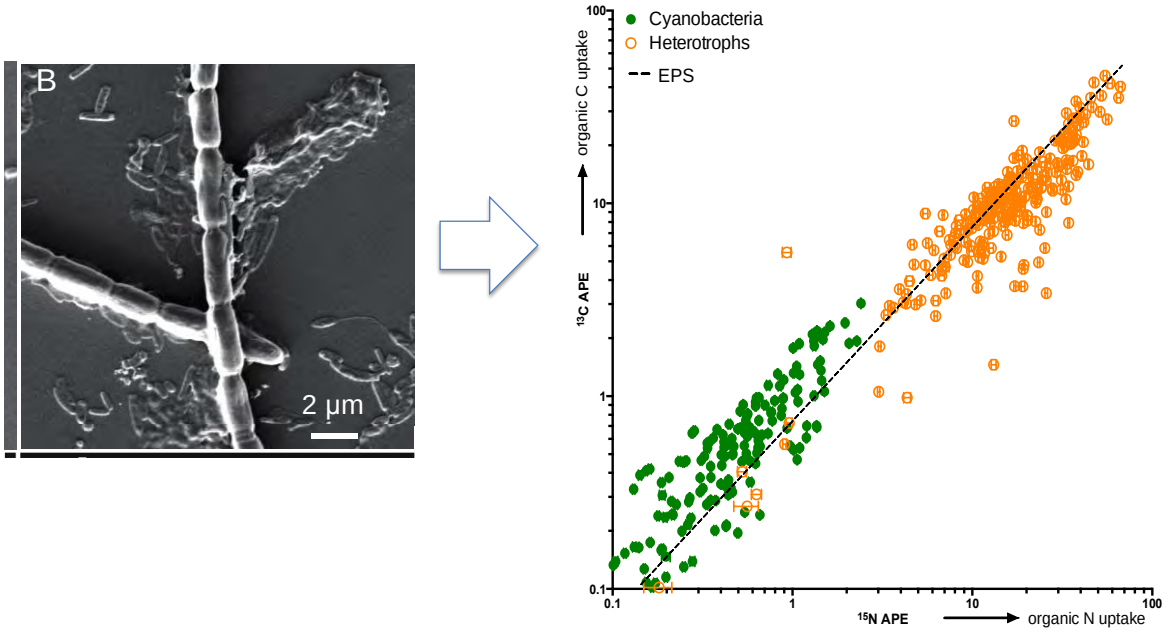


Figure 1: Stable isotope probing allows us to track extracellular organic matter incorporation in a mixed biofilm community. (Left) SEM image of cyanobacterial trichomes and associated microbes in a biofilm. (Right) ^{13}C and ^{15}N enrichment of cells analyzed via NanoSIMS. Each point represents ^{13}C and ^{15}N atom percent excess (APE) for a single trichome (solid) or bacterial cell (outlined). Dotted line indicates the ratio of ^{13}C-^{15}N labeled EPS-substrate, as determined by IRMS.

References

1. Stuart, R.K., Lee, J.Z., Everroad, R.C., Mayali, X., Lipton, M., Boggs, M., Stannard, W.L., Bebout, L., Bebout, B., Weber, P., Pett-Ridge, J. and Thelen, M. 2015 Resource allocation in photosynthetic microbial mats: light-dependent cyanobacterial extracellular polymeric substances (EPS). *ISME Journal,* doi: 10.1038/ismej.2015.180 **Top 10 most downloaded.**
2. Stuart, Rhona K., Xavier Mayali, Amy Boaro, Mary Lipton, Jackson Z. Lee, R. Craig Everroad, Brad M. Bebout, Peter K. Weber, Jennifer Pett-Ridge, and Michael P. Thelen. Light dependent recycling of cyanobacterial extracellular matrix products in a simplified microbial mat community. *ISME Journal*, in review.

This research was supported by the LLNL Biofuels Scientific Focus Area, funded by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under FWP SCW1039. Work at LLNL was performed under the auspices of the U.S. Department of Energy under Contract DE-AC52-07NA27344 and at Pacific Northwest National Lab supported by the OBER/GSP under the Pan-Omics project.

Taxonomic composition of simplified bacterial communities conferring growth and biomass enhancements to biofuel-producing algae

Ty J. Samo*[1] (samo1@llnl.gov), Jeffrey A. Kimbrel[1], Daniel J. Nilson[1], Erin E. Nuccio[1], **Jennifer Pett-Ridge[1]**, Anthony Siccardi[2], Todd W. Lane[3], and **Xavier Mayali**[1]

[1]Lawrence Livermore National Laboratory, Livermore CA; [2]Texas A&M Agrilife, Corpus Christi, TX; [3]Sandia National Laboratories, Livermore, CA, USA

Project Goals: The LLNL Biofuels SFA seeks to support robust and sustainable microalgae fuel production through a systems biology understanding of algal-bacterial interactions. We hypothesize that by understanding the factors that control cellular physiology and biogeochemical fluxes in and out of algal cells, particularly through the phycosphere, we can advance the efficiency and reliability of algal biofuel production. Our research includes studies of probiotic traits of phycosphere-associated bacteria, systems biology studies of model algae, and genome-enabled metabolic modeling to predict the interspecies exchanges that promote algal growth, lipid production and healthy co-cultures. Our overall goal is to develop a comprehensive understanding of complex microbial communities needed to advance the use of biological properties for practical energy production.

Mutualistic algal-bacterial interactions may arise when bacteria provide metabolically beneficial substances to their algal partners in exchange for energy-dense organic compounds. We seek to understand these positive exchanges by establishing closely interacting algal-bacterial cultures, simplifying the resultant algal-attached bacterial communities through enrichments, and monitoring the algae for elevated growth and biomass characteristics. After initial screening and downselection, we have analyzed the bacterial community composition of nearly 100 samples to correlate the presence/absence of specific bacteria with measured of algal health, identify putative modes of metabolite transfer from bacterium to alga, and inform upcoming genome sequencing endeavors which will inform metabolic modeling of the effects of individual beneficial bacteria.

We used two model biofuel-producing microalgal strains, *Phaeodactylum tricornutum* (*Pt*) and *Nannochloropsis salina* (*Ns*), to enrich for growth-promoting bacteria from established algal raceway ponds in Texas and the coastal Pacific Ocean. Enrichments were evaluated for increases in chlorophyll fluorescence (an estimator for health and growth). Eventually, 13 *Ns* and 36 *Pt* enrichments were established and their productivity was evaluated under 24 h light and 12 h light-dark cycles. The DNA of bacterial communities of the algal phycospheres versus the total community were collected, the 16S rRNA gene partially sequenced and compared.

Results indicate enhanced algal growth and yield in many enrichments, highlighting the successful establishment of beneficial bacterial communities. We confirmed bacterial attachment to algal phycospheres via fluorescence microscopy, scanning electron microscopy, and NanoSIMS imaging. Light cycle affected the prevalence of beneficial interactions in both the *Ns* versus the *Pt* enrichments, although in different directions. As expected, bacterial community richness of enrichments was significantly reduced relative to source communities. In all simplified communities, the majority of bacteria belonged to the *Rhodobacteraceae*. At the

genus level, *Phaeobacter* were abundant regardless of algal host. In *Pt* cultures *Labrenzia*, *Loktanella* and *Hyphomonas* were abundant members, whereas the *Ns* cultures were enriched in *Marivita*, *Marinobacter* and *Algoriphagus*. These abundant taxa were not always correlated with a probiotic effect. Interestingly, two *Loktenella* species appeared to have opposing effects depending on the host, with *L. vestfoldensis* being beneficial to *Ns* and detrimental to *Pt*, while *L. rosea* had the opposite effect.

The positive growth effects observed in these enrichments suggest that maintenance of algal-bacterial mutualisms may assist with establishing robust algal biofuel cultures. The high proportion of *Rhodobacteraceae* in the enrichments is an encouraging observation since several sub-groupings of this lineage are major contributors to algal health in natural ecosystems. For example, it has been demonstrated that some species release growth-enhancing hormones[1] or vitamins[2] to an algal cohort and, in some cases, antibiotics to kill algicidal bacteria[3]. Future work will further characterize these relationships using experimental and bioinformatic approaches and identify the chemical and ecological mechanisms underpinning these complex symbioses.

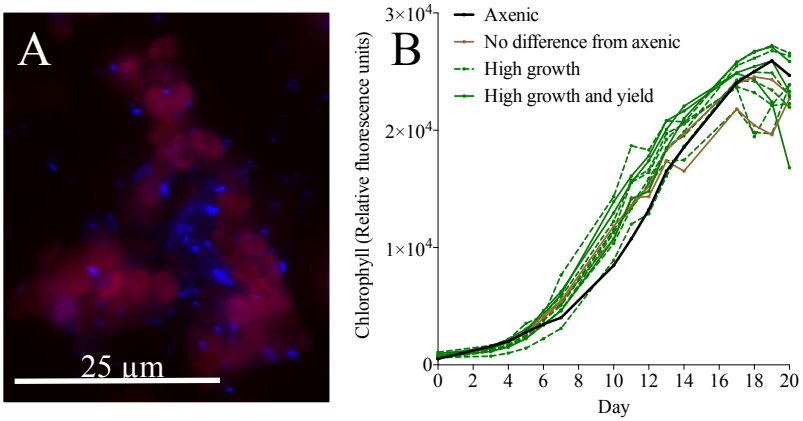


Figure 1. (A) Fluorescence micrograph of *N. salina* (red) with associated bacteria (blue). (B) Growth and yield enhancement measurements of *N. salina* enrichments.

Citations

Amin, S. A., L. R. Hmelo, H. M. van Tol, and others. 2015. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. Nature **522**: 98–101

Durham, B. P., S. Sharma, H. Luo, and others. 2015. Cryptic carbon and sulfur cycling between surface ocean plankton. Proc. Natl. Acad. Sci. **112**: 453–457

Seyedsayamdost, M. R., R. J. Case, R. Kolter, and J. Clardy. 2011. The Jekyll-and-Hyde chemistry of Phaeobacter gallaeciensis. Nat. Chem. **3**: 331–335

This research was supported by the LLNL Biofuels Scientific Focus Area, funded by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under FWP SCW1039. Work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344.

A Trait Based Dynamic Energy Budget Approach to Explore Emergent Microalgal Community Structure

Yiwei Cheng[1*] (yiweicheng@lbl.gov), Ulas Karaoz[1], Nicholas J. Bouskill[1], **Jennifer Pett-Ridge**[2], **Xavier Mayali**[2] and **Eoin L. Brodie**[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Lawrence Livermore National Laboratory, Livermore, CA.

Project Goals: The LLNL Biofuels SFA seeks to support robust and sustainable microalgae fuel production through a systems biology understanding of algal-bacterial interactions. A goal of the SFA is the predictive modeling of complex microbial communities in open pond microalgae dominated ecosystems. Towards achieving that goal, a microbial model capable of representing key mechanistic properties that determine fitness of microbial populations in dynamic systems, and with robust mechanistic representation of intra-cellular resource allocations is key to improve realism of the simulated microbial community dynamics. **Mechanistic models also have the potential to enhance fundamental understanding of the underlying ecological principles of algal-bacterial interactions. Through integration with empirical knowledge, mechanistic models also provide a more robust platform valid for wider range of scenarios.**

Research has been underway for decades to realize the full potential of algal biofuels at the commercial scale, however, monoculture algal ponds are typically threatened by collapse due to microalgal grazing and parasite invasions. Recently, it has been proposed that functionally diverse microalgal-bacterial communities can achieve higher biomass and/or lipid yields, and are more stable (less susceptible to invasion) than monocultures. Similar positive diversity-productivity relationships have been observed in a wide range of ecosystem studies, but the purposeful maintenance of a diverse microbiome is less common in managed systems.

We have developed a trait based dynamic energy budget model to explore emergent microalgal community structure under various environmental (e.g. light, temperature, nutrient availability) conditions. We initially reduce the complex algal community into functional groups (guilds). Each microbial guild (algae or bacteria) is characterized by distinct combination of physiological traits (e.g. nutrient requirement, growth rate, substrate affinity, lipid production) constrained by biochemical trade-offs. These trait values can be derived from literature or genomic data. Metabolism of the algae and the bacterial species (symbiotic or non-symbiotic) are described within a dynamic energy budget framework. Currently, the trait values follow those from the literature and will be refined based on the results from experimentation and flux balance and multi-objective flux modeling, represent initially *Phaeodactylum tricornutum*, *Chlamydomonas reinhardtii* and *Nannochloropsis salina* as model algal species Our model offers a mechanistic framework to predict the optimal microalgal community assemblage towards achieving higher productivity and resistance to invasion under prevailing environmental conditions.

In our initial investigation into the mathematical representation of uptake kinetics, we compared simulation results of two uptake formulations: equilibrium chemistry approximation kinetics (ECA) and Michaelis-Menten (MM) kinetics. The ECA formulation has previously been shown to be more accurate in certain cases such as a single microbe utilizing multiple substrates and multiple microbes utilizing multiple substrates. In this work, initial simulation results showed that MM kinetics tend to overestimate algal biomass when compared to ECA kinetics.

This research was supported by the LLNL Biofuels Scientific Focus Area, funded by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under FWP SCW1039. Work was performed at Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-05CH11231 in coordination with Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344.

Multi-omics Approach Unveil Microbial Transformations of Lignocellulose in the Gut of the Wood-Feeding Beetle *Odontotaenius disjunctus*

Javier A. Ceja-Navarro (JCNavarro@lbl.gov) [1*], Ulas Karaoz[1], Zhao Hao[1], Markus Bill[1], Richard A. White III[2], Mary S. Lipton[2], Joshua N. Adkins[2], Timothy Filley[3], Meredith Blackwell[4], Xavier Mayali[5], Peter Weber[5], **Jennifer Pett-Ridge[5], Eoin L. Brodie[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Pacific Northwest National Laboratory, Richland, WA; [3]Purdue University, West Lafayette, IN; [4]Louisiana State University, Baton Rouge, LA. [5]Lawrence Livermore National Laboratory, Livermore, CA.

https://envmicro.wordpress.com/research-projects/passalid-beetles-natures-efficient-lignocellulosic-biorefineries/

Project Goals: Our research aims to develop an integrated analysis of energy flow in complex microbial communities by combining multi-scale approaches including biogeochemical, stable isotope probing, metagenomic/transcriptomic, proteomic and computational analyses, to understand carbon and nutrient transformation relevant to biofuel production in complex microbial communities. A comprehensive understanding of such communities may help in the development of efficient, industrial-scale processes for lignocellulose degradation.

Odontotaenius disjuctus is a wood feeding beetle that processes large amounts of hardwoods and plays an important role in forest carbon cycling. In its gut, plant material is transformed into simple molecules by sequential processing during passage through the insect's digestive system. Fourier Transformed Infrared Spectroscopy – Attenuated Total Reflectance (FTIR-ATR) demonstrated the sequential transformation of cellulose, xylan and lignin, and the accumulation of reduced nitrogen through the beetle digestive system. To identify the organisms and pathways contributing to these processes, we used multiple 'omics approaches to analyze the distribution of the different symbiotic communities and their specific functions in lignocellulose deconstruction within the insect's gut.

Fosmid clones were selected and sequenced from a pool of clones based on their expression of plant polymer degrading enzymes, allowing the identification of a wide range of carbohydrate degrading enzymes from different microorganisms associated with the beetle. Comparison of metagenomes from four gut regions demonstrated that lignin-degrading genes were more abundant in the first two gut sections, the foregut and midgut. Cellulose, starch, and xylan degradation genes were more abundant in the midgut and posterior hindgut. Genes for hydrogenotrophic methanogenesis and for nitrogen fixation were more abundant in the anterior hindgut. Assembled scaffolds were binned into 127 genome bins representing Bacteria, Archaea, Fungi, and Nematoda. Eleven nearly complete genomes were reconstructed, allowing us to identify linked functions/traits, including organisms with cellulosomes, and a combined potential for cellulose, xylan, starch hydrolysis and nitrogen fixation. A mixed eukaryotic bin containing different lignin peroxidases, catalase peroxidases and lacasses of fungal origin was also identified.

A metaproteomic study was conducted to determine the expression of these pathways. Preliminary analyses suggest enrichment of pathways related to hemicellulosic degradation in the midgut and anterior hindgut. A complete xylan degradation pathway was reconstructed and complementary GC-MS/MS based metabolomics identified xylobiose and xylose as major metabolite pools. To test 'omic generated hypotheses of *in situ* metabolism in the beetle gut, we used Chip-SIP stable isotope tracing, to analyze the isotopic composition of microbial RNA from beetles fed ^{13}C-cellulose. Multiple bacterial groups (including the Spirochaetaceae, Ruminococcaceae, Rhodospirillaceae, Thermotogaceae, and Promicromonosporaceae) were ^{13}C enriched, mainly in the midgut. Ongoing cultivation studies are focusing on these community members.

Our combined multi-omics and analytical chemistry approach demonstrates the continuous transformation of lignocellulosic materials through the beetle gut and the contribution of organisms belonging to multi-trophic levels to the different metabolic processes.

References

This research was supported by the LLNL Biofuels Scientific Focus Area, funded by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under FWP SCW1039. Work was performed at Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-05CH11231, at the Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344, and at Pacific Northwest National Lab supported by the OBER/GSP under the Pan-Omics project. We also acknowledge EMSL support to E. Brodie through proposal #47988.

In silico analyses of interactions between systems objectives for engineering of biological communities

Ali Navid,[1]* (navid1@llnl.gov), Yongqin Jiao,[1] and **Jennifer Pett-Ridge**[1]

[1]Lawrence Livermore National Laboratory, Livermore, CA

Project Goals: Genome-scale models of metabolism have been developed for a wide-range of different organisms with different modes of metabolism and growing in a diverse set of environments. Using Flux Balance analysis (FBA) to assess the metabolic phenotypes of these systems has been very successful. However, as the quantity and quality of in situ data for complex biological systems improve, we have the opportunity to assess the trade-offs among various biological objectives of systems and predict deviations from the maximum growth paradigm that has dominated nearly all FBA analyses. To this end, we are developing a high-performance computing tool for Multi-Objective Flux Analysis (MOFA) of biological systems. This mode of analysis becomes particularly important for studying multi-cellular communities as well as engineering of systems to produce compounds of interests.

Genome-scale models of metabolism have become a standard tool for analyzing metabolic capabilities of biological systems. Examination of the models with Flux Balance Analysis[1] (FBA) has informed us about metabolic activities of individual organisms as well as general truths about universal metabolic characteristics. For FBA, available annotated genomic information about a system is used to develop a system-level reconstruction of the metabolic network. The model is constrained using fundamental physico-chemical principles, experimental observations, and an assumption that the system is at steady state. Linear programming is then used to optimize for a metabolic flux pattern that results in the optimum value of a biological function (typically growth).

While FBA models have proven very useful in identifying factors that are critical for optimum cellular growth (or other biological objectives) under well defined environmental situations, the necessity to optimize only one objective limits the use of this method for analysis of multi-cellular communities or examining the trade-offs among competing cellular objectives.

Although a number of FBA-based methods have been developed to study inter-cellular interactions (*e.g.*, COMETS[2] and OPTCOM[3]), none of them systematically examines the tradeoffs among the system's objectives. This knowledge about how an alteration in activity of a cellular process would affect other important biological processes is particularly crucial for design of metabolic engineering projects that aim to enhance a cellular activity. The results can also be used to inform trait-based models of multi-cellular communities[3]. This type of examination requires development of genome-scale multi-objective flux analysis (MOFA) models that would account for the different metabolic objectives in a system or for each member of a community**.**

Multi-objective optimization (MO) is a critical tool in a number of fields where a decision maker needs to consider tradeoffs between various conflicting objectives. Simulating genome-scale MOFA models of metabolism requires use of high performance parallel computing. LLNL's

extensive computational capabilities allow us to develop tools that would permit such undertakings.

As a test case, we used MOFA to examine the diverse metabolic phenotypes of purple non-sulfur bacterium, *Rhodopseudomonas palustris* (RP). We have developed a highly curated genome-scale model of RP's metabolism and constrained it with a large volume of experimental measurements. While FBA analyses provided us with information about which metabolic pathways were used for carbon fixation, hydrogen production, as well as proton economy of RP for a variety of different nutrient sources, MOFA analyses informed us that under light-anaerobic conditions the observed metabolic behaviors are not in agreement with FBA's central paradigm of cells growing at maximum feasible rates.

Our MOFA analyses show that in light-anaerobic conditions RP grows at slightly lower than theoretical maximum rates. Mapping flux measurements onto the MOFA predicted multi-dimensional Pareto front show that: 1) growth under these conditions is limited by the amount of light the cell absorbs; 2) depending on the nutrient-source, proton metabolism of the cell can effect the rate of growth; 3) RP's metabolism is geared toward optimum carbon efficiency and not maximum rate of growth; 4) hydrogen production by RP results in reduced rate of carbon fixation and cellular growth.

References

1. Orth et al. (2010), Nat. Biotechnol., **28**, 245.
2. Harcombe et al. (2014), Cell Rep., **7**, 1104.
3. Zomorrodi et al. (2012), Plos Comp. Biol, 8.
4. Bouskil et al. (2012), Front. Microbiol, **3**, 364.

This research was supported by the LLNL Biofuels Scientific Focus Area, funded by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under FWP SCW1039. Work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344.

Into the deep: variability in soil microbial communities and carbon turnover along a tropical forest soil depth profile

Jennifer Pett-Ridge[1*] (pettridge2@llnl.gov), Karis McFarlane[1], Katherine Heckman[1,2], Ashley Campbell[1], Amrita Bhattacharyya[1,3], Sasha Reed[4], Peter Nico[3], Tana Wood[5]

[1]Lawrence Livermore National Laboratory, Livermore, CA; [2]U.S. Forest Service Northern Research Station, Houghton, MI; [3]Lawrence Berkeley National Lab, Berkeley CA, [4]USGS Southwest Biological Science Center Moab, UT; [5]U.S. Forest Service, International Institute of Tropical Forestry, Rio Piedras Puerto Rico.

Project Goals: This Early Career research examines the genomic potential and expression of tropical soil microorganisms as they experience shifts in soil temperature, moisture, and depth and oxygen availability. By also tracking the degradation and fate of organic carbon compounds, this work will increase the accuracy of predictions about how microbial processes affect whether organic carbon is retained or lost from tropical systems. The mechanistic understanding produced by this research will directly benefit attempts to improve the predictive capacity of mathematical models that forecast future tropical soil carbon balance.

Tropical forest soils store more carbon (C) than any other terrestrial ecosystem and exchange vast amounts of CO_2, water, and energy with the atmosphere. Much of this C is leached and stored within deep soil layers, but we know exceedingly little about the fate of this C or the microbial communities that drive deep soil biogeochemistry. From data that do exist, most organic matter (OM) in tropical soils appears associated with mineral particles, suggesting deep soils may provide greater C stabilization due to organo-metal co-precipitation and mineral-surface interactions. However, few studies have evaluated sub-surface soils in tropical ecosystems, the turnover times of deep soil C, and sensitivity of this C to global environmental change.

To address this critical research need, we quantified C pools, microbial communities and soil radiocarbon turnover times in bulk soils and soil fractions [free light (unprotected), dense (mineral-associated)] from 0-140 cm in replicate soil pits in the Luquillo Experimental Forest, Puerto Rico. Unsurprisingly, we found soil C, nitrogen, and root and microbial biomass all declined exponentially with depth; total C stocks dropped from 5.5 % at the surface to <0.5% at 140cm depth. Soil OM ^{14}C and mean turnover times were variable across replicate horizons, ranging from 3-1500 years at the surface (0-20 cm), to 5000-40,000 years at 140 cm depth. Soil C in the mineral associated fraction was much older than the free light fraction C, which reflected modern ^{14}C at all depths. In comparison to temperate deciduous forests, these ^{14}C values reflect far older soil C, and OM decomposition that highly favors free light C pools, even at depth. While previous work suggests these low C tropical subsoils contain small but metabolically active microbial communities at depths of ~100cm, these organisms appear highly OM limited, and preferentially degrade recent inputs.

In the coming half century, tropical forests are predicted to see a 2 – 5 ° C temperature increase and substantial differences in rainfall amount and timing. The data described here represent baseline data for a site now undergoing a 4°C warming experiment; upcoming research will

examine soil C storage and mean residence times during and post-warming to improve numerical models of ecosystem warming effects in tropical forests.

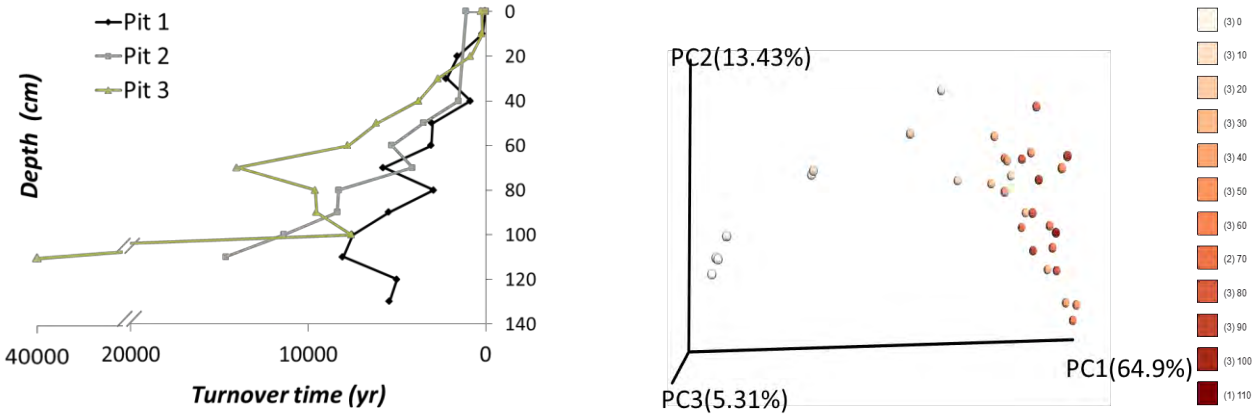


Fig 1. Left: $\Delta^{14}C$ turnover times increased significantly with depth, reaching up to 40,000 yrs in some of the deepest samples. Right: PcoA ordination highlighting the distinction between surface soils and those below the active root zone (using weighted Unifrac distance).

This research was supported by an Early Career Research Program award to J. Pett-Ridge, funded by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under FWP SCW1478. Work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344.

Characterization of a Metal-induced Sensory Transduction System in *Caulobacter crescentus*

Dan Park,[1] Wes Overton,[1] Amy Boaro, [2] Mary Lipton,[2] Grant Bowman,[3] and **Yongqin Jiao[1]*** (Jiao1@llnl.gov)

[1]Lawrence Livermore National Laboratory, Livermore CA; [2]EMSL, Pacific Northwester National Laboratory, [3]University of Wyoming, Laramie, WY

Project Goals: Depleted uranium (U) is a widespread environmental contaminant that poses a major threat to human health. In contrast to humans and animals where trace amount of U can cause damage to kidneys, liver and heart, it is well known that some bacteria can tolerate high levels of U and influence its mobility and bioavailability in the environment. As a non-pathogenic bacterium, *Caulobacter crescentus* is an attractive bioremediation candidate due to its high tolerance to heavy metals, and its ability to mineralize U by facilitating uranium phosphate precipitation. Our goal is to decipher the physiological basis for U tolerance in *C. crescentus*, and provide insight into the effect of aerobic bacteria on U biogeochemistry and assess the utility of them in biomineralization applications.

Maintaining homeostasis for biologically required metal ions and detoxifying the incursion of toxic metal ions require extensive regulatory and response machineries. These processes are essential for cell survival especially under conditions of metal deprivation or overload. Earlier work from Dr. L. Shapiro group at Stanford U. identified a small periplasmic protein UrcA in *Caulobacter crescentus* whose expression level increases in responding to the presence of U. To further understand the specificity of this U response, we examined the change in expression of *urcA-lacZ* in responding to several metals including Ag, Cu, Ni, Cd, Co, Cr, and Zn, along with several other stress conditions. We observed that besides U, *urcA* expression was also induced sharply by Zn. We do not yet fully understand why uranium and Zn induce *urcA* expression; Through RNA-seq and proteomic analyses, we found both U and Zn cause cell envelope stress in *Caulobacter*.

To identify regulators involved in controlling the expression of *urcA*, we performed a genetic screen to examine disruption of which gene affects *urcA-lacZ* expression. A total of 60,000 transposon mutants were screened and we identified a preciously uncharacterized two-component system, named urcS/urcR, that serves as a direct activator of *urcA*. Inactivation of either *urcS* or *urcR* completely abolished the *urcA-lacZ* expression in the presence of Zn. Furthermore, direct binding of cytoplasmic response regulator UrcR with DNA upstream of *urcA* was confirmed by an electrophoretic mobility shift (EMSA) assay.

Besides the urcS/R two component system that is a direct activator of *urcA*, we found three other proteins that are indirect repressors of *urcA*, all of which act through urcS/R. These include an ABC transporter with a peptidase domain (pepN), a conserved membrane protein and two general stress regulators belong to the MarR family. Genetic disruption of any of these proteins caused constitutive expression of *urcA*, independent of Zn or U. Zn addition further increased the *urcA* expression in these mutant backgrounds however, suggesting an alternative mechanism for Zn-induced *urcA* expression

independent of these repressors. In particular, disruption of key amino acids of the peptidase domain of the ABC transporter-pepN protein or the ATP-binding domain of its neighboring ATPase caused upregulation of *urcA,* suggesting a fully functional protein is required to keep *urcA* at the off state in the absence of the environmental signal (U or Zn).

To identify others genes belong to the UrcR regulon other than *urcA*, we performed RNA-seq analysis comparing genomic-wide mRNA expression profile between wild type and the *urcR* null mutant with or without Zn. Our results showed that UrcR is a global regulator of 165 ORFs, the majority of which encode non-cytoplasmic proteins and contain the previously predicted m_5 motif (metal specific genes). Besides proteins of unknown functions, UrcR also regulate 9 putative non-cytoplasmic peptidases and a few multi-drug efflux pumps. Current work focuses on deciphering the nature of the signal(s) generated by metals (Zn or U) that result in *urcA* induction. The signaling pathway identified will help aid the sensing and stress response studies in *Caulobacter* and has important applications for environmental metal detection and remediation.

This study was supported by a Department of Energy Early Career Research Program award from the Office of Biological and Environmental Sciences (to Y.J.). This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 (LLNL-ABS -648058).

The Application of Ecosystem Modeling at Subsurface Sites to Study Carbon, Nutrient and Metal Cycling in Microbial Communities

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

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

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

To meet the goals of this ongoing project, two subsurface sites from the Department of Energy (DOE) Rifle Integrated Field Research Challenge Site (RIFRC) are being interrogated using a suite of metaomic approaches. The first site consists of sediments from the Winchester 2007 gallery, 'JB" well locations and was chosen due to the occurrence of natural attenuation of uranium (uranium reduction in the absence of biostimulation or other remedial interventions). The second and more recent sites of study within this project have been collected from Colorado River Floodplain (CORFP) sediments representing recent sediment depositions. Overbank deposits in the floodplain have become enriched in C, Fe and S minerals. Aggradation processes have led subsequent burial of these enriched sediments creating "hotspots" of biogeochemical activity which serve as analogs to the buried naturally reduced sediments at the JB sites.

In this current work, we describe the generation of a framework for microbial community modeling that couples collections of genome-scale models of bacterial metabolism from the subsurface ecosystem for the purpose of predicting how a complex microbial community may respond to changes in the subsurface environment. This effort is critical as it provides an important analytical strategy for integrating and interpreting metaomics data as well as the capacity to simulate various interactions between environmental conditions and microbial communities. Guided by results from previous metagenomic and metatranscriptomic analyses from a site identified as undergoing natural attenuation of uranium at the RIFRC (JB-05) and the CORFP site, collections of genome-scale metabolic models representing seven bacterial classes were chosen including (a) Beta-proteobacteria (representative species: *Thiobacillus denitrificans*, *Rhodoferax ferrireducens* and *Variovorax paradoxus*); (b) Delta-proteobacteria (representative species: *Geobacter metallireducens* and *Desulfovibrio vulgaris*); (c) Bacilli (representative species: *Bacillus subtilis*); (d) Gamma-proteobacteria (representative species: *Pseudomonas stutzeri*); (e) Alpha-proteobacteria (representative species: *Rhodopseudomonas palustris*); (f) Actinobacteria (*Arthrobacter* sp.) and (g) Clostridia (*Clostridium cellulolyticum*), and metabolic models constructed. This framework was then divided into two main components: 1) the

bacterial community and 2) the extracellular subsurface ecosystem. The mixed microbial community model was placed in the subsurface ecosystem (outer environment), and interactions among community members as well as the exchange between the community and the outer environment were captured by a dynamic multi-species metabolic modeling (DMMM) approach.

The ecosystem model indicates that *T. denitrificans* may dominate the community at the JB site due to its ability to use inorganic electron donors for energy and fix CO_2 as its carbon source effectively bypassing any limitations of bioavailable organic carbon. Through electron transport with cytochrome bc1 complexes and NADH-Q oxidoreductase, a tight coupling between Fe(II) oxidation and nitrate reduction can be established to support CO_2 fixation as the main carbon source. This is in accordance with the bioinformatic analyses of taxonomic and functional profiles from the JB site indicating that *T. denitrificans* possesses a modified complete Calvin cycle for CO_2 fixation, and is the most abundant microorganism available among the seven bacterial classes. In contrast, the ecosystem model shows that in acetate amended sites, the delta-proteobacteria, *G. metallireducens*, would be numerically dominant over other bacterial classes, due to its ability to fix N_2 to complement or overcome limitations of organic nitrogen allowing for rapid growth during biostimulation.

In addition to capture of the community structure in terms of the relative abundance approximating the RIFRC, the ecosystem model is also able to predict how the components of the microbial community may respond to changes in the environment they inhabit by altering conditions in the extracellular subsurface ecosystem component of the model. Results from the model simulations show for example, that the community structure is highly responsive to acetate concentrations in the subsurface. As acetate concentrations increase (from 0 nmol/l to 300 nmol/l) the relative abundance of the gamma-proteobacteria also increases (from ~4%-40%) while that of the beta-proteobacteria decreases (from ~76%-35%). Further, the model suggests that *R. ferrireducens* may become numerically dominant over *T. denitrificans* as acetate concentrations increase, a finding which has been supported by experimental evidence. Similarly, the model simulations show that nitrate concentrations are are also important in driving community dynamics. When nitrate concentrations are increased (0 nmol/l to 750 nmol/l) and organic carbon such as acetate is low (180 nmol/l) the model predicts a community structure similar to that measured using metaomic approaches at the JB site in which the beta-proteobacteria class (in particular *T. denitrificans*) are numerically dominant over the other bacterial classes (increasing to ~43%).

Overall, the community model presented here captures, at least in part, the microbial community structure that was observed using metaomic approaches at RIFRC sites and provides an important framework for continued community modeling development. The model as created here is capable of predicting the response of the community structure in changing environments such as anoxic/oxic conditons or limitations by carbon or nutrients which is critical to understanding carbon and energy flows in an ecosystem leading to improved predictions that can be used to design more efficient remediation and management strategies and better understand the implications of environmental perturbations such as climate change.

This project is supported by the Office of Biological and Environmental Research in the DOE Office of Science DESC0006837.

Title: Phylogenetic signal in soil fungal and bacterial communities in response to experimental nitrogen addition

Rebecca C. Mueller* (rmueller@lanl.gov), La Verne Gallegos-Graves[1], **Cheryl R. Kuske**

[1] Los Alamos National Laboratory, Los Alamos, NM
[2] United States Geological Survey, Moab, UT
[3] University of New Mexico, Albuquerque, NM

http://www.lanl.gov/science-innovation/capabilities/bioscience-biosecurity-health/environmental-microbiology/soil-carbon.php

Project Goals: Anthropogenic N deposition is a chronic and increasing condition in temperate regions that may strongly influence C cycling dynamics. Large increases in N addition have been seen in eastern forests, which have the potential to act as either a source of C or as a major terrestrial C sink. One major theme of our Science Focus Area is to determine the influence of chronic N deposition on microbial C cycling processes in temperate regions, particularly mesic forests. In temperate forest ecosystems, fungal and bacterial biomass is concentrated in shallow surface soil strata where C and N cycling are major processes. The goals of this project are to correlate the resident fungal and bacterial communities, enzyme activities, and local geochemistry across strongly stratified forest soils, determine the impacts of chronic N amendment on soil bacterial and fungal communities across soil layers within a pine forest, and use phylogenetic methods to quantify the effect of N on patterns of soil microbial community assembly. Achieving these goals will provide an understanding of the responsive members of the microbial community within forest soils, and provide insight into whether community shifts and responses vary with environmental conditions, as well as how those changes may impact C cycling. This poster focuses on our recent results from a long-term N deposition experiment conducted in a temperate pine forest in North Carolina.

We used high throughput sequencing methods to target the soil bacterial and fungal communities, which together are major players in N and C cycles, and are applying multivariate statistical approaches to examine shifts in the phylogenetic structure of these communities, as well as the phylogenetic signal in N response among community members. To conduct this study, we designed and validated a new fungal PCR primer (LR22R) to target an approximately 300–400 bp region of the D2 hypervariable region of the fungal LSU for use with the Illumina MiSeq platform. Both *in silico* and empirical analyses showed that the LR22R–LR3 pair captured a broad range of fungal taxonomic groups with a small fraction of non-fungal groups. Based on analysis of known and environmental communities, the new primer had broad utility for constructing accurate phylogenetic trees for fungi, which allowed us to utilize phylogenetic metrics to quantify shifts in both the bacterial and fungal communities. Using highly replicated rDNA surveys we found strong phylogenetic signal in the axonomic response to N deposition for both the bacterial and fungal communities, and the response was stronger for the bacterial community. Strong shifts in bacterial communities, with relatively minor response in the fungal community suggest that nutrient cycling will be increasingly fungal-dominated in response to N additions. By combining sequence-based community analyses with soil chemistry and enzyme activity measures, we aim to identify key responsive community members with relevance as indicators of community change and with utility for modeling soil processes.

References

Mueller RC, LV Gallegos-Graves, CR Kuske (2015) A new *fungal large subunit ribosomal RNA primer for high-throughput sequencing surveys*. FEMS Microbiology Ecology http://dx.doi.org/10.1093/femsec/fiv153

Funding statement: The information in this poster was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (2016LANLF260).

Title: Responses of bacterial and fungal communities under nitrogen amendment differ by microhabitat in arid ecosystems

Cheryl R. Kuske* (Kuske@lanl.gov), Rebecca Mueller, La Verne Gallegos-Graves[1], Jayne Belnap[2], Sasha C Reed[2], Robert Sinsabaugh[3]

[1] Los Alamos National Laboratory, Los Alamos, NM
[2] United States Geological Survey, Moab, UT
[3] University of New Mexico, Albuquerque, NM

http://www.lanl.gov/science-innovation/capabilities/bioscience-biosecurity-health/environmental-microbiology/soil-carbon.php

Project Goals: Anthropogenic N deposition is a chronic and increasing condition in temperate regions that may strongly influence C cycling dynamics. One major theme of our Science Focus Area is to determine the influence of chronic N deposition on microbial C cycling processes in two major biomes of Earth's temperate regions, forests and arid grass/shrublands. In both biomes fungal and bacterial biomass is concentrated in shallow surface soil strata where C and N cycling are major processes. Arid lands cover ~40% of the terrestrial surface of the earth, and are expanding in many regions as a result of desertification. Historical inputs of N to drylands are generally low, but are increasing due to human activities. This poster will illustrate correlations among the resident fungal and bacterial communities, their enzyme activities, and local geochemistry in microhabitats of biocrust-dominated soils of an arid shrubland (Nevada) and an arid grassland (Utah), and the ability of phylogenetic rRNA gene surveys and soil enzyme assays to detect shifts in community structure and concomitant changes in C cycling processes in response to altered N conditions. Overall our goal is to provide an understanding of the active and responsive components of arid land soils that contribute to C cycling, their collective responses to environmental change.

We are using high throughput sequencing methods to target the soil bacterial and fungal communities, which together are responsible for the majority of nutrient cycles, including the N and C cycles. These studies were conducted at experimental N deposition field experiments located in an arid shrubland (Nevada) and an arid grassland (Utah) in the Southwestern US. To predict the functional effects of community shifts, we are simultaneously measuring relative microbial biomass, soil chemistry and enzyme assays to link specific environmental shifts to community responses and their functional consequences.

In the shrubland, bacterial and fungal biomass was significantly higher in association with interspace biocrusts (0-1 cm depth) when compared to homogenized 0-10 cm depth, and was higher in association with shrub canopies than in the interspace. Both communities showed differences in taxonomic composition across soil depth and with plant or biocrust association. For example, the bacterial phyla Acidobacteria, Actinobacteria, and Chloroflexi and the fungal classes Leotiomycetes, Sordariomycetes, and Eurotiomycetes were significantly higher in soils below the biocrust layer. Members of these taxa harbor a wide variety of C decomposition abilities, from use of C_1 compounds to complex lignocellulose, and their stratified location suggests microhabitat-specific C cycling activities. Although N amendment (at 8 or 15 kg/ha/yr) significantly increased available N and P by 30% and 25%, respectively, the responses of the fungal community were limited. Fungal richness changed and the Sordariomycetes increased. In

contrast to the fungal response, the bacterial community responded with a decline in richness and shifts in taxonomic composition. Carbon use efficiency (CUE) calculated from soil enzyme activity ratios was increased at the highest N application rate in all microhabitats, illustrating a shift in soil C cycling pattern.

In congruence with the shrubland, soil fungal and bacterial community composition was significantly different in the interspace biocrusts, 5 cm below the biocrusts, and in association with grass root zones in the arid grassland. However, the response patterns of fungal and bacterial communities to N amendment (at 2-8 kg/ha/yr) were not significantly different in this system when assessed two weeks or several months after application. Although plant richness, diversity, and cover showed no response to N addition, there were strong linkages between plant and soil properties and microbial community structure.

In the shrubland there is evidence for shifts toward fungal-dominated nutrient cycling in response to N additions. By combining sequence-based community analyses with soil and enzyme activity measures, we may identify key responsive community members with relevance as indicators of community change and with utility for modeling soil processes. Both of the field experiments were short-term (2 years), using ecologically relevant (low) concentrations of N that may actually be deposited in these regions. Extending such studies out for longer time periods would likely demonstrate more consistent responses.

The climatic conditions and nutrient cycles in arid ecosystems differ dramatically from well-studied mesic systems. They are characterized by highly variable environmental conditions due to sporadic precipitation events, extreme temperature fluctuations and UV radiation stress, as well as low productivity, patchy distributions of biotic resources (microhabitats) and atypical sources of nutrient inputs. Given these unique characteristics, we hypothesized that responses seen in mesic forests and grasslands will apply in drylands. We conducted a meta-analysis of 15 recently N deposition studies in arid lands, and calculated N-effects from soil microbial biomass and metabolic responses. In contrast to our expectations, the critical N concentrations separating positive from negative treatment effects were comparable to microbial biomass, decomposition rates and respiration reported in broader meta-analyses of N amendment effects in mesic ecosystems. However, the large effect sizes at low N addition rates indicate that arid lands are sensitive to modest increments in anthropogenic N deposition.

References

Mueller RC, J Belnap, CR Kuske (2015) *Soil bacterial and fungal community responses to nitrogen addition across soil depths and microhabitat in an arid shrubland.* Front Microbiol doi: 10.3389/fmicb.2015.00891.

Sinsabaugh RL, J Belnap, J Rutgers, CR Kuske, N Martinez, D Sandquist (2015) *Soil microbial responses to nitrogen addition in aridland ecosystems.* Front Microbiol doi: 10.3389/fmicb.2015.00819.

Maier S, L Muggia, CR Kuske, M Grube (2016) *Fungi and bacteria in biological soil crusts*. In: Biological Soil Crusts: Structure, Function and Management, 3rd edition. Ed. J Belnap. (in press)

Reed SC, FT Maestre, R Ochoa-Hueso, CR Kuske, M Oliver, A Darrouzet-Nardi, B Darby, *et al.* (2016) *Biocrusts in the context of global change*. In: Biological Soil Crusts: Structure, Function and Management, 3rd edition. Ed. J Belnap. (in press)

Funding statement: The information in this poster was supported by the U.S. Department of

Energy Biological System Science Division, through a Science Focus Area Grant (2016LANLF260).

Title: Influence of N availability on interactions between Ascomycete fungi and Actinomycete bacteria during litter decomposition

John Dunbar[1]*(dunbar@lanl.gov), Jean Challacombe[1], La Verne Gallegos-Graves[1], Andrea Porras-Alfaro[2], Chris Yeager[1], Rebecca Mueller[1], Cedar Hesse[1], Cheryl R Kuske[1]

[1] Los Alamos National Laboratory, Los Alamos, NM
[2] Western Illinois University, Macomb, IL

http://www.lanl.gov/science-innovation/capabilities/bioscience-biosecurity-health/environmental-microbiology/soil-carbon.php

Project Goals: Atmospheric N deposition has increased N availability in natural ecosystems by 2-fold, on average, but as much as 100-fold in specific locations. Increased N availability has been widely reported to reduce decomposition of plant litter, at least on short time-scales. Studies of ecosystem response to N deposition have reported declines in plant litter decomposition rates and declines in the relative abundance of Basidiomycota fungi. Members of this phylum are root-associated plant symbionts, plant pathogens and plant biomass decomposers with key roles in acquiring N for plant growth and cycling N in surface soils. When Basidiomycota relative abundance declines, the relative abundance of Ascomycota fungi often increases. Thus, other fungi and bacteria can become more important in plant litter decomposition. Our work aims to determine how N gradients influence the activity and interactions of two broad decomposer groups---the Ascomycete fungi and Actinobacteria---that are known to harbor lignocellulosic decomposition traits and are potential key players in litter decomposition. While increases in relative abundance of these two groups have been documented in some studies, other studies suggest that increased N availability might inhibit Ascomycete activity.

We are monitoring the collective activities of mixed communities of Ascomycota and Actinobacteria in time course experiments, where defined mixtures of five fungal and five bacterial genera, decompose plant litter (arid-land grasses or pine litter) in sand microcosms under five nitrogen treatments. Given the difficulty of manipulating natural communities, defined mixtures provide the best approach to decipher functional responses, interactions, inherent biological barriers, and relevant mechanistic phenomena. To identify *general* patterns, instead of the eccentric response of one or two specific mixtures, we are documenting trends across many independent mixtures. Fungal and bacterial isolates for the mixtures were obtained from arid grassland sites and a pine forest field experiment where N application was an experimental variable. Measurements of the defined mixtures include initial biomass, CO_2 evolution over a 30-50 day time-course, initial and final community composition (rDNA surveys of fungal and bacterial composition), and metatranscriptome analyses.

Preliminary results show evidence of composition-dependent behavior. Nonetheless, strong patterns are emerging across fungal-bacterial mixtures that show a) the N concentration

can change the interaction between fungi and Actinobacteria and b) the phenomenon depends on the plant litter.

Cultures of five fungal of the mixed community species (*Aspergillus* CK392, *Chaetomium* CK152, *Coniochaeta* CK134, *Embellista* CK46, and *Phoma* CK108) were grown in the presence of different carbon sources: chitin, grass, sucrose and wood. Genomes were sequenced, assembled and annotated. Proteomic analyses of secreted proteins present in the culture supernatants were performed at EMSL, resulting in protein abundance data for each species and each carbon source. Initial analysis of the protein abundance data for the five fungal genomes revealed patterns of secreted protein abundance that correlated with the carbon growth substrate.

Comparisons of the most abundant proteins in the supernatants *Aspergillus* CK392 and *Coniochaeta* CK134 cultures revealed more total proteins present in the supernatants of the *Aspergillus* cultures than in *Coniochaeta* cultures. Examination of subsets of proteins in the Aspergillus cultures showed differences in the abundances of secreted enzymes, dependent on the carbon source. For example, the grass cultures showed higher levels of enzymes mediating degradation of complex polysaccharides, including cellulose, starch and xylan, which are all present in grasses.

Funding statement: The information in this poster was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (2016LANLF260).

Denitrification revisited: Contributions of chemodenitrifiers and fungi to soil denitrification

Jenny Onley,[1] Steven A. Higgins,[1] Luis Orellana,[2] **Konstantinos Konstantinidis**,[2] **Joanne Chee-Sanford**,[3] **Robert A. Sanford**,[4] and **Frank E. Löffler**[1,5]* (frank.loeffler@utk.edu)

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

Project Goals:
The goals of this project are to fill existing knowledge gaps in our understanding of N flux and associated C turnover in soils and sediments. Novel information about the diversity, distribution, abundance and expression of genes contributing to N transformation is required to link desirable (i.e., N retention) and undesirable (i.e., N loss, such as N_2O emissions) activities with measurable microbial parameters. Correlating molecular- and organismal-level information with environmental factors that control N and C turnover can predict the impact of land management practices on greenhouse gas emissions. Such integrated approaches generate novel information at multiple scales of resolution and contribute to system-level understanding of key nutrient cycles in soils. In the present work, we show that coupled biotic-abiotic processes contribute to denitrification, and we demonstrate the utility of new *p450nor*-targeted PCR primers to assess the fungal contribution to N_2O formation in soils.

Abstract:
To meet the needs of a growing human population, the use of N-based fertilizers has increased substantially in the last 50 years. Denitrification of nitrate is a major pathway for nitrate turnover and is associated with soil N loss and nitrous oxide (N_2O) emissions (Sanford *et al.*, 2012). The assessment of denitrification potential has been largely based on the enumeration of bacterial *nirK* and *nirS* genes, which encode nitrite reductases (nitrite → nitric oxide, NO). NO is subsequently enzymatically reduced to N_2O and dinitrogen. An alternate pathway for nitrite reduction is chemodenitrification, the abiotic reduction of nitrite to N_2O mediated by ferrous iron.
Anaeromyxobacter dehalogenans, a common and metabolically versatile soil bacterium, utilizes nitrate and ferric iron as electron acceptors. Consistent with gene content (i.e., the presence of *nrfA* encoding the ammonia-forming cytochrome *c* nitrite reductase), the organism reduces nitrate to ammonium via nitrite. Copies of the key denitrification genes *nirK* and *nirS* are absent on the genome, but an atypical Clade II *nos* operon encoding a functional nitrous oxide reductase (NosZ) is present (Sanford *et al.*, 2012). Interestingly, ammonium was not the major product of nitrate reduction in axenic cultures amended with ferric iron, and nitrate was predominantly reduced to N_2O in an abiotic reaction with ferrous iron. The N_2O formed was subsequently reduced by the activity of the Clade II NosZ. Even though *A. dehalogenans* lacks *nirS* or *nirK*, the organism contributes to complete denitrification through a combination of direct enzymatic and coupled biotic-abiotic reactions. The prevalence of ferric iron minerals and iron-reducing bacteria in soils suggest that chemodenitrifiers have relevant contributions to N turnover and N_2O flux.

Fungi are key contributors to carbon (C) cycling in soils and their role in C turnover is well established; however, the broader roles of fungi for soil N turnover remain largely unexplored. Recent evidence suggested a significant role for dominant soil fungi in denitrification and associated N_2O production (Chen *et al.*, 2014). Fungi are distinguished from their denitrifying bacterial counterparts due to the formation of N_2O as the major end product of denitrification, and mounting evidence suggests that fungal denitrification could be a major source of N_2O in soils. Fungi possess a unique gene, *p450nor*, encoding an approximately 44 kDa cytochrome P450 protein that catalyzes the reduction of NO to N_2O by direct electron transfer from NADH or NADPH (Shoun *et al.*, 2012). Cultivation-based approaches to assess the fungal contribution to N_2O production in soils are highly biased, and culture-independent tools for detecting fungal denitrifiers are desirable. Therefore, we designed novel PCR primer sets targeting the fungal *p450nor* gene. Amplification of *p450nor* from DNA of 37 denitrifying fungal isolates validated the approach and application to agricultural soil yielded 23 unique *p450nor* amplicons (Higgins et al., 2016). Phylogenetic analysis demonstrated monophyly and provided insights into the taxonomic diversity of denitrifying fungi in the soils studied. Interestingly, *p450nor* genes were not detected in metagenomes generated from the same agricultural soil samples, emphasizing the value of the novel PCR-based approach for assessing potential fungal contributions. Collectively, our studies demonstrate that N cycling (and associated C turnover) cannot be predicted based on gene content (e.g., *nirS, nirK*) alone, and chemodenitrifiers may be major contributors to N_2O flux in soils. The new *p450nor*-targeted primer set complements the molecular toolbox for studying N_2O formation in soils, and has broad utility for assessing fungal denitrification activity.

References:

Chen H, Mothapo NV, and Shi W. 2014. The significant contribution of fungi to soil N_2O production across diverse ecosystems. *Appl Soil Ecol* 73: 70-77.

Higgins SA, Welsh A, Orellana LR, Konstantinidis K, Chee-Sanford J, Sanford RA, and Löffler FE. 2016. Detection and diversity of fungal nitric oxide reductase genes (*p450nor*) in agricultural soils. In review.

Sanford RA, Wagner DD, Wu Q, Chee-Sanford JC, Thomas SH, Cruz-García C, *et al.* 2012. Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. *Proc Natl Acad Sci*. USA. 109:19709-19714.

Shoun H, Fushinobu S, Jiang L, Kim S-W, and Wakagi T. 2012. Fungal denitrification and nitric oxide reductase cytochrome P450nor. *Philos Trans R Soc Lond B Biol Sci* 367: 1186-1194.

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

Impact of Agricultural Practices on Nitrogen Cycle Genes and Nitrous Oxide Emissions from Midwestern Soils

Luis Orellana*[1](lhorellana@gatech.edu**), Joanne Chee-Sanford**[2], **Robert Sanford**[3], **Frank Löffler** [4,5], and **Konstantinos Konstantinidis**[1].

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

Project Goals:
The goals of this project are to fill existing knowledge gaps in our understanding of N-flux and associated C-turnover in soils and sediments. Novel information about the diversity, distribution, abundance and gene expression contributing to N-transformation is required to link desirable (i.e., N-retention) and undesirable (i.e., N-loss, such as N_2O emissions) activities with measurable microbial parameters. Linking molecular- and organismal-level information with environmental factors that control N- and C-turnover can predict the impact of land management practices on greenhouse gas emissions. Such integrated approaches generate novel information on multiple scales of resolution and contribute to system-level understanding of key nutrient cycles in soils. In the present work, we analyzed the response of microbial communities to agricultural practices (e.g., addition of N-species) in two agricultural sites, an important soil ecosystem for bioenergy crop production in Midwest US.

Abstract:
Assessing the impact of fertilizer overuse on microbial soil communities is important for a better understanding of the cycling of nitrogen and carbon and the emissions of potent greenhouse gasses. We analyzed short-read metagenomes obtained from two agricultural sites with contrasting soil textures (sandy versus silty-loam) during four seasons in 2012 at two depths: surface (0-5cm) and deep (20-30 cm). The predicted protein-coding sequences recovered in the 16 metagenomes, based on the SEED and Gene ontology databases, revealed a clear separation between surface and deep samples. For instance, genes related to light-dependent stress, DNA repair and nutrient uptake were more abundant (> 2-fold) in the surface samples in both soils (p-adjusted < 0.05). Distinct archaeal populations and nitrogen metabolism genes were characteristic of the deep samples. To overcome the limitations of fixed e-values cut-offs for annotation of short-read metagenomes and to reduce false positive matches, we developed a novel computational approach, called ROCker, that employs the receiver operating characteristic (ROC) curve to minimize the false discovery rate (FDR) based on how simulated shotgun metagenomic reads of known composition map onto well-curated reference protein sequences. ROCker typically showed 60-fold lower false positive rates compared to the common practice of using fixed e-values and hidden Markov models. Application of the ROCker approach to the time series metagenomes showed that most N cycling genes (e.g., *nosZ*, *amoA* and *nirK*, among others) varied in abundance over the course of the year. For instance, we found a remarkably high abundance of metagenomic reads related to the Clade II *nosZ* (reduction of N_2O to N_2) sequences, accounting for approximately 90% of the total *nosZ* reads found in both soil layers. Approximately 12% of the *nosZ* reads were taxonomically assigned to the *Anaeromyxobacter* genus, indicating their potential relevance in N_2O reduction. In addition, six *amoA* (ammonia

oxidation) genes, each encoded by distinct archaeal and bacterial populations, became abundant in the deep sandy samples when seasonal nitrogen fertilization was applied. Population binning allowed the recovery of several draft genomes for novel (at least at the species level) ammonia-oxidizing archaea, bacteria and nitrite oxidizing bacteria. The most abundant populations (ranging from 10 to 35X coverage), related to the *Thaumarchaeota* and *Nitrospira* phyla, were observed to sharply increase in abundance upon N fertilizer application. The activity and abundance of these populations were more closely examined by combining metatranscriptomic, metagenomic and metaproteomics approaches in ^{15}N-labeled ammonia incubations under controlled laboratory settings. This analysis allowed the quantification of the activity of each population, and assessed the strength of each feature (gene, transcript or protein) in explaining ammonia oxidation and N_2O emission rates. Collectively, our study identified key microbial populations and genes responding to seasonal and human-induced perturbations (e.g., fertilization) and controlling the fate of nitrogen in agricultural soils. It also advanced the molecular toolbox for studying N cycling in soils, and is applicable to other important environmental processes.

References:

1. Robert A. Sanford, Darlene D. Wagner, Qingzhong Wu, Joanne C. Chee-Sanford, Sara H. Thomas, Claribel Cruz-García, Gina Rodríguez, Arturo Massol-Deyá, Kishore K. Krishnani, Kirsti M. Ritalahti, Silke Nissen, Konstantinos T. Konstantinidis, and Frank E. Löffler. 2012. **Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils.** Proc. Natl. Acad. Sci. U. S. A. **109**: 19709–14.

2. Orellana, L. H., L. M. Rodriguez-R, S. Higgins, J. C. Chee-Sanford, R. A. Sanford, K. M. Ritalahti, F. E. Löffler, and K. T. Konstantinidis. 2014. **Detecting nitrous oxide reductase (*nosZ*) genes in soil metagenomes: method development and implications for the nitrogen cycle**. mBio **5**: e01193-14.

3. Orellana, L. H., L. M. Rodriguez-R and K. T. Konstantinidis. **ROCker: a pipeline for accurate detection and quantification of target genes in short-read metagenomic datasets**. *Under review.*

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

Nitrous Oxide Source Identification from a Fertilized Bioenergy Crop Soil

Kelley A. Meinhardt,[1] Stuart E. Strand[1*] (sstrand@uw.edu), Manmeet W. Pannu,[1] Steven C. Fransen,[2] Karen L. Casciotti,[3] and **David A. Stahl**[1]

[1]University of Washington, Seattle, Washington; [2]Washington State University, Prosser, Washington; [3]Stanford University, Stanford, California

Project Goals: Over 60% of global anthropogenic nitrous oxide (N_2O), a significant greenhouse gas, originates from fertilized agricultural soils. The major goals of this project were: 1) to resolve the relative contributions of ammonia-oxidizing microbes to nitrification and the production of N_2O from a fertilized soil growing an important bioenergy crop, 2) to determine factors that stimulate N_2O production, and 3) identify methods to suppress N_2O production in such systems. This knowledge would allow for the implementation of more sustainable agricultural practices that would reduce negative environmental consequences currently facing large-scale bioenergy production.

Nitrous oxide (N_2O) is a powerful greenhouse gas that is largely a consequence of the extensive use of synthetic nitrogen (N) fertilizer in agricultural practices. As the global population continues to grow, there will be an increasing demand for food, livestock, and bioenergy, requiring even further intensification of agricultural practices, N fertilizer use, and resulting N_2O emissions. The processes responsible for most N_2O production are microbially-controlled, and under aerobic conditions the controlling populations are ammonia-oxidizing archaea (AOA) and bacteria (AOB). However, the extent to which each group contributes to N_2O production is not known, with published results often varying.

AOA and AOB population dynamics and activities in soils planted with the bioenergy crop switchgrass (*Panicum virgatum*) were characterized both in the laboratory and experimental field studies to resolve the influence of soils, soil chemistry, and management practices on N_2O production. The relative contribution of each group was constrained by laboratory studies combining transcription, isotope fractionation, and selective inhibitor analyses. Real-world significance of the laboratory experiments was confirmed in a field study. The results, described below, provide knowledge that is central to mitigating agricultural greenhouse gas emissions.

Soil from switchgrass fields that previously received synthetic N amendments (fertilized) or no amendment (control) were used to establish a microcosm experiment, where new synthetic N was applied to the fertilized soil, and gas and soil samples were collected over the course of 10 days. The consumption of ammonia and balanced production of nitrate in the soils confirmed activity of ammonia-oxidizers during the experiment. Functional gene counts differed significantly between the two treatments for both ammonia oxidizers, with AOB gene counts increasing 45-fold after 10 days. Conversely, AOA gene counts increased only 1.5-fold after 10 days. Both AOA and AOB transcripts in the fertilized soil increased during the experiment, however AOB transcripts increased over 85-fold from day 0 to 10. Total N_2O and total natural abundance ^{15}N ($\delta^{15}N^{bulk}$) from the gas samples were measured on an isotope ratio mass spectrometer (IRMS). Total N_2O increased beginning on day 5, and reached a 75-fold difference

at day 7 from day 0 concentrations. Another 2.5-fold increase was observed from day 7 to 10. Relatively little N_2O was produced in the unfertilized control soil microcosms, which retained a $\delta^{15}N^{bulk}$-N_2O signature between -7 to +3 per mil. The $\delta^{15}N^{bulk}$-N_2O from fertilized microcosms transitioned from values of -10 to -49 per mil during the course of incubation, indicating a shift in N_2O production by AOA to AOB. Typical $\delta^{15}N^{bulk}$-N_2O values from soil AOA range from -11 to -34 but can be as low as -68 per mil in cultures of AOB. Keeling plots indicated the major source of N_2O had a $\delta^{15}N^{bulk}$ value of -45.5 per mil, representing a fractionation of -45 per mil relative to the NH_4^+ supplied. The distinct increase in AOB abundance and activity paired with rising N_2O stemming from a single source within the fractionation range of AOB provided strong evidence that AOB dominated N_2O production in these fertilized soils.

A second microcosm experiment was established to confirm the above results, as well as test the effectiveness of nitrification inhibitors on reducing N_2O emissions from fertilized soils. Four different treatments were applied to the same soils types (control and fertilized) used in the first experiment: 1) acetylene to inhibit both AOA and AOB, 2) PTIO to selectively inhibit AOA, 3) 1-octyne to selectively inhibit only AOB, and 4) no inhibitor (positive control). Soil and gas samples were collected over a 10 day period. No decrease in soil ammonia or increase in nitrate or N_2O in the acetylene-treated soil indicated that both ammonia oxidizers had been successfully inhibited. The 1-octyne treated soils produced nitrate and N_2O at rates similar to those from the acetylene treatment, suggesting that AOA produced very little of these compounds. The no inhibitor control showed a balanced consumption of ammonia and production of nitrate, while N_2O production was high. Importantly, the PTIO treatment showed slightly higher nitrate and N_2O production than the no inhibitor control treatment. This suggested that inhibition of ammonia oxidation by the AOA served to direct more ammonia to the AOB and again indicated AOB dominated N_2O production in these fertilized soils.

To evaluate the real-world significance of these findings, a field study was executed to measure AOA and AOB abundance in relation to N_2O flux. Switchgrass plots receiving synthetic N fertilizer or no amendment (control) were sampled for soil and gas flux in April, July, and September 2013. The results were consistent with laboratory findings. The abundance of AOB correlated with ammonia application and seasonal N_2O flux, implicating AOB as the major producers. There was no difference in AOA abundance between the two treatments. Other environmental factors, such as temperature, seemed to be a larger driver of their abundance in the field.

The anticipated increase in N_2O emissions associated with agriculture is a great challenge for society. The results of our studies suggest that management practices that promote AOA and suppress AOB will help reduce microbial production of this atmospherically active gas.

This research was supported by the Department of Energy's Genomic Science and Technology for Energy and the Environment grant DE-SC0006869.

Anthropogenic N deposition slows decay by favoring bacterial metabolism: Insights from metagenomic analyses

Zachary Freedman[1]* (zacf@umich.edu), Rima Upchurch,[1] **Donald Zak**,[1] and Lauren Cline[2]

[1]University of Michigan, Ann Arbor; [2]University of Minnesota, St. Paul

https://sites.google.com/a/umich.edu/dr-don-zak-s-lab/

Project Goals: With support from DoE BER, we have been able to study the molecular and microbial mechanisms by which experimental N deposition has decreased plant litter decay and increased the accumulation of organic matter in a northern hardwood forest ecosystem located in Michigan, USA. Experimental plots have been exposed of increased NO_3 deposition for 20 years at a rate expected by 2050 across some portions of eastern North America. Using biogeochemical and molecular analyses, we have been able to test our hypotheses that the ecosystem response to experimental N deposition (*i.e.*, greater soil C storage) is governed by the environmental regulation of microbial gene expression. Here, we present findings that suggest saprotrophic bacteria, which degrade organic matter less efficiently than their fungal counterparts, may be favored by future rates of N deposition.

Litter decomposition is an enzymatically-complex process that is mediated by a diverse assemblage of saprophytic microorganisms and can be suppressed by anthropogenic N deposition. In a northern hardwood forest ecosystem located in Michigan, USA, 20 years of experimentally increased atmospheric N deposition has reduced forest floor decay and increased soil C storage. Here, we paired extracellular enzyme assays with shotgun metagenomics to assess if chronic N deposition has altered the functional capacity of microbial communities inhabiting decaying forest floor. Experimental N deposition significantly reduced the activity of extracellular enzymes mediating plant cell wall decay, which occurred concurrently with changes in the relative abundance of metagenomic functional gene pathways. Experimental N deposition altered the relative abundance of 60 functional gene pathways mediating the metabolism of carbohydrates, aromatic compounds, as well as respiration. Moreover, experimental N deposition increased the relative abundance of 50 of the 60 gene pathways, the majority of which were associated with saprotrophic bacteria. Conversely, the relative abundance and composition of fungal genes mediating the metabolism of plant litter was not affected by chronic N deposition. Future rates of atmospheric N deposition have favored saprotrophic soil bacteria, whereas the metabolic potential of saprotrophic fungi appears resilient to this agent of environmental change.

Funding through DOE's Office of Biological and Environmental Research program.

Linking Microbial Community Structure, Activity and Carbon Cycling in Biological Soil Crust

Tami L. Swenson[1]* (tlswenson@lbl.gov), Ulas Karaoz[1], Rebecca Lau[1], Richard Baran[2], **Trent Northen[1,3]**

[1]Lawrence Berkeley National Laboratory, [2]present address: Thermo Fisher Scientific, [3]DOE Joint Genome Institute

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

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

Soils play a key role in the global carbon cycle, but the relationships between soil microbial communities and metabolic pathways are poorly understood. Our overall aim is to develop soil metabolomics methods and statistical models to link active microbes to the abundance and turnover of soil metabolites and examine the detailed substrate and product profiles of individual soil bacteria. To achieve these goals, we are using two different soil systems- biological soil crusts (biocrusts) and grassland soils. Biocrusts are communities of organisms inhabiting the upper layer of soil in arid environments. The crust itself is essentially microbial exopolysaccharide linked sand particles and is critical to soil stabilization. Biocrusts persist in a dessicated dormant state for extended periods with rare pulsed activity events following precipitation. *Microcoleus vaginatus*, a non-diazotrophic filamentous cyanobacterium, is the key primary producer in bacterially-dominated biocrusts in the Colorado Plateau and is an early pioneer in colonizing arid environments. Over decades, biocrusts proceed through developmental stages with increasing complexity of constituent microorganisms and macroscopic properties. Since *Microcoleus vaginatus* does not fix nitrogen, metabolic interactions with other biocrust microorganisms in the *Microcoleus vaginatus*-associated 'cyanosphere' presumably play a key role in the cycling of soil organic matter and in determining biocrust community dynamics.

To develop soil metabolomics approaches, a series of extractants (aqueous vs. organic solvents) were compared with soils that were either fumigated with chloroform vapor (to release metabolites from microorganisms) or left unfumigated. Analysis by gas chromatography-mass spectrometry and liquid chromatography-mass spectrometry revealed a breadth of metabolites including sugars, sugar alcohols, amino acids, fatty acids, dicarboxylic acids, sterols, nucleobases and osmolytes. Fumigation prior to extraction had a significant effect on the range and intensity of most metabolites and water was one of the most effective extractants. The inclusion of organic solvent (methanol) facilitated the extraction of fatty acids and sterols.

Exometabolite profiling was used to investigate the utilization of soil metabolites by sympatric bacterial isolates from biocrust (Baran *et al*, 2015). From this we found that *Microcoleus vaginatus* releases a broad range of metabolites. Many of these metabolites were found to be uptaken by heterotrophs but there were surprisingly few metabolites uptaken by all bacteria. This points to competition for a small set of central metabolites and specialization of individual heterotrophs towards a diverse pool of available organic nutrients. We are now

extending these studies to intact soil communities. Specifically, our soil metabolomics methods are being used to analyze the correlations between community structure, activity and soil metabolite dynamics following a laboratory pulsed activity (wetting) event. Biocrusts were wetup with water and metabolites (from porewater) and DNA were extracted at various timepoints up to 49.5 hours post-wetup. Exometabolite analysis revealed a similar breadth of metabolites as soil extracts. In general, many metabolites (*e.g.* amino acids) immediately increased in abundance following wetup and then steadily decreased. However, a few continued to increase over time (*e.g.* xanthine). Interestingly, we have observed xanthine to be released by some *Bacilli sp.* isolated from the biocrust (webofmicrobes.org) and metagenomics and metatranscriptomics show that members of the Paenibacillaceae family increase in abundance in late wetup samples. Previous 16S amplicon data also show a “Firmicutes bloom” following wetup with the new metagenomic data resolving this at genome-level. Ultimately, these approaches will provide an important complement to sequencing efforts linking soil metabolites and soil microbes to enable genomic sciences approaches for understanding and modeling soil carbon cycling.

References

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

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

Linking microbes to soil metabolism

Kirsten S. Hofmockel[1,2*] (kirsten.hofmockel@pnnl.gov), Racheal Erb[1], Sheryl Bell[1], Adina Howe[1], Will Christer[2], Galya Orr[2]

[1]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA;[2]Iowa State University, Ames, IA

*kirsten.hofmockel@pnnl.gov

Project Goals: Our project (Microbial drivers of global change at the aggregate scale: linking genomic function to carbon metabolism and warming) works towards a basic understanding of the microbial ecology that regulates cellulose decomposition in grassland soils. We seek to identify the functional traits and community interactions that are responsible for the decomposition of root-derived C and the implications this has for long term soil C storage and release.

Abstract text. Grasslands are a critical yet understudied component of land-atmosphere C exchange. Globally, grasslands represent ca. 30% of global land area and terrestrial net primary productivity (NPP)(Chapin et al. 2002), contain ca. 20% of the world's soil C (Schlesinger 1977, FAOSTAT 2009), and store the majority of their C belowground in root biomass (Neely et al. 2009). Yet, little is known about the microbial pathways regulating cellulose decomposition under field conditions due to the extreme microbial diversity, and the temporal, chemical, and structural complexity of soil ecology. We seek to identify the underlying molecular biology and community ecology that regulates cellulose decomposition in grassland soils.

We examined the microbial interactions and metabolic functions involved in the degradation of cellulose in field experiments, lab incubations, enrichment cultures and modeling experiments. To identify key organisms and enzymes involved in root decomposition, we used fluorescently labeled cellulose nanocrystals (Grate et al) in combination with metatranscriptomics in controlled enrichment experiments. We demonstrate that the community level responses (OD, respiration, 16S amplicons) were not significantly different when soil communities were grown with cellulose or fluorescently labeled cellulose nanocrystals over a 10d incubation, suggesting our experimental platform can be used to track the fate of cellulose through soil decomposer communities. Our results reveal that Proteobacteria, which represent 10% of the native soil community, dominated the community in our cellulose degrading experiments, comprising 32% of the enrichment culture and 80% of the community by day 10. Pseudomonadales were the most abundant, increasing from 2% of the community to greater than 50% during the cellulose enrichment experiment. In contrast Firmicutes comprised 40% of the native soil community, but reduced to 27% by the end of the incubation experiment. Bacillales were the dominant order and decreased from 69% of the community in the soil inoculum down to 20% during the incubation. Based on metatranscriptomic analyses we found that transcripts associated with cellobiosidase (CB; 3.2.1.91) increased in abundance over the experiment, with no significant differences in to

β-1,4-glucosidase (BG; 3.2.1.21) and β-1,4-xylosidase (BX; 3.2.1.37). Organisms assimilating the fluorescently labeled cellulose will be analyzed in conjunction with metatranscriptomic data to understand the basic ecology regulating cellulose decomposition in diverse soil communities. By targeting extracellular enzymes typically measured in biogeochemical field studies and organisms native to the soil community, we aim to link reduced lab studies to field approaches and predictive models.

References

1. Chapin, F. S., III, P. A. Matson, and H. A. Mooney. 2002. Principles of Terrestrial Ecosystems. Springer-Verlag.
2. FAOSTAT. 2009. Statistical Database 2007.
3. Grate JW, Mo KF, Shin Y, Vasdekis A, Warner MG, Kelly RT, Orr G, Hu D, Dehoff KJ, Brockman FJ, Wilkins MJ. Alexa Fluor-Labeled Fluorescent Cellulose Nanocrystals for Bioimaging Solid Cellulose in Spatially Structured Microenvironments. Bioconjugate chemistry. 2015 Mar 2;26(3):593-601.Neely, C., S. Bunning, and A. Wilkes. 2009. Review of evidence on drylands pastoral systems and climate change: Implications and opportunities for mitigation and adaptation. Land and Water Discussion Paper. Food and Agriculture of the United Nations, Rome, Italy.
4. Schlesinger, W. H. 1977. Carbon balance in terrestrial detritus. Annual Review of Ecology and Systematics 8:51-81.

This research was supported by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research under Contract No. DESC0010775.

Metagenomics-Enabled Predictive Understanding of Microbial Communities to Climate Warming: Results from Long Term Soil Incubations and Modeling Simulations

Elaine Pegoraro[1], Junyi Liang[4], **Edward A.G. Schuur[1]*** (Ted.Schuur@nau.edu), Rosvel Bracho[2], Susan Natali[3], Ji Chen[4], Chang Gyo Jung[4], Wenting Feng[4], Mengting Yuan[4], Xue Guo[4], Lauren E. Hale[4], **Liyou Wu**[4], Jiajie Feng[4], Cong Wang[4], Xishu Zhou[4], **Zhili He**[4], Jim Cole[5], **James M. Tiedje**[5], **Konstantinos Konstantinidis**[6], Katherine Todd-Brown[4], **Yiqi Luo**[4] **and Jizhong Zhou**[4]

[1]Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ; [2]School of Forest Resources & Conservation, University of Florida, Gainesville, FL; [3]Woods Hole Research Center, Falmouth, MA; [4]Institute for Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, Norman, OK; [5]Center for Microbial Ecology, Michigan State University, East Lansing, MI; [6]Center for Bioinformatics and Computational Genomics and School of Biology, Georgia Institute of Technology, Atlanta, GA.

http://ieg.ou.edu

Project goal: The overall goal of this project is to advance system-level predictive understanding of the feedbacks of belowground microbial communities to multiple climate change factors and their impacts on soil carbon (C) cycling processes. The main objectives of this integrative project are to (i) determine the responses of microbial community structure, functions and activities to an increased input of easily decomposable C substrates to soil (priming effects); (ii) determine the extent to which priming enhances mineralization of native soil C; (iii) determine what proportion of the increased mineralization of native soil C is old C; (iv) determine if substrate input with different C quality distinctively affects microbial activity and soil organic matter decomposition; and (v) develop integrated bioinformatics and modeling approaches to scale information across different organizational levels. This study focuses on using laboratory incubations of soil as an isolated system to understand the influence of microbial processes on the release of C, and their response to changes in easily decomposable C substrate inputs.

Warming of tundra ecosystems due to climate change is predicted to thaw permafrost and increase plant biomass and litter input to soil. Additional input of easily decomposable carbon (C) can alter microbial activity by providing a much needed energy source to microbes, thus accelerating soil organic matter decomposition. This phenomenon, known as the priming effect, can increase CO_2 flux from soil to the atmosphere; however, the extent to which it could decrease soil C stocks in the Arctic is unknown. This project investigates priming effects on permafrost soil. We hypothesized that priming would increase and change microbial activity and composition, thus increasing mineralization of old and slowly decomposing C. We are conducting a long-term (> 1 year) incubation experiment that started in July 2015. Soil cores were collected in 2013 from a moist acidic tundra site in Healy, Alaska, from surface (0-15 and 15-25 cm) and deep permafrost layers (45-55, 65-75, and 75-85 cm). Samples were incubated aerobically, at 15°C. We amended soil samples with uniformly ^{13}C labeled glucose and cellulose to quantify changes in C mineralization rates attributable to the added substrates. Carbon dioxide flux and $^{13}CO_2$ measurements were coupled and measured every 24 hours for the first 5 days, every 35 hours for 8 days, and every couple of weeks for 4 months. We also sampled $^{14}CO_2$ at days 0, 15, and 105 of incubation to identify the age of respired C. Data shows that substrate

additions resulted in higher respiration rates in amended soils; however, priming was only observed in deep layers amended with glucose, where on average 22%, 29%, and 10% more soil C was respired at 45-55, 65-75, and 75-85 cm, respectively. This suggests that microbes in deep layers are limited in energy due to greater fractions of slowly decomposing C; therefore, additional input of easily decomposable C increases native organic matter decomposition. Glucose and cellulose will be added every 4 months to simulate field input of root exudates, new root biomass, and dissolved organic C leachate, and the same measurements will be performed until Fall 2016. Microbial composition, structure, and dynamics will be measured during the second substrate amendment at days 7, 15, and 65 of incubation using two techniques: the functional gene structure analysis, GeoChip and phylogenetic composition using 16S rRNA gene sequencing on Miseq.

In the same time, model-data assimilation is used to explore effects of possible microbial activity change by long-term field warming on soil C dynamics. Soils were sampled from a long-term warming experiments. In each treatment of both control and warming, two soils, bulk soil and soil from a deep collar, which was inserted in 2000, were sampled. Totally, there are four field treatments, bulk soil at control, deep collar at control, bulk soil at warming, deep collar at warming. Each treatment has 6 replicates. After pre-treatment in lab, the soil samples were incubated at 15 ℃ for 406 days and 25 ℃ for 365 days. A three-pool model, with active, slow, and passive pools, is used to simulate the soil C dynamics during incubation. A Markov Chain Monte Carlo (MCMC) technique is applied to do the model-data assimilation. Results show that CO_2 emission is different among treatments, indicating long-term field warming may change soil microbial activity, which determines soil C dynamics. In addition, long-term field warming significantly increased the decay rates of slow and passive pools. Our results suggest that the changed microbial activity by long-term field warming can accelerate the decomposition of soil C pools with relatively long residence time, which can potentially affect the feedback to long-term climate change. During the incubation, microbial composition, structure, and dynamics were measured at 2 weeks, 3 months, and 9 months using GeoChip. We are analyzing these data to explore the effects of the functional genes on soil C dynamics.

This work is supported by the US Department of Energy, Biological Systems Research on the Role of Microbial Communities in Carbon Cycling Program (DE-SC00010715).

Metagenomics recovers 100s of population genomes from Alaskan permafrost and Oklahoma prairie soils and provides insights into their roles in microbial community response to warming.

Konstantinos T. Konstantinidis[1*] (kostas@ce.gatech.edu), Eric R. Johnston[1], C. Ryan Penton[2], Julian Yu[2], **Liyou Wu**[3], Shi Zhou[3], Xue Guo[3], Ping Zhang[3], **Zhili He**[3], Mengting Maggie Yuan[3], **Yiqi Luo**[3], **Edward A.G. Schuur**[4], James R. Cole[5], **James M. Tiedje**[5], **Jizhong Zhou**[3].

Georgia Institute of Technology, Atlanta, GA 30332, USA [1]; Arizona State University, Phoenix, AZ 85004, USA [2]; University of Oklahoma, Norman, OK 73019, USA [3]; Northern Arizona University, Flagstaff, AZ 86001, USA [4]; Michigan State University, East Lansing, MI 48824, USA [5]

Project goals: The overall goal of this project is to advance system-level predictive understanding of the feedbacks of belowground microbial communities to multiple climate change factors and their impacts on soil C cycling processes. Regarding this goal, we are pursuing the following objectives: **(i)** To improve our understanding of soil microbial communities indigenous to temperate and tundra ecosystems through whole-community analysis, and through the classification of novel taxa recovered directly from metagenomes and/or obtained in pure culture; **(ii)** To determine the microbiological basis underlying temperature sensitivity of soil organic matter decomposition; (**iii)** To determine the ubiquity of recovered bacterial populations and genes across large geographic regions spanning several hundred kilometers; and (**iv**) To develop integrated bioinformatics and modeling approaches to scale information across different organizational levels towards predictive understanding of ecosystem responses to multiple climate change factors, which will be collaborated and integrated with the K-Base.

Abstract: Under this project, we have begun investigations on microbial communities from Alaskan tundra permafrost (AK) and Oklahoma temperate grassland (OK) soils, both of which have been experimentally warmed 2 to 4 ℃ above ambient temperature *in-situ.* Well-replicated whole-community shotgun metagenomic sequencing of soils collected after 1 and 5 years of warming yielded near-complete representation of microbial community 'sequence richness' at AK and OK sites. A custom-made assembly and contig binning strategy has allowed for the recovery of many near-complete bacterial population genomes from both locations. In particular, populations recovered from AK soils collectively made up to ~15% of the total microbial community. These genomes represented diverse taxonomic groups and metabolic lifestyles tuned toward sulfur cycling, hydrogen metabolism, methanotrophy, and organic matter oxidation. While short-read analysis of soil metagenomes collected after 1-year of warming revealed small shifts in pathways related to SOM-decomposition (Xue et al., *in press*), recently-sequenced metagenomes collected after 5-years of experimentation revealed dominant bacterial populations shifting in abundance by as much as 80% in response to the warming treatment (Johnston et al., *in preparation*). Further, several bacterial populations

recovered from AK tundra soils were also present and/or dominant in geographically distant (~100-530 kilometers apart) tundra habitats (full genome representation and >98% genome-derived average nucleotide identity). Therefore, their relative contribution to various ecosystem functions is expected to be high and their individual responses to climate warming may be of significance to large geographic regions.

In addition to studying the *in-situ* response to warming, we have also incubated soil taken from both sites and two different depths under elevated temperatures in the laboratory for 3 years. By combining shotgun metagenomic sequencing with respiration data and soil indices, we hope to gain a more detailed view of the soil taxa, and the underlying mechanisms, modulating responses to warming and the activities responsible for greenhouse gas release. To this end, we have assembled >200 draft genomes, most of which represent previously uncharacterized (novel) taxa, and collectively making up ~50% of the incubation metagenomes. We will report on the significant differences in the abundance of these genomes over time in the laboratory incubations as well as the differences between soil ecosystems and depths. To further support this work and enable testing of the emerging hypotheses from comparative metagenomics, we have been implementing cultivation on dilute nutrient, minimal salt and soil extract media under reduced oxygen stress. A total of 660 culture plates under varying salt conditions (0-10% NaCl) and dilutions (10^{-6} – 10^{-8}) were cultivated for >8 weeks from Alaskan shallow (<55 cm) and deep (>55 cm) soils at 25°C. The majority of colonies appearing after 2 weeks remained small (<2 mm diameter), with varying colony morphologies. Our current workflow to identify target cultures for physiological characterization and sequencing involves amplification and identification of 16S rRNA genes in order to relate each culture to the metagenomic binning results.

References

1. C. Luo, L. M. Rodriguez-R, E. R. Johnston, L. Wu, L. Cheng, K. Xue, Q. Tu, Y. Deng, Z. He, J. Zhou Shi, M. M. Yuan, R. A. Sherry, D. Li, Y. Luo, E. A. G. Schuur, P. Chain, J. M. Tiedje, J. Zhou, and K. T. Konstantinidis. Soil Microbial Community Responses to a Decade of Warming as Revealed by Comparative Metagenomics. Applied & Environ. Micriobiol. 03712-13. 2014.

2. E. R. Johnston, L. M. Rodríguez-R, C. Luo, M. M. Yuan, L. Wu, Z. He, E. A. Schuur, Y. Luo, J. M. Tiedje, J. Zhou, and K. T. Konstantinidis. Metagenomics reveals pervasive bacterial populations and reduced community diversity across the Alaska tundra ecosystem. Frontiers in Microbiology. *Under Review*.

3. K. Xue, M. M. Yuan, Z. J. Z. J. Shi, Y. Qin, Y. Deng, L. Cheng, L. Wu, Z. He, J. D. Van Nostrand, R. Bracho, S. Natali, E. A. G. Schuur, C. Luo, K. T. Konstantinidis, Q. Wang, J. R. Cole, J. M. Tiedje, Y. Luo, and J. Zhou. Rapid microbial feedbacks reveal vulnerability of tundra soil carbon to climate warming. Nature Climate Change. *In Press*.

Funding statement: Our work is supported by the U.S. DOE Office of Science, Biological and Environmental Research Division (BER), Genomic Science Program, Award No. DE-SC0010715.

Successional Dynamics of Grassland Microbial Communities in Response to Warming, Precipitation Alternation, and Clipping

Xue Guo[1], Mengting Yuan[1], Zhou Shi[1], Xishu Zhou[1], Jiajie Feng[1], Cong Wang[4], **Liyou Wu[1]*** (lwu@rccc.ou.edu), **Zhili He**[1], Chang Gyo Jung[4], Joy D. Van Nostrand[1], Lauren E. Hale[1], Qichao Tu[1], Jie Deng[1], Jianjun Wang[1], Zhenxin Li[1], Tong Yuan[1], **Jim Cole**[2], **James M. Tiedje**[2], **Edward A.G. Schuur**[3], **Konstantinos Konstantinidis**[4], **Yiqi Luo**[1], and **Jizhong Zhou**[1]

[1]Institute for Environmental Genomics and Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK; [2]Center for Microbial Ecology, Michigan State University, East Lansing, MI; [3]Center for Ecosystem Sciences and Society, Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ; [4]Center for Bioinformatics and Computational Genomics and School of Biology, Georgia Institute of Technology, Atlanta, GA http://ieg.ou.edu/

Project goal: The overall goal of this project is to advance systems-level predictive understanding of the feedbacks of belowground microbial communities to multiple climate change factors and their impacts on soil carbon (C) cycling processes. The specific objectives are to: (i) reveal the responses of soil microbial communities to climate warming and soil moisture alteration in both tundra and temperate grassland ecosystems; (ii) determine temperature sensitivity of recalcitrant C decomposition and characteristics of the microbial degraders; and (iii) develop integrated bioinformatics and modeling approaches to scale information across different organizational levels.

As a part of the integrated project, here we present: i) results from field experiments in a temperate grassland located in central Oklahoma established in 2008 to reveal the influence of elevated temperature, altered precipitation and plant biomass clipping on long-term and seasonal succession of plant and microbial communities, and ii) a primer design tool for highly parallel qPCR analysis of microbial communities.

Long-term succession of plant and microbial communities. To understand the long-term successional dynamics of microbial communities in response to warming, clipping and alternated precipitation, the 264 annual soil samples were analyzed by sequencing of 16S rRNA gene and ITS, and a functional gene array (GeoChip 5.0). Dissimilarity and permutational analyses of variance indicated that the three treatments all significantly ($P<0.05$) affected the structure and functions of the microbial communities. Warming was the most influential factor on bacterial and fungal communities over time. Interestingly, a warming effect started to be significant ($P<0.05$) from the third year of operation on bacterial communities, while it was significant ($P<0.05$) from the very first year on fungal communities. Species richness values (Shannon index) of warmed microbial communities were similar in 2009 and 2010, but gradually decreased ($P<0.05$) in the last three years whereas the control plots remained unchanged. The community composition was altered over time with *Actinobacteria* and *Firmicutes* ($P<0.05$) becoming more abundant, and *Proteobacteria*, *Bacteroidete*s, and *Acidobacteria* showing reduced abundances in warmed samples in 2014, compared with those in 2009 and 2010. In the first two years, warming increased ($P<0.05$) the relative abundance of genes involved in C degradation, nitrogen (N) cycling and phosphorus (P) utilization, while in the third year, warming had no significant effects on these genes. Then, in the most recent two years, these genes decreased in response to the warming treatment ($P<0.05$). Precipitation alternations significantly ($P<0.05$) affected the phylogenetic compositions of bacterial and fungal communities, but not their functional gene structures. Annual clipping significantly ($P<0.05$) changed some bacterial and fungal populations, such as *Actinobacteria*, *Bacteroidetes*, *Ascomycota*, *Zygomycota*. Cumulative annual clipping effects on functional genes were observed over time. From the second year of operation, significant ($P < 0.05$) increases of the relative abundances of genes involved in degradation of both labile and recalcitrant C were observed. However, in the last two years, only the genes involved in the degradation of recalcitrant C increased ($P<0.05$). Other

genes involved in nutrient-cycling processes including N cycling and P utilization were also increased ($P<0.05$) by annual clipping. Soil microbial communities under different treatments had different ($P<0.05$) temporal turnover rates, such as species-time relationships (STR) and time-decay relationships. The temporal turnover rates of bacterial communities were stimulated by warming, but decreased by double precipitation, and not significantly changed by clipping. The temporal turnover rates of fungal communities were accelerated by warming, half and double precipitation, but decreased by clipping. These results indicate that warming, altered precipitation, and clipping have differential effects on the diversity, composition, and structure of soil microbial communities over time.

Short-term microbial succession. Surface (0-15cm) soil samples were collected monthly during 2012 from both warmed and control plots, and were analyzed by sequencing of 16S rRNA genes and GeoChip 5.0. Soil respiration and geochemical properties were also measured to link the soil microbial community structures with environmental factors. Both warming and sampling month significantly ($P < 0.005$) affected the soil microbial taxonomic groups and functional genes. *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Verrucomicrobia,* and α- and δ-*Proteobacteria* had significant monthly variations under warming, while only *Gemmatimonadetes*, δ-*Proteobacteria* and *Chloroflexi* showed monthly changes in the control. This higher temporal divergence of warmed communities was confirmed by a higher taxonomic β-diversity (Sørensen index, $P = 0.02$) of warmed samples compared with the control. Network analysis indicated the network for warmed bacterial communities exhibited more links, a higher average degree, and a higher average clustering coefficient than that of control communities. The functional gene intensities also showed a significant interactive effect of sampling month and warming, in which the C degradation genes tended to increase, but the C fixation and N cycling genes tended to decrease during peak plant biomass months (Apr. to May and Sept. to Oct.). Together, these results revealed a higher temporal variation of soil microbial communities related to seasonal succession in a warmer environment. The species-time relationship exponent (STR-w) was slightly higher in warmed than in control plots, suggesting a faster species accumulation over time in response to warming. These exponents were higher than most reported values from long-term studies on soil microbial communities, suggesting a quicker species accumulation under short term than long-term succession.

Efficient high-throughput primer design tool for highly parallel qPCR. To provide a new quantitative and high throughput microbial community analysis approach, over the past year, by utilizing our FunGene database and repository (http://fungene.cme.msu.edu), we have been developing and testing an efficient high-throughput primer design tool. Protein-coding genes are, in general, less conserved than structural RNA genes, meaning that often no single probe or primer pair is able to target a gene's full range of diversity. Our tool helps with the design of multiple primers from potentially large reference sets of 3,000 sequences or greater. We cast the problem as a variant of the well-known "maximum coverage problem" from computer science. Since this problem has no practical exact solution, we use a "greedy" algorithm to choose a set of primer pairs from the candidates that maximizes the diversity covered by the primer sets. During tool testing, we developed new primer sets for nitrogen cycling genes (amoA, nifH), recalcitrant carbon degradation genes (*cutC*, *cntN*), antibiotic resistance genes (tet_sul2, tetA-G), an integrase gene involved in mobile elements (intI1). We have experimentally validated a set of three non-degenerate primer pairs targeting *cntN* sequences. The sequencing results showed adequate sensitivity, satisfactory amplicon size, and 99% PCR efficiency with the three primers used. This tool is also being employed in the design of primers targeting ACC deaminase (*acdS*) genes, involved in reducing plant stress, for investigations into the role of disease suppressive soil microbial communities in promoting plant health. The current tool is already in use by several research groups. We have developed the tool to be "KBase ready" and intend to help integrate this functionality both into FunGene and into KBase.

Funding statement: *Our work is supported by the U.S. DOE Office of Science, Biological and Environmental Research Division (BER), Genomic Science Program, Award No. DE‐SC0010715.*

Disentangling the Activity of Anaerobic Methane-Oxidizing Archaea from Their Syntrophic Sulfate-Reducing Bacterial Partner

Hang Yu[1*](hyu@caltech.edu), Grayson L. Chadwick,[1] Connor Skennerton,[1] Silvan Scheller,[1] Andy O. Leu,[2] Gene W. Tyson,[2] and **Victoria J. Orphan**[1]

[1]California Institute of Technology, USA; [2] University of Queensland, Australia

Project Goals: Anaerobic oxidation of methane (AOM) with sulfate is mediated by a syntrophic partnership between anaerobic methane-oxidizing archaea (ANME) and sulfate-reducing bacteria (SRB) and is a major methane sink in the global carbon cycle. Our aim is to gain a better understanding of the mechanisms of electron sharing by these syntrophs using a combination of isotope geochemistry, single-cell microscopy and meta-omics techniques. The overall energy yield of AOM with sulfate is low ((ΔG°' = -17 kJ/mol) and ANME have a doubling time of 3-9 months. With an integration of these approaches, we have begun to disentangle the activity of these microbes that are symbiotic in nature and provided a fundamental understanding to the inner workings of AOM.

A consortium of anaerobic methane-oxidizing archaea (ANME) and sulfate-reducing bacteria (SRB) consumes teragrams of methane in marine ecosystems and therefore are critical players in the global carbon cycle. However, basic mechanisms of the metabolic coupling in this microbial symbiosis remains obscure. Two hypotheses exist on how electrons are transferred in anaerobic oxidation of methane (AOM) with sulfate: 1) Milucka and colleagues[1] propose that both methane oxidation and the initial step of sulfate reduction co-occur in ANME, and zero-valent sulfur is subsequently consumed by SRB via a disproportionation reaction into sulfate and sulfide; 2) methane oxidation in ANME and sulfate reduction in SRB occur independently and are coupled via direct interspecies electron transfer using multi-heme cytochromes or pili without a diffusible intermediate compound, such as zero-valent sulfur[2,3].

We designed microcosm incubations to test these hypotheses and are refining three approaches to track activity of environmental microbes by: 1) developing stable isotope geochemistry tools to more accurately track rates of methane oxidation and sulfate reduction, 2) visualizing single-cell spatial activity patterns that link organism identity to biosynthetic activity using fluorescence *in situ* hybridization coupled to nanoscale secondary ion mass spectrometry (FISH-nanoSIMS), and 3) illuminating the pathways involved in AOM with paired metagenomics and metatranscriptomic studies that decouple this symbiosis in the laboratory and induce differential gene expression between ANME and SRB.

Zero-valent sulfur species were found to be inhibitory to AOM at concentrations higher than 0.1 mM, consistent with the idea of product inhibition. However, methane oxidation did not resume after 5 days, the time for expected consumption of sulfur as a substrate by SRB. This suggests that zero-valent sulfur prompts an unknown toxicity effect rather than being an actively exchanged metabolite in the consortia. Additionally, SRB could not be decoupled from ANME

and grown with zero-valent sulfur amendments in our long-term incubations. Instead, we found a decoupling of this syntrophy using artificial electron acceptors such as anthraquinone-2,6-disulfonate (AQDS) in which ANME were catabolically and anabolically active without sulfate or their SRB partner[4]. The theoretical energy yield of AOM coupled to AQDS ($\Delta G^{\circ\prime}$ = -41 kJ/mol) is higher than that coupled to sulfate ($\Delta G^{\circ\prime}$ = -17 kJ/mol). Accordingly, we found that SRB remained biosynthetically and transcriptionally inactive in the presence of both AQDS and sulfate, possibly due to a lack of methane-derived electrons that likely were shuttled to AQDS more favorably. Furthermore, all sulfate reduction genes including sulfate adenylyltransferases, adenosine-5'-phosphosulfate reductases and dissimilatory sulfite reductases were down-regulated in the metatranscriptome, indicating that pathways of methane oxidation and sulfate reduction are not linked in one organism as proposed[1]. Other genes up-regulated with AQDS included those involved in reverse methanogenesis and membrane electron transport. Overall, the combination of different cellular activity probes developed here provides evidence for direct interspecies electron transfer between ANME and their SRB partner as well as insight into the cellular machineries that facilitate the extracellular electron flow in AOM.

References

1. Milucka, J. et al. Zero-valent sulphur is a key intermediate in marine methane oxidation. Nature 491, 541–546 (2012).
2. McGlynn, S. E., Chadwick, G. L., Kempes, C. P. & Orphan, V. J. Single cell activity reveals direct electron transfer in methanotrophic consortia. Nature 526, 531–535 (2015).
3. Wegener, G., Krukenberg, V., Riedel, D., Tegetmeyer, H. E. & Boetius, A. Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. Nature 526, 587–590 (2015).
4. Scheller, S., Yu, H., Chadwick, G. L., McGlynn, S. E. & Orphan, V. J. Artificial electron acceptors decouple archaeal methane oxidation from sulfate reduction. In press.

This work was supported by the Department of Energy BER program (DE-SC0010574 and DE-SC0004940) and funding by the Gordon and Betty Moore Foundation through Grant GBMF3306 (to VJO).

New Insights into Methane-Oxidizing Communities in Lake Sediments through Microcosm Manipulation and Systems Biology Studies

Ludmila Chistoserdova[1]*, Maria E. Hernandez[1], Sascha Krause[1], Zheng Yu[1], Janet B. Matsen[1], David A.C. Beck[1,3], and **Mary E. Lidstrom**[1,2]

Departments of [1]Chemical Engineering and [2] Microbiology and [3]eScience Institute, University of Washington, Seattle, USA; milachis@uw.edu

Project Goals: This project addresses the structure and function of microbial communities active in methane consumption, using lake sediment as a model. We utilize both top-down and bottom-up approaches toward this goal. The top-down approach involves manipulation of native sediment samples under specific environmental conditions, such as methane/oxygen/nitrogen species availability/relative concentrations. The bottom-up approach employs axenic cultures of model bacteria for constructing synthetic communities of different complexity, from very simple (two-species) to relatively complex (50-species). Through manipulations of these communities, combined with systems biology studies, we are striving to understand the molecular mechanisms that form a basis for specific species interactions in microbial oxidation of methane.

In the current phase of the project, as part of the bottom-up approach, we carried out microcosm incubations under two different oxygen tension regimens ('low' versus 'high'), in multiple replicates, over the course of 14 weeks, with microcosm cultures transferred with dilutions once every week. Samples of DNA and mRNA from week 4 to week 14 have been shotgun sequenced, with 4 replicates for each sampling point, in collaboration with the JGI. We are currently in the process of analyzing this extensive dataset that represents species active in methane oxidation, over time, as well as their respective activities. What is clear from the data so far is the complex nature of communities involved in methane metabolism. These communities are represented, in addition to *bona fide* methantrophs (the *Methylococcaceae* species), by methylotrophic species within the family *Methylophilaceae* and non-methylotrophic species mainly belonging to the order *Burkholderiales* and to the phylum *Bacteroidetes*. Preliminary analysis of the metatranscriptomes indicates that the most highly transcribed genes in the microcosms are the ones encoding methane monooxygenase, the first enzyme in methane oxidation. Genes encoding methanol dehydrogenase, catalyzing oxidation of the immediate product of methane oxidation, are also among the most highly expressed genes in both *Methylococcaceae* and *Methylophilaceae*. *Methylobacter* and *Methylotenera* species in addition express respiratory denitrification functions (incomplete pathway in the former and complete pathway in the latter). The *Burkholderiales* also highly expresses the denitrification functions, along with acetate metabolism functions. The *Bacteroidetes* appear to represent the next tier in the food web, utilizing extracellular polymeric substances produced by *Methylococcaceae* and *Methylophilaceae*. The relationships between the core set of organisms active in methane oxidation are brought to the

next level of complexity by the presence and activity of predatory species, among which *Bdellovibrionales* and *Myxococcales* dominate.

The bottom-up approach provides further insights into the metabolic interconnections between major functional guilds. Comparative transcriptomics of model *Methylobacter* and *Methylotenera* species cultivated either axenically or as parts of two-species stable communities revealed differential expression of specific functions, suggesting their involvement in molecular mechanisms of interspecies metabolic/regulatory interdependence. We observed especially dramatic response of alternative methanol dehydrogenases to the cultivation conditions as follows. While in *Methylobacter* cultivated axenically, the XoxF type (lanthanide-dependent) methanol dehydrogenase was preferentially expressed, in cocultures, the MxaFI type (calcium-dependent) enzyme was preferentially expressed. The *Methylotenera* species that tend to encode multiple XoxF enzymes differentially expressed different variants in axenic versus coculture conditions. The physiological meaning of differential choices of methanol-oxidizing enzymes in both partners is currently being addressed via the analysis of knock-out mutants in respective genes.

While our two-species synthetic community experiments concentrated so far on microcosms involving the most prominent *Methylococcaceae* partner, *Methylobacter*, as determined through the top-down approach, manipulation of synthetic communities of increased complexity used multiple (up to 50) species of methanotrophs, non-methanotrophic methylotrophs as well as non-methylotrophic heterotrophs. This experimental setup involves significantly more complexity compared to the two-species communities. However, these communities are completely tractable as they are made up of organisms with known genomic sequences, with predicted physiological traits validated through phenotypes observed in the lab, and these are mixed at predetermined relative abundances. By placing these communities under specific cultivation regimens (‘low’ versus ‘high’ oxygen, ‘low’ versus ‘high’ methane tensions etc.), we observed specific community dynamics, selecting for a smaller subset of originally mixed species, akin to dynamics observed in the top-down experiments. However, the species dominating these dynamics differed somewhat from the ones in the top-down experiments involving natural sediment communities. Most remarkably, *Methylomonas* species, while rapidly outcompeted in the top-down experiments with natural sediment samples by the *Methylobacter* species, persisted under both ‘low’ and ‘high’ oxygen pressures in these experiments, while *Methylosarcina* species persisted under ‘low’ methane. The nature of differential competitiveness of different *Methylococcaceae* species is being further addressed by employing two to three species communities of methanotrophs representing three major genera, *Methylobacter*, *Methylosarcina* and *Methylomonas*, and through increasing complexity of these communities by adding select non-methanotrophs species in different combinations.

This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC-0010556.

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

Ember M. Morrissey[1,2*] (ember.morrissey@mail.wvu.edu), Rebecca L. Mau[1], Egbert Schwartz[1,3], J. Gregory Caporaso[1,3,4], Paul Dijkstra[1,3], Natasja van Gestel[1], Benjamin J. Koch[1], Cindy M. Liu[4,5,6], Jane C. Marks[1,3], Lance B. Price[5,6], George W. Koch [1,3], Steve Blazewicz[7], Jennifer Pett-Ridge[7], Peter K. Weber[7], Xavier Mayali[7], **Bruce A. Hungate**[1,3]

[1]Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ, 86011, USA; [2]Division of Plant and Soil Sciences, West Virginia University, Morgantown, WV, 26505; [3]Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, 86011, USA; [4]Center for Microbial Genetics and Genomics, Northern Arizona University, Flagstaff, AZ, 86011, USA; [5]Center for Microbiomics and Human Health, Translational Genomics Research Institute, Flagstaff AZ, 86001, USA; [6]Department of Environmental and Occupational Health, Milken Institute School of Public Health, George Washington University, Washington DC, 20037, USA; [7] Lawrence Livermore National Laboratory, Livermore, CA, 94550, USA

Project Goals: The goal of this project is to develop new omics-driven technical approaches that couple multiple-element stable isotope probing with phylogenetic analysis to investigate microbial community functional processes involved in soil carbon cycling. These techniques will be used to identify soil bacteria and fungi involved in organic carbon degradation, examine the impacts of shifting environmental variables on their functional processes, and determine if there is a "phylogentic imprint" on the soil carbon cycle. The capability to quantify *in situ* microbial growth rates both at the community scale and for specific taxa will be the primary advantage of this new methodology. The project will leverage several long-term field research sites (ponderosa pine and mixed conifer forests) that have been subject to experimental climate change manipulation. The information generated in this work will help to establish whether phylogenetically specific imprints are observed on soil carbon cycling processes and facilitate better incorporation of omics-derived data into process-scale modeling efforts.

Soils are a huge reservoir of carbon, exceeding phytomass and atmospheric carbon combined. Anthropogenic increases in CO_2 are expected to augment primary productivity and thus enhance carbon transfer from the atmosphere to the soil potentially increasing soil carbon storage. However, the consequences of enhanced primary production on soil carbon storage remain unclear as microbial decomposition activities respond dynamically to fresh carbon substrates. Specifically, the decomposition of native soil organic C can be reduced or enhanced in response to fresh carbon inputs, a phenomenon known as the "priming effect". We used quantitative stable isotope probing with ^{13}C-labeled glucose and ^{18}O-labled water to measure individual and community level activity in order to understand how microbial activity mediates priming in soil. Initially labile carbon addition decreased soil carbon mineralization (negative priming) but over time repeated additions increased the mineralization of soil carbon (positive priming). This shift in activity was associated with an increased relative abundance of Proteobacteria and TM7 and a decrease in the proportion of Acidobacteria and Actinobacteria. By comparing changes in ^{18}O assimilation (growth) due to labile C addition with the amount of ^{13}C assimilation from the added substrate, we assessed the changes in soil carbon utilization

induced by fresh carbon inputs. Initially labile carbon was being consumed in lieu of soil organic matter, a phenomena often called preferential substrate utilization, causing the negative priming. After repeated carbon additions, labile carbon increased the growth of most prokaryotic taxa. This additional growth was achieved using a mixture of the added carbon and the soil organic matter resulting in enhanced native carbon mineralization explaining the positive priming. To understand how responses to labile carbon addition were distributed across bacterial taxa we categorized changes in activity and tested for phylogenetic clustering. Most bacterial taxa were involved in priming, and these organisms were not phylogenetically clustered. This suggests that increased growth and soil carbon utilization in response to fresh carbon inputs is routine among bacteria and does not require specialized physiological or ecological attributes. Consequently, priming may not be strongly constrained by bacterial biodiversity in soil.

This research was supported by the Office of Biological and Environmental Research in the DOE Office of Science

Wetland Microbial Community Response to Restoration

Susanna Theroux[1*], Wyatt Hartman[1], **Susannah Tringe**[1*] (sgtringe@lbl.gov)

[1]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, California

Project goals: This project aims to identify the geochemical and biological controls on greenhouse gas (GHG) production in coastal wetlands. Current projections of wetland response to global climate change are poorly constrained due to a limited understanding of the microbial mediators of carbon cycling in wetland soils. Using large-scale environmental genome sequencing, we aim to identify the microbial sentinels of GHG production and consumption in historic and restored wetlands of the San Francisco Bay-Delta. An enhanced comprehension of microbial metabolic arsenals will allow us to recommend management practices for future wetland restoration projects to maximize carbon sequestration and minimize microbial production of greenhouse gases.

Wetland environments play a critical role in the global carbon cycle, storing up to 35% of all terrestrial carbon and producing up to 75% of all non-anthropogenic methane. However, estimates of wetland GHG budgets are difficult to constrain given large variations in salinity, tidal regimes, soil saturation, organic carbon content, restoration status, and other geochemical and biological variables. Belowground microbial communities are the main drivers of greenhouse gases (GHG) cycling and their response to climate change will dictate whether a wetland serves as a net carbon sink or source. Wetland restoration has been proposed as a potential long-term carbon sequestration strategy, however wetland site selection and management practices are critical for ensuring restored wetlands sequester more GHG than they emit.

In an effort to better understand the underlying factors that shape the balance of carbon flux in wetland soils, we targeted the microbial communities along a salinity gradient ranging from freshwater tidal marshes to hypersaline ponds in the San Francisco Bay-Delta region. Using 16S rRNA gene sequencing and shotgun metagenomics, coupled with greenhouse gas monitoring and soil biogeochemical characterization, we sampled sixteen sites capturing a range of salinities and restoration status. As expected, freshwater wetland soils produced more methane than brackish and saline sites, since sulfate in seawater encourages sulfate reduction and discourages methane production. Restoration status also significantly affected GHG cycling: notably, restored freshwater and brackish wetlands produced orders of magnitude more methane than their historic counterparts, possibly due to differences in trace metal and organic carbon content in younger wetlands. However, unrestored former industrial salt ponds produced methane at rates rivaling those of brackish restored wetlands, and reconnection to the Bay resulted in a decrease in methane production.

The results of our microbial diversity survey showed that sampling location, plant type, and salinity were the primary drivers of both methane production and belowground microbial community composition. While overall methanogen abundance was only weakly correlated with methane production, both 16S and metagenome sequencing allowed for the identification of hallmark species whose relative abundance trended consistently with methane production. In unrestored salt ponds, the primary methane producers were methylotrophic methanogens utilizing substrates not accessible to sulfate reducers. Our study links belowground microbial communities with their aboveground greenhouse gas production and provides a benchmark for predicting wetland soil microbial response in the face of both natural and unnatural disturbances.

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.This project was funded by the DOE Early Career Research Program, grant number KP/CH57/1 to Susannah Green Tringe.

Gene- and genome-centric metagenomic analysis of complex microbial communities along a permafrost thaw gradient

Ben J. Woodcroft*[1] (b.woodcroft@uq.edu.au), Joel A. Boyd[1], Caitlin M. Singleton[1], Paul N. Evans[1], Rachel Wilson[2], Malak Tfaily[3], William Cooper[2], **Patrick Crill[4], Jeff Chanton[2], Scott Saleska[5], Virginia Rich[6], Gene W Tyson**[1]

[1] Australian Centre for Ecogenomics (ACE), School of Chemistry and Molecular Biosciences, University of Queensland, Australia
[2] Department of Earth, Ocean, and Atmospheric Science, Florida State University, Tallahassee, Florida, USA
[3] Pacific Northwest National Laboratory, Richland, Washington, USA
[4] Department of Geological Sciences, Stockholm University, SE-106 91 Stockholm, Sweden
[5] Ecology and Evolutionary Biology Department, University of Arizona, Tucson, Arizona, USA
[6] Department of Microbiology & Department of Civil, Environmental, and Geodetic Engineering, The Ohio State University, Columbus, Ohio, USA

URL: http://ecogenomic.org/melting-permafrost

Project goals: **As a result of climate change, large amounts of carbon sequestered in permafrost (~50% of global soil carbon) are becoming available for microbial degradation. One of the primary results of permafrost thaw is the production of the potent greenhouse gas methane, forming a positive feedback to climate change. Here, meta-omic and geochemical techniques are used to elucidate tangible links between microbiology and geochemistry along a natural permafrost thaw gradient in northern Sweden.**

We analysed ~2 Tbp of shotgun metagenomes from 214 samples taken across the permafrost thaw gradient transitioning through intact palsa, partially thawed sphagnum-dominated bog and fully thawed eriophorum-dominated fen. The samples span several depths from triplicate soil cores taken monthly during the summer thaw over several consecutive years. A comparison of metagenome-derived community profiles with 16S rRNA amplicons showed strong amplification biases against numerous lineages. The most notable differences between these profiling techniques was the absence of the candidate phylum WPS-2 in the amplicons

(despite comprising >5% of the community according to the metagenomes) and the overamplification of methanogens (up to 5-fold).

Assembly and genome binning of the metagenomes enabled the recovery of 1,529 high quality population genomes (>70% completeness & <10% contamination), including genomes from three phyla that currently lack sequenced representatives. Using a newly developed method, SingleM, which uses conserved single copy marker genes to examine community composition at high resolution, we assessed how representative the recovered population genomes were of these communities. Our results indicate that >60% of these complex microbial communities are represented by at least one population genome within the same genus.

Through the recovery of population genomes representative of the majority of microbes from these complex peat environments, we are able to couple macroscale biogeochemical measurements to predicted microbial metabolic potential. These genome-centric approaches are furthering our understanding of thawing permafrost communities, allowing us to more fully elucidate the relationships between the geochemistry, microbial ecology and global climate models.

This work is supported by the U.S. Department of Energy under funding opportunity announcement number DE-FOA-0000866. Pathways to carbon liberation: a systems approach to understanding carbon transformations and losses from thawing permafrost and iVirus is supported by the US Department of Energy Office of Biological and Environmental Research under the Genomic Science program (Award DE- SC0010580).

An alternate H_2 sink in Anaerobic Environments Explains Non-Stoichiometric Greenhouse Gas Production Ratios in Boreal Peatlands and Thawing Permafrost.

R.M. Wilson[1*](rmwilson@fsu.edu) and M.M. Tfaily[2], V. Rich[3], J. Kostka[4], S. Bridgham[5], J. Keller[6], C. Medvedeff[6], L. Meredith[7], P. Hanson[8], M. Hines[9], L. Pfeifer-Meister[5], G. Tyson[10], B. Woodcroft[10], S. Saleska[7], P. Crill[11], W.T. Cooper[1], and **J.P. Chanton[1]**

[1]Florida State University, Tallahassee, FL, [2]Pacific Northwest National Lab, Richland WA, [3]The Ohio State University, [4]The Georgia Institute of Technology, Atlanta GA, [5]University of Oregon, Eugene, OR, [6]Chapman University, Orange, CA, [7]University of Arizona, Tuscon, AZ, [8]Oak Ridge National Lab Knoxville, TN, [9]University of New Hampshire, Durham NH, [10]University of Queensland, Australia, [11]Stockholm University, Sweden

Project Goals: ***Our objective is to discover how microbial communities mediate the fate of carbon in thawing permafrost under climate change.*** **We propose a systems approach integrating *(a)* molecular microbial and viral ecology, *(b)* molecular organic chemistry and stable and radiocarbon isotopes, and *(c)* state-of-the-art modeling, along a chronosequence of permafrost thaw in subarctic Sweden.**

Following the depletion of inorganic electron acceptors, organic matter in anoxic environments decomposes by hydrolysis, fermentation, and methanogenesis requiring syntrophic interactions among microbes to achieve energetic favorability. In the classic anaerobic food chain, hydrogenotrophic methanogenesis is the terminal electron acceptor (TEA) and ultimately produces equimolar CO_2 and CH_4 for each molecule of organic matter degraded. However, CO_2:CH_4 production in deep anoxic peat often exceeds this 1:1 ratio, in seeming contradiction of thermodynamic theory. Here we present evidence that, in peatlands, the ubiquitous hydrogenation of a range of diverse unsaturated compounds serves as an alternative to methanogenesis as a terminal electron acceptor and H_2 sink. This mechanism lowers the partial pressure of dissolved H_2, maintains electron flow through syntrophic reactions, and promotes CO_2 production without concomitant CH_4 production. While organic TEAs have been proposed before to drive microbial respiration of organic matter through the reversible reduction of quinone moieties, the mechanism proposed herein differs from those earlier mechanisms by also acting as a requisite step in the degradation of the organic compounds that are reduced. The implication of this proposed mechanism is that it has the potential to control CO_2:CH_4 production and emission ratios from peatlands which ultimately determines their global warming potential.

M-tools and iVirus: Software tools and a cyberinfrastructure for meta-omic analyses

Joel A. Boyd[1*] (joel.boyd@uqconnect.edu.au), Benjamin Bolduc[3*] (bolduc.10@osu.edu), Paul G. Dennis[1], Michael Imelfort[1], Timothy Lamberton[1], Donovan H. Parks[1], Simon Roux[3], Connor T. Skennerton[2], Ben J. Woodcroft[1], Ken Youens-Clark[4], Phil Hugenholtz[1], Bonnie L. Hurwitz[4], **Gene W. Tyson[1]**, **Matthew B. Sullivan[2]**

[1]Australian Centre for Ecogenomics (ACE), School of Chemistry & Molecular Biosciences, University of Queensland, Queensland, Australia; [2]Division of Geological & Planetary Sciences, California Institute of Technology, Pasadena, California, USA; [3]Department of Microbiology & Department of Civil, Environmental, and Geodetic Engineering, The Ohio State University, Columbus, Ohio, USA; [4]Department of Agricultural & Biosystems Engineering, University of Arizona, Tucson, Arizona, USA

URL: http://ecogenomic.org/software; http://ivirus.us

Project Goals: Over the last decade, analysis of microbial and viral communities has undergone a major shift as the result of improvements in sequencing technology. The transition from single gene amplicon to large-scale meta-omic studies has allowed us to capture not only the phylogenetic but functional diversity of a community. However, the size and complexity of meta-omic datasets often require that multiple specialized tools be applied to address the different aspects of microbial and viral community analysis. Here we present two software suites, M-tools and iVirus, that are being developed as part of an ongoing interdisciplinary project focused on exploring the ecological and biogeochemical implications of climate change induced permafrost thaw.

The M-tools suite is organised around the analysis of microbial metagenomes, with a primary goal of providing a user-friendly pipeline for extracting high quality population genomes. The M-tools suite also provides software for community profiling of unassembled metagenomic data. iVirus is focused on collecting viral datasets and deploying the most commonly used tools for viral meta-omics, creating a publicly available, community resource ideal for sharing and collaboration. M-tools is comprised of six programs that span the workflow from raw data to high quality population genomes and iVirus has four apps developed for interrogating meta-omic datasets.

M-tools:

- **GraftM:** Creates community profiles from raw meta-omic sequences using Hidden Markov Models (HMM) to identify genes of interest which are classified using

phylogenetic tree insertion methods. Provides the tools for the creation of custom gene packages for the analysis of sequence data.

- **SingleM:** Provides highly resolved community composition from metagenomes using conserved single copy marker genes. By not heavily relying on reference databases of sequenced genomes, SingleM can be used to accurately profile communities that contain novel lineages. It can also be used to determine how representative a set of population genomes is of a community.
- **GroopM:** Recovers population genomes from large metagenomic datasets using differences in population abundance across metagenomic samples (differential abundance binning).
- **CheckM:** Assesses the quality of isolate, single cell, and population genomes using lineage specific single copy marker gene sets. Includes utilities for comparing genomes and exploring features such as GC content, sequence length, and tetranucleotide signatures.
- **RefineM:** Refines isolate, single cell or population genomes using qualitative and quantitative features such as GC content, coverage and coding density.
- **OrfM:** Rapidly predicts ORFs in raw metagenomic reads.

iVirus:

- **vContact/vContact-PCs:** Generates Protein Clusters (PCs) using a Markov clustering algorithm and incorporates metadata annotations. Then assigns contigs to taxonomic groups using the presence or absence of shared PCs along the length of the contig.
- **PCpipe:** Compares ORFs from user-defined datasets to existing viral PCs as a means to organize viral sequence space into functional units that can serve as (i) a universal functional diversity metric for viruses, (ii) a scaffold for iterative functional annotations, and (iii) input for ecological comparisons.
- **Fizkin:** Performs Bayesian network analyses based on the amount of shared sequence content in viromes and contextual data about the sample's environment.
- **VirSorter:** Identifies viral sequences in microbial genomes and metagenomic datasets

This work is supported by the U.S. Department of Energy under funding opportunity announcement number DE-FOA-0000866. Pathways to carbon liberation: a systems approach to understanding carbon transformations and losses from thawing permafrost and iVirus is supported by the US Department of Energy Office of Biological and Environmental Research under the Genomic Science program (Award DE- SC0010580).

Title: DNA-SIP enabled community genomics of cellulose degraders in an agricultural soil

Charles Pepe-Ranney[1*] (chuck.peperanney@cornell.edu), Ashley N Campbell,[2] and **Daniel H Buckley[1]**

[1]School of Integrative Plant Science, Cornell University, Ithaca; [2]Lawrence Livermore National Laboratory

Project Goals: Short statement of goals. (Limit to 1000 characters)
Project Goals: This research program will reveal fundamental aspects of soil C-cycling and provide ecological and metabolic insights on diverse non-cultivated soil microorganisms that play major roles in the global C-cycle. Specific goals include: 1) Map the C assimilation dynamics for thousands of non-cultivated microorganisms in soil by harnessing a full cycle microbial food web mapping approach that employs an array of ^{13}C-labeled molecules; 2) Map the C assimilation dynamics of soil microorganisms across soil systems as a function of soil characteristics; and 3) Evaluate ecological and seasonal patterns of activity and abundance for discrete microbial taxa across gradients of soil characteristics and as a function of their C-assimilation dynamics. These goals will be achieved by employing a newly developed microbial food web mapping approach, enabled by advances in ^{13}C-stable isotope probing of nucleic acids and next generation sequencing.

Microorganisms drive biogeochemical cycles and because soil is a large global carbon (C) reservoir, soil microorganisms are important players in the global C-cycle. Frustratingly, however, many soil microorganisms resist cultivation and soil communities are astoundingly complex and hence the dynamics of soil C metabolism remain incompletely described. Stable isotope probing (SIP) is a useful approach for establishing identity-function connections in microbial communities but has been challenging to employ in soil due to the inadequate resolution of microbial community fingerprinting techniques. High resolution DNA sequencing improves the resolving power of SIP transforming it into a powerful tool for studying the soil C cycle. We conducted a DNA-SIP experiment to track flow of cellulose-C, the most abundant global biopolymer, through a soil microbial community.

We found that uncultivated bacterial lineages among *Spartobacteria*, *Chloroflexi*, and *Planctomycetes* assimilated ^{13}C from ^{13}C-cellulose. These lineages are cosmopolitan in soil but little is known of their ecophysiology. In addition to cataloging the SSU rRNA genes of cellulose responsive microorganisms we used DNA-SIP as an approach for targeted community genomics. DNA-SIP enriches DNA of targeted microorganisms. For example, *Verrucomicrobia* cellulose degraders were enriched by nearly two orders of magnitude in the labeled DNA pool, and this "enriched" DNA can serve as template for community genomics. We employed a stripped down binning approach that coupled dimensional reduction of contig oligonucleotide signatures with density based clustering followed by complete linkage clustering of assembly contigs based on paired read links. This approach scales to large data, is easily interpreted and visualized, and

allows for flexible bin definitions such that contigs from strains highly similar in genomic content (that might be overlooked by conventional community genomics analyses) can be clustered into useful groups alongside strain specific genome bins. Using this approach, we produced draft genomes from soil cellulose degraders including microorganisms belonging to *Verrucomicrobia*, *Chloroflexi*, and *Planctomycetes*.

The metagenomic assembly of ^{13}C-enriched DNA yielded approximately 3,100 genes that are associated with the deconstruction of cellulose and 57 pangenome bins contained at least two of these cellulose-associated CAZymes. We also recovered 14 high quality single genome bins (>75% complete and < 10% contaminated as determined by assessment of single copy genes). For example, we recovered a nearly complete novel genome (96% complete) belonging to soil *Chloroflexi* which shared no more than 78% SSU rRNA gene identity to any cultured relative, and we also recovered a draft genome (75% complete) belonging to a lineage of *Verrucomicrobia* that possessed 14 cellulose-associated CAZymes. In addition, we identified a pangenome bin for *Cellvibrio* – a well-studied, model cellulose degrader – which consisted of several polymorphic strains all of which appear to have DNA labeled by 13C-cellulose. This study demonstrates how DNA-SIP can be used to study the ecophysiology of microbes important in terrestrial C cycling and to target guilds of microorganisms for characterization by community genomics.

Funding statement.
This material is based upon work supported by the Department of Energy Office of Science, Office of Biological & Environmental Research Genomic Science Program under Award Numbers DE-SC0004486 and DE-SC0010558.

Title: Soil microbial food web mapping with high resolution stable isotope probing

Authors: N. D. Youngblut* (nyoungb2@gmail.com), C. Koechli, **D. H. Buckley**

School of Integrative Plant Science, Cornell University, Ithaca

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

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

Soils make up the largest active carbon pool on the planet. Although carbon cycling in soil is largely mediated by microbial life, the specific taxonomic groups that perform each role in the soil microbial food web have not been well resolved. High-resolution stable isotope probing (HR-SIP) leverages highly multiplexed high-throughput 16S rRNA sequencing to simultaneously map *in situ* substrate assimilation dynamics to potentially thousands of finely resolved microbial taxa. In this study, HR-SIP was performed with nine ^{13}C isotopes (cellulose, xylose, glucose, glycerol, vanillin, palmitic acid, amino acids, lactate, and oxalate) in order to identify carbon assimilators at multiple stages in the breakdown of plant biomass in soil.

During the 48 day incubation, we observed a coinciding succession of ^{13}C-substrate respiration and also incorporation into bacterial biomass. Relatively labile substrates (e.g., glucose) were utilized first (days ~1-6), followed by oxalate (days ~6-14), and finally by cellulose and palmitic acid (days ~14-30). The total amount of ^{13}C respired varied substantially between substrates, from ~35% for vanillin to nearly 100% for lactate. The number ^{13}C-incorporating taxa ("incorporators") also varied among treatments, with ~4-6% of taxa incorporating cellulose or palmitic acid, compared to <2% of taxa incorporating any other substrate. Phylogenetic similarity was very high between cellulose and palmitic acid incorporators and also between lactate and oxalate incorporators. Interestingly, nearly all of the lactate incorporators identified on day 6 were also identified as oxalate incorporators, but not until day 14, suggesting diauxic growth is common for soil bacteria consuming these fermentation products. While certain *Gammaproteobacteria* and *Firmicutes* taxa consumed almost all substrates, many other taxa specialized on one or two substrates. For instance, certain *Firmicutes* taxa solely

incorporated glucose although all other substrates were present. These results suggest pervasive niche partitioning among bacterial taxa in the soil carbon cycle, with partitions for both the relatively transient dissolved organic matter pool and the more persistent particulate organic matter pool. More generally, these findings will help define ecologically relevant taxonomic groups of microbes with coherent functional roles in the soil microbial food web.

Funding statement.
This material is based upon work supported by the Department of Energy Office of Science, Office of Biological & Environmental Research Genomic Science Program under Award Numbers DE-SC0004486 and DE-SC0010558.

Title: Tillage history drives changes in the dynamics of microbial respiration, assimilation and growth in soil upon addition of dissolved organic carbon (as ^{13}C-xylose)

C. Koechli* (cnk29@cornell.edu), N. D. Youngblut, **D. H. Buckley**

School of Integrative Plant Science, Cornell University, Ithaca

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

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

Bacteria are essential to the cycling and storage of carbon in the soil ecosystem. Tillage decreases soil organic matter content and changes the composition of soil microbial communities. Differences in microbial ecology between no-till vs tilled soils may contribute to differences in organic matter loss pathways, however mechanistic linkages between microbial community structure and function remain unclear in soils.

Microbial contributions to the degradation of both dissolved (as ^{13}C-xylose) and particulate (as ^{13}C-cellulose) carbon compounds were contrasted in no-till and tilled soils by using high resolution DNA stable isotope probing (HR-SIP) to evaluate the temporal dynamics of carbon assimilation and respiration in a series of soils from a long-term tillage experimental in Chazy, NY. Additionally, we used high throughput sequencing of 16S rRNA gene amplicons to assess seasonal variation in microbial community composition at field scale in relation to tillage history. For each OTU, patterns of isotope incorporation in HR-SIP experiments were assessed in relation to variation in OTU relative abundance as a function of season, tillage history, and their interaction.

Using analysis of variance of Unifrac distance matrices (PERMANOVA), we see that bacterial communities vary significantly (R^2 = 0.12, p = 0.001) with tillage, as well as with season (R^2 = 0.06, p = 0.001). We find that no-till soil has significantly higher rates of soil respiration and higher rates of ^{13}C-xylose, but not ^{13}C-cellulose, mineralization relative to tilled soil. The set of bacteria that incorporated ^{13}C xylose differed between tilled and no-till soils at the beginning of the experiment (days 1,3) but became more similar over time. In contrast, the set of bacteria that incorporated ^{13}C cellulose remained similar between tilled and no-till soils throughout the

experiment. However, the number of OTUs that incorporated ^{13}C cellulose differed, with 195 OTUs incorporators in the no-till treatment as compared to 136 in the plow-till treatment.

Comparing the incorporation of ^{13}C from cellulose and xylose into the bacterial community in tilled vs. no-till soils shows that the bacteria participating in carbon transformation differ as a function of soil management history, with implications for carbon fate. The diversity of bacteria that incorporate xylose, and rates of xylose respiration varied with respect to tillage. The diversity of bacteria that incorporate cellulose also varied with tillage but, in contrast, no corresponding differences were observed in rates of cellulose respiration. These results suggest that changes in the structure of the microbial community affects xylose degradation but not cellulose degradation. It is possible that this outcome may be a consequence of the form in which the carbon is delivered as either dissolved or particulate organic carbon.

Funding statement.
This material is based upon work supported by the Department of Energy Office of Science, Office of Biological & Environmental Research Genomic Science Program under Award Numbers DE-SC0004486 and DE-SC0010558.

Multi-'Omic' Analyses of the Dynamics, Mechanisms, and Pathways for Carbon Turnover in Grassland Soil

Cristina N. Butterfield[1]* (butterfc@berkeley.edu), Zhou Li[2], Peter Andeer[3], Susan Spaulding[1], David Burstein[1], Susannah G. Tringe[3,4], Brian C. Thomas[1], Robert Hettich[2], Trent Northen[3,4], Chongle Pan[2], and **Jillian F. Banfield**[1]

[1]University of California, Berkeley; [2]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]Lawrence Berkeley National Laboratory, Berkeley, CA; [4]DOE Joint Genome Institute, Walnut Creak, CA

http://ggkbase.berkeley.edu

Project Goals:
Climate change will alter terrestrial ecosystems. However, the strength and the direction of change will be shaped by feedbacks, most of which will be difficult to predict. Of primary importance in this regard is how the distribution of carbon between the atmosphere and the subsurface will change in response to altered rainfall, temperature and vegetation patterns. Metagenomics, proteomics, transcriptomics, and metabolomics are used to compare the membership and functioning of soil communities at three different depths below the root zone in a grassland that experiences a Mediterranean climate. We track microbial community composition and activity during the period of major carbon turnover in this ecosystem under two rainfall scenarios, identify key carbon currencies released from the soil zone and provide a basis for prediction of how grassland ecosystems will respond to future climate change.

Abstract (2 page limit):
We are investigating how varied rainfall events impact the dynamics of carbon storage in grassland soil, and the potential consequences for the transport of carbon and other nutrients from these soils to the underlying vadose zone and streams. However, little is known of the microorganisms that play vital roles in the processing of this largely vegetation-derived soil carbon and how the metabolic activities that occur at different soil depths ultimately impact DOC discharged into flanking environments. We are studying the period around the first Fall rainfall event, when soil-associated carbon fixed during Spring growth is rapidly metabolized, focusing on climate manipulations that differ 1) in the amount of Spring rainfall (above-ground carbon stocks), 2) on the period of time following the first Fall rain events (soil microbial communities), and 3) soil depth.

We couple genome-resolved metagenomics with proteomics and metabolomics to determine the metabolic capabilities of soil microbial communities and to map carbon flow through the sub-root soil. Our research is conducted in a well-monitored grassland in the Angelo Coast Range Reserve in Northern California that has been subjected to a long term rainfall manipulation to simulate climate change. This grassland experiences a

Mediterranean climate, which means that mild rainy winters drive spring plant growth, and hot dry summers drive the senescence of these plants. Thus, the period immediately following plant death, when the first rainfall events occur, leads to the rapid transformation and mobilization of plant-derived organic carbon and its microbial degradation products within this system as well as to lower soil depths. In our pilot year (2013), we observed a disturbance in the overall architecture of the bacterial and archaeal communities and a subsequent rapid return to pre-rain, baseline levels. We reconstructed 198 genomes, including 46 near-complete genomes from the relatively abundant organisms from every bacterial and archaeal phylum, and binned the remaining DNA sequence based on taxonomy of predicted genes. This information was used to create a protein sequence database for proteomics. We identified and quantified 2880~4700 proteins and their microbial origins from 10 soil communities. We found that much of the community was actively degrading plant biomass following the first rain event of the season. Some of the most abundant functions were rarely reported in the literature, and included the aromatic carbon degradation by Thermoplasmatales archaea and methylotrophy by Gemmatimonadetes and Rokubacteria bacteria. Metabolomics analyses also illuminated that most organic and nitrogenous compounds are efficiently degraded by the 30 cm depth level.

Currently, we are automating methods that were successfully applied to the 10 samples collected in the 2013 Fall rainy season to analyze 60 soil metagenomes and 20 soil proteomes collected in Fall 2014. This increase in bioinformatics capacity will make processing of the 148 samples from 2015 possible. We have both automated and optimized the application of the metagenomic assembly program IDBA-UD and achieved a significant improvement in both assembly throughput and recovered scaffold lengths for the 2014 samples over the 2013 samples. We have explored how both read sequencing and physical sample depth relate to the difficulty of assembly of sequence information from a highly complex soil environment. Finally, we are exploring the application of differential coverage binning from very high numbers of samples and the integration of multiple automated binning programs to improve both the throughput and quality of genome binning significantly. In particular, differential coverage binning allows us to leverage sequence information from 2013 samples to improve genome bins of 2014 data, and will subsequently improve the binning of all samples in future years.

The 2014 dataset includes samples from plots amended with water over the last 15 years to simulate the longer spring rainfall predictions indicated in some climate change models. Comparison of microbial communities present in the control vs. manipulations should indicate to us a shift in lifestyle associated with environmental change. Preliminary metagenomics analyses indicate that the 2013 and 2014 pre-rain samples are similar in terms of community composition. A priority objective for analysis of these and post rain samples is to improve functional predictions to facilitate metabolic substrate specificity assignment and link this information to metabolomics measurements.

To aid in the identification of specific transformation pathways, we have developed a strategy using isotopically-labeled metabolites to follow transformations of particular carbon substrates by the soil community. The identification of these transformations

should facilitate the putative identifications of relevant proteins and microbes within our larger in situ field data sets. We can then validate these assignments using targeted-, multi-omic analyses of stable isotope probing experiments, which will enable the tracking of nutrients through community members and thus a model of the system.

Overall, “omics” methods are being integrated to capture the dynamics of microbial communities in soil below the grass root zone. Below ground terrestrial system processes will impact grassland ecosystem function and the global carbon cycle.

This research is supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy Grant DOE-SC10010566.

EcoFun-MAP: an Ecological Function Oriented Metagenomic Analysis Pipeline

Zhou Shi (jason.shi@ou.edu)[1], Zhili He[1], Erin Nuccio[3], **Jennifer Pett-Ridge[3], Mary Firestone[2]** and **Jizhong Zhou**[1]

[1]University of Oklahoma, Norman, OK; [2]University of California, Berkeley CA; [3]Lawrence Livermore National Laboratory, Livermore, California;

Project Goals: The EcoFun-MAP has been developed as an accurate, efficient and highly accessible strategy to fish out reads of important environmental functional genes from shotgun metagenome sequence data. EcoFun-MAP will facilitate functional analysis of large complex data by 1) constructing reliable and comprehensive reference databases specializing on the functional genes that are important to the ecological functions and geochemical processes and 2) providing efficient tools, graphical interface and automated workflow to easily access the databases for reducing repetitive effort by metagenome researchers.

Increasingly large amount of next generation sequencing (NGS) data has on one hand resulted in unprecedented insights into microbial ecology studies, and on the other hand, created a burden onto the computational analysis due to a lack of quick, accurate and dedicated tools. Here we present EcoFun-MAP, an Ecological Function oriented Metagenomic Analysis Pipeline, which was developed for automatic analyses of metagenome sequencing data from an ecological function perspective. EcoFun-MAP was built upon two functional gene databases, including a protein sequence based Hidden Markov Model database, and nucleotide sequence based functional OTU database, and both of them were manually curated and specifically tailored for fitting the applicable scope. EcoFun-MAP allows to profile a large amount of raw reads down to the functional OTU level, and annotating them into hierarchical ecological functional categories.

We refine the applicable scope of EcoFun-MAP to the functional genes encoding proteins/enzymes that play crucial roles in the major geochemical processes and ecological functions, including carbon (C), nitrogen (N), sulfur (S), and phosphorus (P) cycling, electron transferring, metal homeostasis, organic remediation, stress responses, secondary metabolism, and virus and virulence activity. Within those categories, a total of 1399 functional genes (e.g., *nifH, nirS, nirK, amoA*) were selected, and for each of them, a keyword-based query was manually crafted and submitted to the National Center for Biotechnology Information (NCBI) online databases for the retrieval of both protein and nucleotide based candidate reference sequences. The number of sequences retrieved for each gene could be ranged from a few to tens of thousands.

Two reference databases were constructed for the EcoFun-MAP to fully function: a Hidden Markov Models (HMMs) based database (HMMDB), and functional OTUs (FOTU) based database (FOTUDB). To construct HMMDB, we manually selected a minimum of five to a few hundred distinguished representative sequences as seed sequences (SS's) from protein based candidate sequences for each gene. Then, the selected seed sequences were aligned in ClustalW,

and the produced alignments were manually verified and later used as inputs for another program HMMBUILD to build function gene HMMs. The process has been done repeatedly for all functional genes and resulted in HMMDB finally. To construct FOTUDB, the candidate reference sequences of each gene were searched back against corresponding HMM from HMMDB using HMMSEARCH with applying an e-value cutoff to ensure that irrelevant sequences were excluded from further procedures. Due to the heterogeneity among the sequence sets of different genes, the most appropriate cutoff value for each gene could differ from others, therefore needs tremendous human interference by make repeated adjustments. After that, the output sequences were considered to be highly reliable reference functional gene sequences. Next, FOTUs were generated by clustering the confirmed sequences into a number of OTUs using CD-HIT99 with group similarity threshold of 95%, and corresponding BLAST databases were also constructed using MAKEBLASTDB. To this end, both of two reference databases have been established.

In the annotating workflow, HiSeq sequence results were resampled in each sample based on the minimal reads number in samples. The resampled sequences were input as raw unknown nucleotide sequences and were trimmed by Btrim with setting window size to 5 and average quality to 20, so as to remove unreliable sequences indicated by poor quality score. Next, all trimmed nucleotide sequences were translated into protein sequences using FragGeneScan with an error ratio 10%, which is widely accepted as estimated normal Illumina sequencing error ratio. Then HMMSEARCH was used for annotating the predicted protein sequences with the HMMDB database and an e-value cutoff can be customized by the users, and both global and local model hits were counted as valid results. In an additional filtering step, all confirmed sequences were compared together back against the FOTUDB with BLASTN. Only the best hits (Rank No. 1 in BLAST results) were kept as final fish-out results. All processing steps, statistical analysis methods, and bioinformatics tools are organized into a pipeline and will be integrated into a DOE KBase.

This research is based upon work supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Numbers DE-SC0004730 and DE-SC0010570 at UC Berkeley and by UC-subcontract number 00008322 at the University of Oklahoma.

Determining key physiological and metabolic traits of soil microorganisms that regulate C assimilation and transformation

Kateryna Zhalnina[1,2*] (kzhalnina@lbl.gov), Ulas Karaoz[2], Heejung Cho[1], Katherine B. Louie[2], Benjamin P. Bowen[1], Nasim Mansoori[3], Zhao Hao[1], Richard A. White III[4], **Mary S. Lipton**[4], **Trent R. Northen**[2], **Jennifer Pett-Ridge**[5], **Mary K. Firestone**[1,2], and **Eoin L. Brodie**[1,2]

[1]University of California, Berkeley; [2]Lawrence Berkeley National Laboratory, Berkeley, California; [3]Joint BioEnergy Institute, Emeryville, California; [4]Pacific Northwest National Laboratory, Richland, Washington; [5]Lawrence Livermore National Laboratory, Livermore, California

http://www.lbl.gov

Project Goals: Our project works towards a fundamental understanding of C cycling in soil as mediated by soil microorganisms and their interactions with plants. How do 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? We seek to gain a mechanistic understanding of the conversion of root-derived C to stabilized soil C, clarify the impacts of microbial activities on soil C sequestration, and substantially expand our understanding of molecular regulation of terrestrial C cycling.

Plant-soil-microbial interactions may strongly impact the stability of soil organic carbon (SOC). Plants exude variety of compounds, supplying rhizosphere microorganisms with readily assimilable substrates and driving microbial succession in response to developing plant roots. Uptake of root exudates may stimulate microorganisms to produce enzymes that degrade more complex SOC and potentially decrease the stability of older C in soil. Alternatively, it has been shown that root exudates may liberate C from protected mineral surfaces in soil making it more available for microbial mineralization. Together these contribute to the phenomenon known as “rhizosphere priming”. Although rhizosphere priming of SOM decomposition has been widely demonstrated, the mechanisms underlying this effect and our ability to accurately predict it remain uncertain. We combine modeling and experimentation to determine key traits of soil microorganisms relevant to their fitness in the rhizosphere and transformation of carbon.

To define putative microbial traits we used metagenomic and isolate sequences from a Mediterranean grassland soil to identify changes in microbial composition and function in response to plant growth. First, we carried out genome-centric analyses using 38 bacterial isolates representative of the dominant organisms identified in metagenomes of this Mediterranean grassland soil. To determine the trajectories of these bacterial heterotrophs during key developmental phases of our model plant (*Avena fatua*), we obtained DNA from rhizosphere and bulk soil sampled over 12 weeks of plant growth with ~5GB of metagenome reads obtained per time point. Reads were aligned to the isolate genomes and coverage of genomes and marker genes were used to track bacterial dynamics during the root growth. Second, a genome-centric approach was taken to reconstruct genomes from metagenomic reads obtained from field soil sampled from dry and wet seasons and mini-rhizotron bulk/rhizosphere soils sampled during root growth. A total of 851 GB of sequence were co-assembled and binned based on differential coverage and sequence composition. Curation of these bins resulted in 92 organismal bins (82 bacterial, 7 archaeal, and 3 fungal), each representing a reconstructed genome. These reconstructed genomes were classified based on their

response to root growth into general classes: positive responders (early, late, and gradual), and negative responders. The genomes were annotated with a specific focus on genes for degradation and transformation of complex carbohydrates, organic C-oxidation, C-fixation, and transporters; therefore linking the organisms and their response patterns with functional traits.

To test putative functional traits we used the sequenced heterotrophic isolates. Each genome was analyzed for key traits, such as minimum generation times, polymer degrading enzymes and transporters for different substrates (amino-, fatty-, organic-acids, sugars, nucleotides and plant hormones). Clustering of genomes according to these functional traits suggested that certain classes (e.g. α-proteobacteria) had a greater representation of rhizosphere fitness traits and may be expected to respond positively to root development.

We tested predicted traits using variety of functional analyses. Predicted minimum generation times (based on codon usage bias) were experimentally validated and demonstrated as a reliable approach for growth rate prediction in uncultivated microorganisms. Polymeric substrate preferences and enzymatic activities of different taxa were analyzed using nanostructure initiator mass spectrometry and fourier transform infrared spectroscopy. Based on these analyses Actinobacteria were shown to metabolize more recalcitrant C, such as lignin while α-proteobacteria demonstrated preferences towards cellulose when incubated with root litter. Enzymes involved in degradation of specific polymers were identified through secretome proteomic analysis.

The ability of rhizosphere microorganisms to utilize simple substrates and their metabolic response to plant root exudates was tested using a complex exudate medium. *Avena* exudates were collected, and identified my LC-MS/MS during plant development. Exudation composition showed significant differences both across vegetative stages and between vegetative and senescence stages. These patterns across plant developmental stages were associated with successional changes in rhizosphere microbial communities. Bacterial substrate preferences were tested by growing isolates on the exudate-derived medium and substrate uptake/release determined by LC-MS/MS. This approach provides an ability to link genome predictions of microbial substrate preferences with substrate utilization and niche specialization in the rhizosphere.

To predict the metabolic response of bacterial heterotrophs *in silico* to plant exudates and their impact on C turnover we developed genome-scale metabolic models of both positive and negative rhizosphere responders. The models were manually curated and gapfilled using BIOLOG and exometabolomics and are currently being adapted to represent extracellular enzyme production. Together these approaches are being used to predict and test how key traits of soil bacteria interact with root metabolism and soil organic matter to impact the phenomenon of rhizosphere priming. Our goal is to develop predictive metabolic and trait based models of the rhizosphere to simulate how climate driven changes in vegetation will feedback to soil C cycling and nutrient availability.

This research is based upon work supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Numbers DE-SC0004730 and DE-SC0010570 at UC Berkeley and FWP SCW1421 at LLNL. Support to UC Berkeley was also provided by EMSL project #48912. Funding was also provided by an LLNL LDRD "Microbes and Minerals: Imaging C Stabilization". Work at LLNL was performed under the auspices of the U.S. Department of Energy under Contract DE-AC52-07NA27344 and at Lawrence Berkeley National Laboratory under the auspices of the University of California - contract DE-AC02-05CH11231.

The interconnected rhizosphere: Roots stimulate high microbial network complexity and alter the expression of enzymes related to organic matter decomposition in soil

Erin Nuccio[1]* (nuccio1@llnl.gov), Shengjing Shi,[2,3] Jason Shi,[3] Ulas Karaoz,[4] **Eoin Brodie**,[4] Zhili He,[3] **Mary Lipton**,[5] **Jizhong Zhou,[3] Jennifer Pett-Ridge,[1] and Mary Firestone[2]**

[1]Nuclear and Chemical Sciences Division, Lawrence Livermore National Laboratory, Livermore, California; [2]University of California, Berkeley, California; [3]University of Oklahoma, Norman, Oklahoma; [4]Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, California; and [5]Pacific Northwest National Laboratory, Richland, Washington

Project Goals: The soil surrounding roots, the rhizosphere, is the primary nexus of belowground carbon cycling in terrestrial systems. Our project focuses on a fundamental understanding of C cycling in soil as mediated by soil microorganisms, their interactions, and how these interactions are changed in response to plants. In particular, our work investigates how the interactions between roots and soil microorganisms alter transformations of root derived C, and explores how roots alter the networks of microbe-microbe interactions in soil. The primary goals of this work are to determine how organic C decomposition is impacted by the interactions of the soil microbial community with living roots, and substantially expand our understanding of the microbial ecology and molecular regulation of terrestrial C cycling.

The soil surrounding plant roots, the rhizosphere, has long been recognized as a zone of great functional importance to plants and the terrestrial ecosystems they inhabit. Plants play a central role in transferring atmospheric CO_2 to belowground soil C pools, while microbes are primary mediators of C transformation and mineralization in the soil. However, the molecular mechanisms underlying soil-plant-microbial interactions are poorly understood. Rhizosphere microorganisms can alter the breakdown of plant tissues and root litter, and in many cases have been shown to accelerate the decomposition of detrital plant biomass. The soil microbial communities and carbohydrate and lignolytic gene transcripts mediating the decomposition of root litter in soil are largely unidentified.

We hypothesized that root exudates stimulate the expression of enzymes that are involved in decomposition of macromolecular C compounds. To assess how the abundance and diversity of decomposition enzymes differs in the rhizosphere relative to the surrounding bulk soil, we sequenced the metatranscriptomes of rhizosphere and bulk soil over time. Our results suggest that enzymes involved in the breakdown of plant polysaccharides were more highly expressed in the rhizosphere compared to the bulk soil. In addition, we are using genomic and proteomic approaches (including stable-isotope techniques) to identify key metabolic pathways responsible for C transformation and mineralization during root in-growth and root death/decay. By integrating stable isotopes as tracers of natural resource utilization (i.e. root litter), and analysis of the functional properties of the communities that respond to those resources, we can identify

the microbial communities that are stimulated in the soil microbiome in response to root litter, living roots, and their intersection.

While there has been much study of the interactions between plant roots and soil microorganisms, we know little about the interactions among the microbial members of the root microbiome and how these relationships change over time. To identify networks of potential microbe-microbe interactions within the rhizosphere, we examined the co-occurrence patterns of bacteria in the rhizosphere soil of wild oat (*Avena fatua*), and compared them to patterns in the surrounding bulk soil. Random matrix theory (RMT) based network analysis was employed to identify such interactions using rDNA gene sequencing over two growing seasons.

Our results revealed that bacterial networks in rhizosphere were substantially more complex than those in the surrounding bulk soils, and the complexity of the bacterial networks increased as the plants grew, even as univariate diversity decreased. In contrast, network patterns in the bulk soils remained relatively static over the experimental period. Increased network complexity coupled to decreased diversity highlights that interactions are a crucial dimension of community organization overlooked by univariate diversity metrics. Within the rhizosphere networks, groups of highly connected modules formed over time, which most likely represent both interactions as well as microbial niches developing in response to root-induced changes of the soil environment. Consistent with the hypothesis that extensive mutualistic interactions occur among rhizosphere bacteria, covariations were predominantly positive (>80%); we identified quorum-based signaling as one potential interaction strategy. Highly connected taxa identified as putative keystone species (module hubs and connectors) often had low relative abundance in the rhizosphere (< 0.1%), which suggests that focusing on abundant taxa may overlook organisms that could play important roles in maintaining rhizosphere community structure and function. Network complexity appears to be a defining characteristic of the rhizosphere microbiome, and is a previously undescribed property of this habitat.

This research is based upon work supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Numbers DE-SC0004730 and DE-SC0010570 at UC Berkeley and FWP SCW1421 at LLNL. Support to UC Berkeley was also provided by EMSL project #48912. Funding was also provided by an LLNL LDRD "Microbes and Minerals: Imaging C Stabilization". Work at LLNL was performed under the auspices of the U.S. Department of Energy under Contract DE-AC52-07NA27344.

Soil Bacteria Transport Systems: Ligand and Regulatory Specificity

Philip D. Laible* (laible@anl.gov), Peter Korajczyk, Eliza L. Zielazinski, Nicholas P. Dylla, Sarah Zerbs, Frank R. Collart and **Philippe Noirot**

Biosciences Division, Argonne National Laboratory, Lemont, IL

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

Project Goals: The Argonne "Environment Sensing and Response" Scientific Focus Area (SFA) program seeks to identify the molecular basis of cellular transport and sensory pathways that mediate the response of terrestrial ecosystems to environmental nutrients. The mechanistic links between and within ecosystems comprised of plants, fungi, and soil bacteria involved in the production of biomass for fuel are currently very poorly defined. The effects of nutrient availability, closely linked to climate, on those mechanistic links, are also inadequately understood. This program will address this knowledge gap by mapping transport and sensor proteins to specific environmental compounds to define their function and biological roles and establish a series of defined connections between the environment and the cell. The knowledge will facilitate the development of system-level models predictive of cellular response to changes in environmental conditions.

A typical bacterial ABC transporter contains three protein components: a solute binding protein (SBP), a transmembrane domain (TMD) dimer spanning the inner membrane, and an ATPase dimer in the cytoplasm (Fig. 1). In bacteria, roughly 2-5% of the genome codes for components of ABC type transporters (*1*). Certain types of soil bacteria contain an even higher percentage of ABC transporters, such as *Pseudomonas fluorescens*, a species that contains well over 350 transport protein components [as identified by Transport DB (*1*)]. Interestingly though, approximately 30% of the ABC transporter proteins in *P. fluorescens* PF-5 are 'orphans' not genomically co-located with other transporter component genes and thus difficult to assign to a specific complex. The abundance of ABC transporter component genes and the presence of orphan genes suggest the potential for crosstalk amongst members. While studies have indicated that TMDs can interact with multiple SBPs with varying affinity and specificity (*2*), there is limited information regarding the specificity of TMD and ATPase protein-protein interactions. The promiscuity of TMD-ATPase interactions may also affect promiscuity of SBP interactions with the transporter complex and refine ligand specificity of the overall transporter. To address this knowledge gap, the SBPs and the membrane-localized transmembrane components of complexes involved in methionine importation from a series of *Pseudomonas fluorescens* strains were investigated *in vitro*. Of significance to these experiments is that ligand binding of methionine to the ATPase domain in an *Escherichia coli* homolog of these systems (MetNI) causes trans-inhibition, where transport is suppressed in a concentration-dependent manner by the substrate.

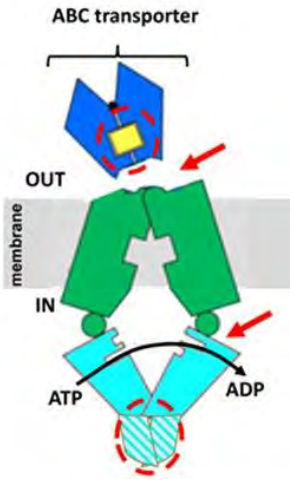


Figure 1. Model of ABC transporters involved in nutrient importation. Red, dashed circles indicate ligand binding sites and red arrows are interfaces where orphaned components may contribute via cross-talk.

Genomic analyses of four *Pseudomonas fluorescens* strains (PF-5, Pf0-1, SBW25, and WH6) have identified a set of ten ATP-binding cassette (ABC)-type amino acid transporters with high sequence similarity to the structurally characterized MetNI from *E. coli*. Recombinant, dual-vector expression as well as polycistronic strategies yield intact complexes of the TMD and ATPase components of the *P. fluorescens* gene targets in *E. coli*. This polycistronic approach allows for limiting tag interference in the complex as well as having more controlled TMD-ABC stoichiometry. Expression experiments systematically varying the combination of ATPase genes with one TMD gene revealed promiscuity of

certain ATPases, where stable "hybrid" complexes could be formed with both a TMD from the same strain as the ATPase or with a TMD from a different *P. fluorescens* strain. This type of functional "crosstalk" between ATPases and TMDs could play a role in rapid nutrient exchange between *Pseudomonas* soil bacteria and other rhizosphere inhabitants. The ability of different ATPases to recognize the same binding site of a TMD also provides an opportunity to study the molecular basis of recognition between the domains of the transporter core.

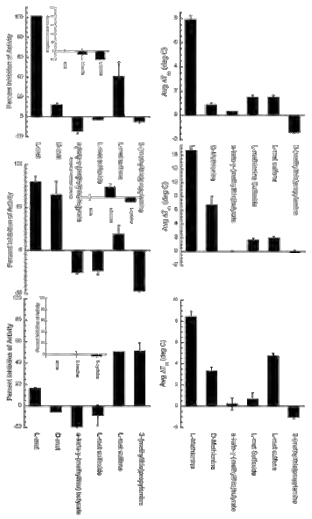


Figure 2. Similarities of ligand binding to SBPs (right panels) and ATPase regulatory domains (left panel) in ABC transporters involved in methionine uptake in soil bacteria. A, *E. coli* MetNI; *P.fluorescens* native ATPase; C, *P. fluorescens* orphaned ATPase.

Few ABC transporters with trans-inhibition regulatory features have been characterized so it is unknown if the regulatory small molecule always matches the transported small molecule. Utilizing ATPase activity inhibition assays, it is possible to identify ligands that have an inhibitory effect on transporter complex function. We determined probable transporter substrates by testing the specificity of associated MetQ-family SBPs using a fluorescence-based thermal shift assay. The ATPase inhibitors were compared with the SBPs ligand-binding profiles to determine if there were patterns between small molecules transported from the environment and the internal regulators of transport activity. Crosstalk at both the external and internal membrane interfaces may have a role in expanding transporter capacity while retaining specificity. Association of non-native ATPases with a TMD could also enable activity regulation by alternative regulatory substrates. Comparison of profiles from the activity inhibition assay and the thermal shift assay for a set of methionine-derivative ligands for native and non-native transporter complexes support these hypotheses (Figure 2). These results demonstrate how multiple complete transporters and orphan components may function together to afford functional advantages to bacteria in complex and highly competitive environments such as the plant rhizosphere. Approaches used for methionine import have been expanded for use in the understanding of how carbon is shuttled between community players within the rhizophere.

References

1. Ren, Q. H., Chen, K. X., and Paulsen, I. T. (2007) TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels, Nucleic Acids Res 35, D274-D279.
2. Thomas, G. H. (2010) Homes for the orphans: utilization of multiple substrate-binding proteins by ABC transporters. Mol. Microbiol. 75, 6-9.
3. Kadaba, N. S., Kaiser, J. T., Johnson, E., Lee, A., and Rees, D. C. (2008) The high-affinity E-coli methionine ABC transporter: Structure and allosteric regulation, Science 321, 250-253.
4. Ferreira, M. J., and de Sa-Nogueira, I. (2010) A Multitask ATPase Serving Different ABC-Type Sugar Importers in Bacillus subtilis, J Bacteriol 192, 5312-5318.

This abstract 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).

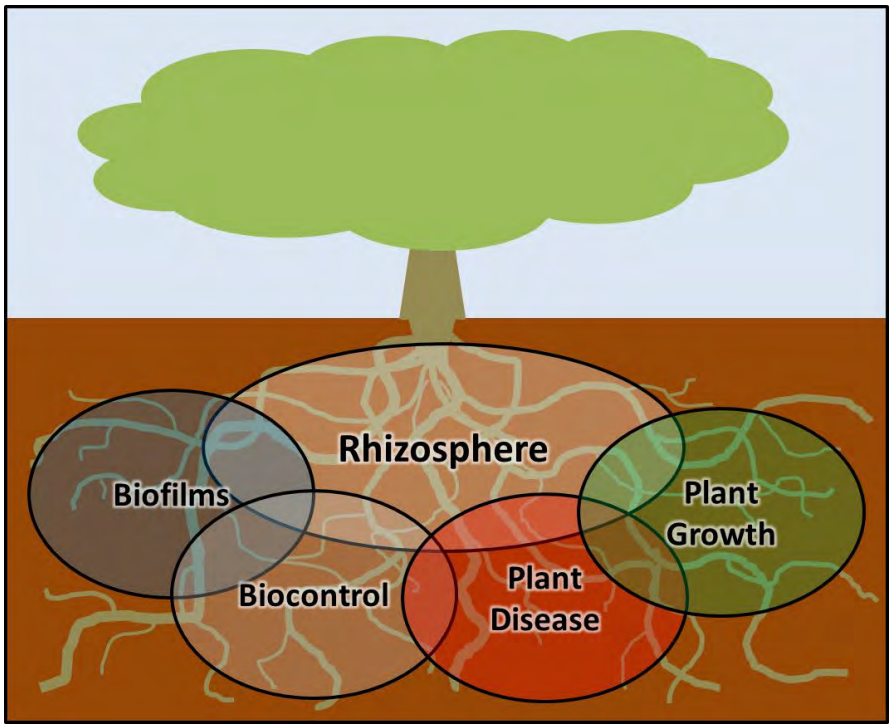
Predicting Pseudomonads' ecological roles in the rhizosphere using machine learning and 'omics' computational modeling

Peter E. Larsen[1,2,*] (plarsen@anl.gov), Yang Dai[2], Mark W. Silby[3], Frank R. Collart[1], **Philippe Noirot**[1]

[1] Biosciences Division, Argonne National Laboratory, Lemont IL; [2] Department of Bioengineering, University of Illinois at Chicago, Chicago IL; [3] Department of Biology, University of Massachusetts Dartmouth, N. Dartmouth, MA

The ability to obtain complete genome sequences from bacteria in environmental samples, such as soil samples from the rhizosphere, has highlighted the microbial diversity and complexity of environmental communities. However, new algorithms to analyze genome sequence information in the context of community structure are needed to enhance our understanding of the specific ecological roles of these organisms in soil environments. We present a machine learning approach using sequenced Pseudomonad genomes coupled with outputs of metabolic and transportomic computational models for identifying the most predictive molecular mechanisms indicative of a Pseudomonad's ecological role in the rhizosphere: a biofilm, biocontrol agent, promoter of plant growth, or plant pathogen. The capacity to form biofilms are indicative of Pseudomonad's capacity to form sessile colonies of plant roots. Biocontrol capacity is Pseudomonads' ability to defend plant roots against plant bacterial, fungal, or animal pathogens. Plant pathogenicity is the ability of certain Pseudomonads to cause disease in plants. Plant growth promotion is the capacity of Pseudomonads to increase plant biomass through providing access to nutrients, remediating abiotic stress, or biosynthesis of plant hormone analogues that influence plant growth.

Application of this model requires the input of sequence, annotated Pseudomonads that can confidently be ascribed to one of the selected rhizosphere ecological niche types. Genomes are re-annotated using a custom database of over 754,000 enzymes and 164,000 transporter annotated protein sequences to insure uniformity of annotations across all genomes. From re-annotated genomes, enzyme function profiles, metabolomic models, and transportomic models are generated. For this analysis, a novel modeling approach to quantify a bacteria's realative capacity to transport specific ligands across the membrane, Predicted Relative Trans-membrane Transport (PRTT) has been developed. These datatypes are used to train Support Vector Machines (SVMs) that can determine membership to rhizosphere niche type. Most predictive



features identified by SVM provide valuable insights into the specific molecular meahcnisms by which Pseudomonads are adapted to the rhizosphere environment and their interactions with plant roots.

Computational predictions of ecological niche were highly accurate overall with models trained on transportomic model output being the most accurate (Leave One Out Validation F-scores between 0.82 and 0.89). The strongest predictive molecular mechanism features for rhizosphere ecological niche overlap with many previously reported analyses of Pseudomonad interactions in the rhizosphere, suggesting that this approach successfully informs a system-scale level understanding of how Pseudomonads sense and interact with their environments. Specific metabolic and transportomic functions are identified that are important for Pseudomonad adaptations to rhizosphere ecological niche types. Transport activities that are identified as predictive for inhabiting the rhizosphere involve carbohydrate transporters (e.g. 2-O-alpha-manosyl-D-glycerate) suggestive for osmoregulation in soils and 3-hydroxyphenylpropionic, one of many lignin breakdown products, which are ubiquitous in soils. Biocontrol is most predicted by its transportome, specifically by transport of cobamide coenzyme, and monosaccharides. The most predictive metabolic activities for biofilm formation are for anti-biofilm compounds protoporphyrin and methyglyoxal. Fatty acid biosynthesis pathways were identified as features predictive for plant pathogenicity in Pseudomonads. Metabolomic input type predicts that synthesis of a number of plant signaling compounds is predictive of plant growth promotion by Pseudomonads including indole and flavones eriodictyol, neringenin. C4-dicarboxylate, calcium, and glutathione transport are also predictive of plant growth promotion The ability to transport of a number of simple sugars (i.e. malonate, mannose, sucrose, galactose, and hexose) is found to be predictive of plant growth promotion by Pseudomonads and is suggestive of an ecological niche that is able to take advantage of exuded photosynthetic sugars present in the rhizosphere. The observation that an organism's transportome is highly predictive of its ecological niche is a novel discovery and may have implications in our understanding microbial ecology. The framework developed here can be generalized to the analysis of any bacteria across a wide range of environments and ecological niches important to carbon cycling and plant-rhizosphere community interactions making this approach a powerful tool for providing insights into functional predictions from bacterial genomic data for DOE mission applications.

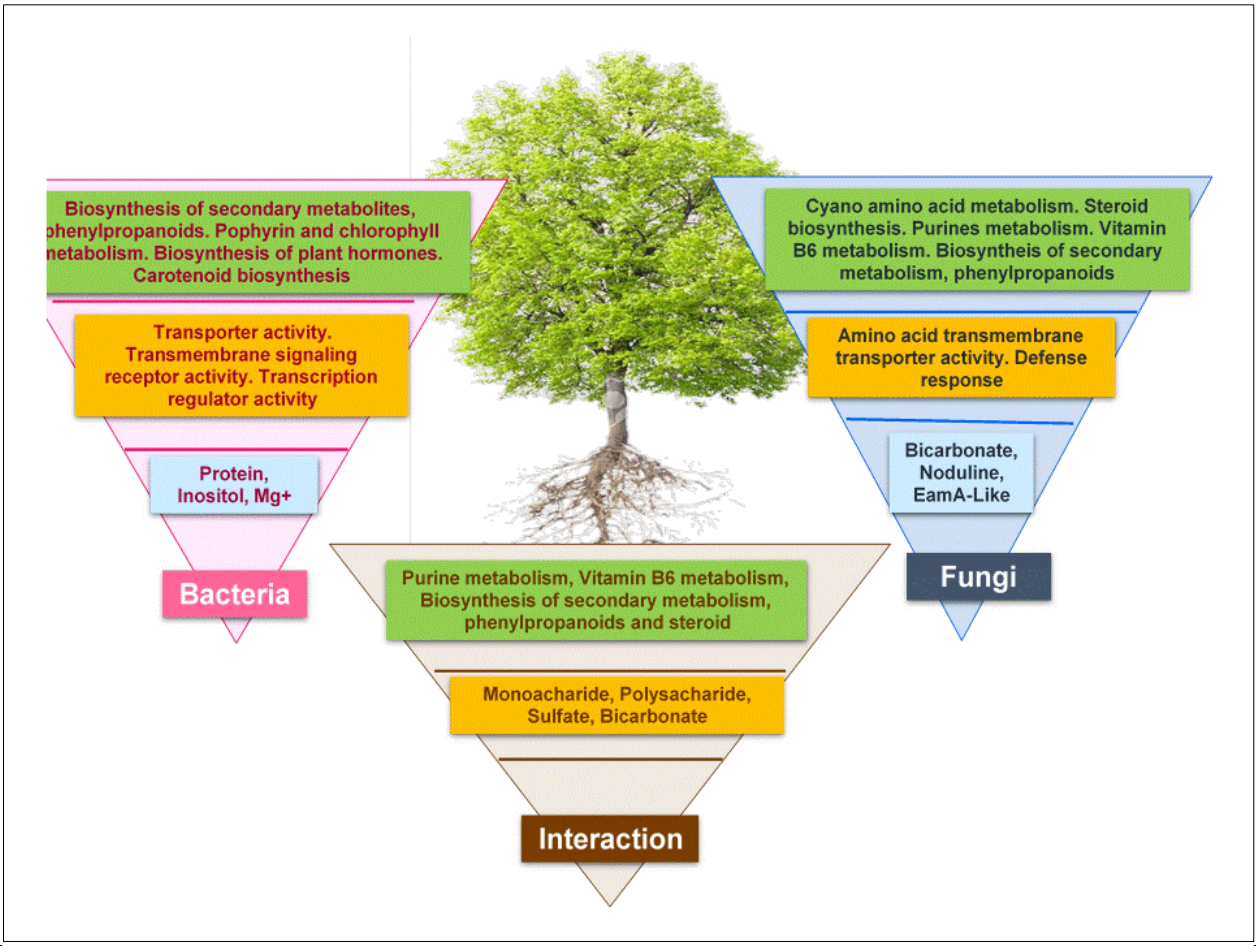
Molecular Mechanisms that Underlie the Beneficial Effects of a Plant-Fungus-Bacterial Community Interaction

Shalaka Shinde[1*] (sdesai@anl.gov), Sarah Zerbs[1], Peter E. Larsen[1], Jonathan. R. Cumming[2], Peter Korajczyk[1], Young-Mo Kim[3], Stephen J.Callister[3], Frank Collart[1] and **Philippe Noirot**[1]

[1]Biosciences Division, Argonne National Laboratory, Lemont, IL, [2]Biological Science Division, West Virginia University, Morgantown. [3]EMSL, Pacific Northwest National Laboratory, Richland, WA

Symbiosis between soil bacteria, mycorrhizal fungi and roots of many ecologically and economically important tree species leads to coordinated resource exchange and enhanced productivity and resiliency in forest ecosystems. Although nutrient cycling and exchange of mineral nutrients are key features of this interaction, very little is known about the molecular mechanisms that underpin this process or the role of these processes in the formation of soil community structure. The Argonne "Environment Sensing and Response" Scientific Focus Area (ESR-SFA) program proposes fundamental research to elucidate the complex plant-microbial interactions between *Populus tremuloides* (Quaking aspen) and its fungal and bacterial symbionts that influence the capture, partitioning, and allocation of carbon (C) under nutrient limitation stresses. We have developed a unique tripartite system comprised of *Populus tremuloides* tree seedlings, *Laccaria bicolor* and *Paxillus involutus* ectomycorrhizal fungi, and *Pseudomonas fluorescens* mycorrhizal helper bacteria, which exhibits beneficial effects on plant biomass under nutrient limitation and is tractable to identify specific molecular mechanisms of community interaction.

A suite of experimental setups is utilized for assessment of phenotypic, biochemical, and physiological responses of plants in the context of transcriptomic, metabolomics and proteomic analyses. This integrated data set--from gene to protein to metabolite to phenotype – will support identification of the molecular mediators of acclimation to stress in a woody plant species and the metabolic networks influenced by rhizosphere communities. We have explored the molecular mechanisms underlying these interactions and have constructed models for the interaction of plants during association with the ectomycorrhizal fungi and/or mycorrhizal helper bacteria. Plant transporters related to inositol, protein and magnesium are up-regulated during bacterial colonization, which can be associated with changes in root growth and morphology and carbon-nutrient exchange at the interface. Plant transporters such as those associated with bicarbonate, auxin/amino acids and potassium are up-regulated during fungal colonization, which can be connected to root structural changes due to mycorrhizae formation and cation exchange capacity in the rhizosphere. This information will be essential to understanding the basis for occupancy of an ecological niche and to define the molecular interactions that occur in communities in changing environments.



The figure shows plants pathways, predicted transportome and secondary metabolites, which are differentially accumulated during bacterial, fungal or bacterial + fungal association. Each triangle represents plants response to bacteria (pink), ectomycorrhizal fungi (blue) and to both microbes (brown). Boxes represent the differentially expressed genes (green), transportome (yellow) and secondary metabolites (light blue). Transportome and secondary metabolites are predicted from omics data modeling as previously described (Larsen PE et al. PLoS One 2015 doi: 10.1371/journal.pone.0132837; see also associated poster).

The molecular function information derived from these studies will guide bacterial genome engineering and synthetic biology approaches aimed at modulating system response by manipulating components of transport or sensory systems. Overall, the capabilities and information derived from the experimental studies will support DOE mission applications in bioenergy, C management, climate change mitigation, and sustainability.

Systems Analysis of the Physiological and Molecular Mechanisms of Sorghum Nitrogen Use Efficiency, Water Use Efficiency and Interactions with the Soil Microbiome

Daniel P. Schachtman[1] (Daniel.Schachtman@unl.edu), Rebecca Bart[2], Thomas Brutnell[2], Daniel Chitwood[2], Asaph Cousins[3], Jeffrey Dangl[4], Ismail Dweikat[1], Andrea Eveland[2], Maria Harrison [5], Stephen Kresovich[6], Peng Liu[7], [2]Todd Mockler, Jessica Prenni[8], Susannah Tringe[9], and Arthur Zygielbaum[1]

[1]University of Nebraska-Lincoln; [2]Donald Danforth Plant Science Center; [3]Washington State University; [4]UNC-Chapel Hill; [5]Boyce Thompson Institute; [6]Clemson University; [7]Iowa State University; [8]Colorado State University; [9]Joint Genome Institute-DOE

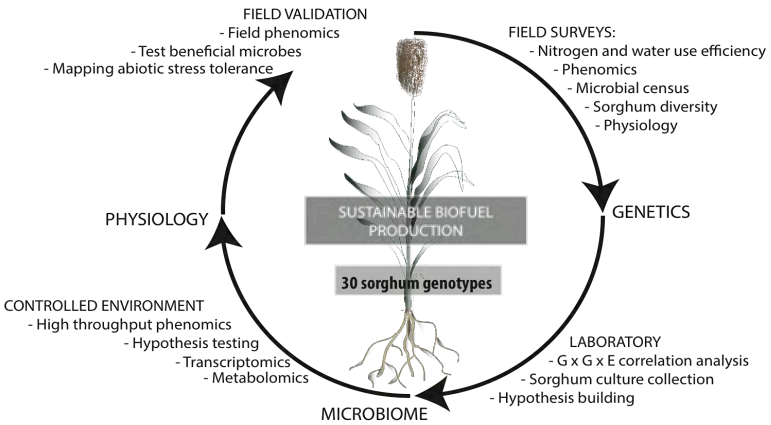
Project Goals:

The *overall project goal* is to establish a foundational, systems-level understanding of plant, microbial, and environmental interactions that will lead to strategies for enhancing growth and sustainability of sorghum through genetic and microbial adaptations to water and nitrogen limited environments.

The *specific goals* are:

1. Conduct deep surveys of root microbiomes concurrent with phenotypic characterizations of a diverse panel of sorghum genotypes to define the associated microbes and the most productive lines under drought and low nitrogen conditions.
2. Associate systems-level genotypic, microbial, and environmental effects with improved sorghum performance using robust statistical approaches.
3. Develop culture collections of sorghum root/leaf associated microbes.
4. Test hypotheses regarding Gsorghum x Gmicrobe x E interactions in controlled environments.
5. Validate physiological mechanisms, map genetic loci for stress tolerance, and determine the persistence of optimal microbial strains under greenhouse and field conditions.

To compete in the biofuel energy market, cellulosic feedstocks will need to be high yielding and carbon neutral or negative while requiring low inputs. To avoid competition with existing food production systems, these crops will also need to be grown on marginal lands. This will require the introduction of novel traits to increase abiotic stress tolerance associated with marginal soils. This project will utilize multiple interdisciplinary approaches in varied settings – including the laboratory, controlled environments, and the field – to identify plant genes and sorghum associated microbes that will enhance the sustainable production of sorghum as a biofuel feedstock. Basic knowledge about physiological and genetic mechanisms involved in nitrogen use efficiency (NUE) and water use efficiency (WUE) and potential mechanisms involved in microbe interactions will be generated. A range of methods will be used, including: classical whole plant physiology, stable isotope detection,



System-wide knowledge and tools to breed sorghum and identify microbes to enhance biomass, NUE and WUE

phenomics, transcript profiling, metabolic profiling, 16S amplicon sequencing, metagenomics, microbial genome sequencing, comparative genomics, microbiology, genetics, and a range of computational methods for data analysis, integration and storage. To conduct these comprehensive studies, we have assembled a multi-institutional, interdisciplinary team with a wide range of expertise in these areas.

This research will increase our knowledge about the genetic and physiological mechanisms involved in WUE and NUE, which will be used to create sustainable biofuel feedstock systems on marginal land. Identification of microbial community membership and testing of culturable microbes, as well as genetic dissection of sorghum genotype X microbe interactions, will result in new strategies for the development of microbial solutions to increase abiotic stress tolerance and sustainable sorghum systems. We will create a sorghum microbe collection which may be the first available for an agricultural crop and a multi-dimensional relational database to house and access the biological materials and data generated in this project.

The project began in October when funding was finalized. Preliminary results will be shared in the poster.

This work is funded by DOE (BER) Award DE-SC0014395

EPICON: Epigenetic Control of Drought Response in Sorghum

Peggy G. Lemaux (lemauxpg@berkeley.edu)[1*], Devin Coleman-Derr[1,2], Jeffrey Dahlberg[3], Robert Hutmacher[4], Christer Jansson[5], Elizabeth Purdom[1], John Taylor[1], Axel Visel[6], and **John Vogel[6]**

[1] University of California, Berkeley CA; [2]USDA-ARS Plant Gene Expression Center, Albany CA; [3]University of California, Kearney Ag Research & Extension Center, Parlier CA; [4]University of California, West Side Research & Extension Center, Five Points CA; [5]EMSL-Pacific Northwest National Laboratory, Richland WA; and [6]DOE Joint Genome Institute, Walnut Creek CA.

Project Goals: EPICON researchers will explore epigenetic control mechanisms in the temporal response to controlled field-based, water-limiting conditions in pre-flowering and post-flowering drought-tolerant *Sorghum bicolor* varieties, and investigate changes in their associated rhizospheric bacterial and fungal microbiomes. EPICON's efforts will focus on unraveling the role epigenetic signals play in acclimation to and recovery from drought through effects on individual transcription factors or transcriptional networks that direct entire metabolic pathways. This research will utilize a wide portfolio of analytical tools, *i.e.*, RNA-Seq, smRNA-Seq, ChIP-Seq, BS-Seq, Orbitrap MS, MALDI-ToF MS, nano-DESI, metagenomics and metatranscriptomics. In performing this work we will identify genes and markers to improve genetic strategies for sorghum and other crops, particularly with respect to drought response. From the cumulative data, we will develop a model to better understand the role of epigenetics and the phytobiome in sorghum's response to drought.

Genetic manipulation of crops to increase the presence or activity of desirable traits has focused primarily on modifications of the plant's DNA sequence. However, there is increasing public research that indicates plant development and environmental responses are also mediated by epigenetics, the process by which heritable changes in phenotype or gene expression are accomplished without changes in the DNA sequence. With particular relevance to the EPICON project, epigenetic changes have been shown to play a major role in regulating plant responses to drought, which is likely to be an increasing problem for world agriculture due to climate change. In general, exposure of plants to abiotic stresses, including water limitation, triggers cascades of epigenetic changes, which include remodeling of chromatin, the network of DNA, RNA and various proteins making up chromosomes, coupled with related changes in regulatory mechanisms, including the involvement of small non-coding RNAs.

This project's research efforts will focus on unraveling the temporal role that epigenetic signals play in acclimation to and recovery from drought through effects on individual transcription factors or transcriptional networks that direct entire metabolic pathways. To achieve these goals, responses to water deprivation will be studied in two sorghum cultivars that differ in their drought responses, pre-flowering versus post-flowering drought tolerance. Sorghum, a widely cultivated cereal, noted for its drought and flood tolerance, offers notable advantages as a bioenergy feedstock because of its relatively reduced environmental footprint compared to its close relative, corn. The study of sorghum's response to drought will be conducted in California's Central Valley; this region's lack of summer rainfall will make well-controlled drought studies in the field possible. Phenotypic analyses will be conducted to chart growth, flowering, grain and biomass yield, and other observable characteristics. Leaf and root samples will be taken at regular intervals during sorghum development to perform molecular phenotyping to track spatiotemporal changes in epigenetic, transcriptomic, metabolomic and proteomic footprints, using RNA-Seq, smRNA-Seq, ChIP-Seq, BS-Seq, Orbitrap MS, MALDI-ToF MS and nano-DESI. As potential molecular mechanisms are identified, targeted engineering will be used to validate suggested findings.

Shifts in the composition and activity of sorghum-associated bacterial and fungal community composition throughout the drought period will also be monitored to determine if changes in membership or functional

capacity within the rhizosphere, root endosphere, and phyllosphere correlate with epigenetic, transcriptional or metabolomic variation in the plant. These investigations into the sorghum microbiome will be done using Illumina itag sequencing of 16S rRNA and ITS2 (internal transcribed sequence) amplicons, specific to prokaryotic and fungal microbes respectively, and via shotgun metagenomic and metatranscriptomic sequencing of rhizosphere communities.

Analysis of the entire data set will provide a better understanding of the epigenetic processes responsible for restructuring the metabolic and regulatory landscape of the sorghum genome, and the relationship of these processes to drought tolerance. These efforts will lead to the achievement of our ultimate goal, which is to identify key transcriptional regulators and pathways that control drought tolerance and to characterize their mechanisms of action, both in the plant and in associated microbial communities. Additionally, these efforts will uncover biomarkers that are associated with drought tolerance, which can be used to monitor and follow phenotypic changes in large populations. The genetic targets and their regulatory pathways will be utilized in future efforts to improve growth and biomass production of sorghum and other crops under water-limiting conditions.

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

***EvoNet*: A Phylogenomic and Systems Biology approach to identify genes underlying plant survival in marginal, low-Nitrogen soils**

Kranthi Varala,[1]* (kranthi.varala@nyu.edu), Rodrigo Gutiérrez,[2] Dennis Stevenson,[3] Robert DeSalle,[4] W. Richard McCombie,[5] Jean-Michel Ané,[6] Heidi Kaeppler[6] and **Gloria Coruzzi**[1]

[1]New York University, New York; [2]Pontifica Universidad Católica de Chile, Santiago, Chile; [3]New York Botanical Garden, New York; [4]American Museum of Natural History, New York; [5] Cold Spring Harbor Laboratory, New York and [6]University of Wisconsin, Madison

http://evonet.org

Project Goals: This DOE BER sustainability project aims to identify the key genes and gene regulatory networks that enable "extreme survivor" plants to adapt and grow in marginal, extremely nitrogen (N) -poor soils in the arid Chilean Andes. These extreme survivor species cover the main branches in flowering plants and include 7 grass species of particular interest for biofuels. We focus on 24 "extreme survivor" Chilean species that are relevant to biofuels and compare their genomes to 24 Californian "sister" species that live in a N replete arid environment. Exploiting a novel phylogenomic pipeline, a "paired species" sampling strategy will identify the genes that support the evolutionary divergence of the extreme survivors in Chile from their sister species in California. The genes thus identified will help to discover the mechanisms underlying physiological and developmental processes that allow efficient assimilation of nitrogen in nitrogen-poor, dry soils. The genes and network modules so uncovered can potentially be translated to biofuel crops to greatly increase biomass and nitrogen use efficiency in marginal, low-fertility soils.

This collaborative project exploits the genomes of "extreme survivor" plants adapted to thrive in marginal, extremely Nitrogen (N) poor soils in the arid Chilean Andes. It uses a previously validated phylogenomic pipeline we developed called BigPlant [1], and a "paired species" sampling strategy, to identify the genes that distinguish these "extreme survivors" in Chile from their related species adapted to similarly dry regions in California (CA) that are not constrained by N. These "extreme survivor" species broadly cover the main branches in flowering plants, and therefore offer a wide range of genomic backgrounds within which the survival traits repeatedly arose i.e., multiple independent origins of trait.

Key to our phylogenomic approach is the "paired species" sampling strategy. This will identify the genes responsible for the divergence of the "extreme survivors" adapted to the marginal low-N soils of Chile, from their most closely related species growing on arid but N-replete soils in CA. To maximize our ability to separate the trait-relevant signature from overall speciation events, our "paired species" sampling will cover multiple independent origins of the low-N adaptive trait. In published studies, we showed that this phylogenomic approach could; i) identify genes that underlie convergent evolution of antioxidant synthesis in Rosids in a study of 150 genomes [1]; and ii) identify 100+ genes associated with the loss of Arbuscular Mycorrhizal

Chile Species (Drought + low-N)	California Species (Drought)
Acantholippia deserticola	*Aloysia wrightii*
Adesmia spinosissima	*Amorpha californica*
Allionia incarnata	*Anulocaulis annulatus*
Ambrosia artemisioides	*Dicoria canescens*
Aristida adscensionis	*Danthonia unispicata*
Bouteloua simplex	*Hilaria jamesii*
Calandrinia sp.	*Montia dichotoma*
Chorizanthe conmisuralis	*Pterostegia drymarioides*
Cistanthe sp.	*Calptridium monospermum*
Cristaria sp.	*Malvella leprosa*
Deyeuxia curvula	*Calamagrostis rubescens*
Euphorbia amandi	*Chamaesyce vallis-mortae*
Festuca sp (chrysophylla)	*Vulpia microstachys*
Junellia seriphioides	*Glandularia gooddingii*
Munroa decumbens	*Dasychloa pulchella*
Nassella nardoides	*Achnantherum occidentale*
Nototriche sps.	*Sphaeralcea rusbyi*
Senecio puchii	*Pluchea sdericea*
Sisymbrium sp.	*Stanleya pinnata*
Stipa frigida	*Jarva speciosa*
Tagetes multiflora	*Pectis papposa*
Tarasa operculata	*Malacothamnus fremontii*
Urbania pappigera	*Verbena menthifolia*
Werneria sp.	*Lepidospartum squamatum*

Table 1. Extreme survivor species in Chile (green) and paired "sister" species in CA (yellow). *Our project studies 24 pairs of species from Marginal (Dry +low-N, Chile) and Dry soils (Dry, California). All the Chilean species have already been sequenced while California species are being collected.*

(AM) symbiosis in the Brassicaceae [2]. We now extend this approach to the study of "Marginal survivor strategies" as follows:

Aim 1. Sample transcriptomes of 24 "extreme survivors" & paired species (NYU, NYBG, Chile). ***Progress:*** We sequenced all 24 species from Chile (Table 1) and are currently sampling the sister species collected in the field in California.

Aim 2. Phylogenomic analysis of 48 "paired species" to identify genes that repeatedly support nodes that distinguish the extreme survivors in Chile from their sister species in CA (AMNH, NYU). ***Progress:*** We have adapted our BigPlant phylogenomic pipeline to improve speed, while reducing the memory footprint.

Aim 3. Combine phylogenomics (protein sequence) and gene networks (gene expression) to identify genes and network modules associated with adaptations to marginal, low-N soils (NYU, Chile). ***Progress:*** To exploit a comparative analysis of gene regulatory networks, we are developing a new module **PhyloExpress** that extends the BigPlant pipeline to include gene expression data.

Aim 4. Functionally validate top-ranked candidate genes for low-N adaptation in Arabidopsis and Brachypodium (NYU, Chile, U Wisconsin). ***Progress:*** We are transforming Brachypodium with the most promising candidate from our preliminary analysis of 5 species pairs.

References

1. Lee E et. al., A functional phylogenomics view of the seed plants. PLoS Genet 7(12):e1002411.
2. Delaux et. Al., Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. PloS Genet 10(7):e1004487.

The EvoNet grant is supported by the Genomics Science program within the Office of Biological and Environment Research in the DOE Office of Science.

Climate adaptation and sustainability in switchgrass: exploring plant-microbe-soil interactions across continental scale environmental gradients

Thomas E. Juenger,[1,*] (tjuenger@austin.utexas.edu) Katherine Behrman,[1] Jeremy Schmutz,[2] Tanja Wolke,[3] Felix Fritschi,[4] Alina Zare,[4] Laura Bartley,[5] Julie Jastrow,[6] Sarah O'Brien,[6] Roser Matamala,[6] Sean Watson,[7] Denise Costich[8]

[1]University of Texas, Austin;[2] HudsonAlpha Institute for Biotechnology, Huntsville, Alabama; [3]DOE Joint Genome Institute, Walnut Creek, California; [4]University of Missouri, Columbia; [5]University of Oklahoma, Norman; [6]Argonne National Lab, Argonne, Illinois; [7]Ladybird Johnson Wildflower Center, Austin, Texas; [8]International Maize and Wheat Improvement Center

https://sites.cns.utexas.edu/juenger_lab/switchgrass

Project Goals: Our collaborative project is focused on understanding switchgrass genetic diversity and adaptation across continental scale environmental gradients. Our goal is to improve the sustainability of switchgrass as a biofuel by gaining insight into the interaction of switchgrass diversity with its associated microbiome and environmental conditions. Our approach will involve 1) the collection and characterization of new switchgrass germplasm from across the species range, 2) the development of a genetic association mapping panel and extensive common gardens to evaluate switchgrass performance, 3) a detailed characterization of the switchgrass microbiome, 4) studies of the impact of switchgrass stands on ecosystem processes, and 5) extensive multiscale modeling to define conditions of sustainability and identify key tradeoffs between genetic diversity, productivity, and ecosystem services. Ultimately, these studies will identify critical plant-microbe-soil traits that may be manipulated, through breeding or agronomic management, to improve the sustainability of biofuel feedstocks.

Less carbon-intensive energy sources are needed to reduce greenhouse gas emissions and their predicted role in climate change. There is growing interest in the potential of biofuels for meeting this need. A critical question is whether large-scale biofuel production can be sustainable over the time scales needed to mitigate our carbon debt from fossil fuel consumption. The carbon balance and ultimately the sustainability of biofuel feedstock production is the result of complex climate-coupled interactions between carbon fixation, sequestration, and release through combustion. The long-term productivity of biofuels depends on the genetic and environmental factors limiting plant growth. These factors are often related to soil resources which involve complex interactions at the plant-microbe-soil interface impacting their availability and cycling.

Our collaborative project will address sustainable switchgrass (*Panicum virgatum*) production through a detailed characterization of plant growth and performance in both individual spaced and stand plantings. The project represents an unprecedented field-based experimental system for a bioenergy grass. We bring together diverse skill sets from plant and microbial genetics and genomics, physiology and ecosystem modeling. An underlying theme of the research is the use of locally adapted plant material to explore plant function, to understand the mechanistic basis of

environmental interactions, and to discover the plant genes important for adaptation and sustainability in the face of climate change. To this end, we have been collecting new genetic diversity in switchgrass from natural populations across the species range in North America. These new genotypes are being characterized through genome resequencing and clonally propagated for inclusion in experimental gardens at 14 field locations. This material will provide a detailed population genomic characterization of switchgrass along with resources for association mapping and genomic selection for future breeding programs.

Our plant-microbiome project will fully characterize the microbial communities associated with switchgrass at our planting locations using genomic tools. Specifically, we will sample field plantings of switchgrass for leaf and root microbial communities with 16S iTAG and metegenomic sequencing in collaboration with the DOE Joint Genome Center. Analyses will center on quantifying the relative importance of switchgrass host genotype, the planting environment, or their interaction on microbial community composition. Ultimately, these data will be linked with plant trait information to evaluate drivers of plant-microbiome interactions and their impacts on ecosystem processes.

Our ecosystem processes research will focus on carbon cycle responses at the ecosystem level using stand plantings of switchgrass diversity. We will couple plant growth and physiology measurements with measurements of key carbon pools and fluxes that integrate carbon dynamics across different temporal and spatial scales. Finally, our modeling will define conditions of a sustainable biofuel system and identify key tradeoffs between genetic diversity, productivity, and ecosystem services. Here, we highlight the background resources for our project, our progress establishing our new experiments, and plans for the coming field season.

This research is supported by funding through the Office of Biological & Environmental Research within the Department of Energy Office of Science and through the DOE Joint Genome Institute Community Science Program.

An Overview of the Switchgrass Associated Microbes Belowground with Mutualistic Fungi and Soil Mesofauna as Potential Models for Sustainable Cultivation

Nhu H. Nguyen[1*] (xerantheum@berkeley.edu), Eoin Brodie[2], Kelly Craven[3], Javier Ceja Navarro[2], Erin Nuccio[4], Jennifer Pett-Ridge[4], and **Mary Firestone[1]**

[1]University of California, Berkeley; [2]Chemical Sciences Division, Lawrence Livermore National Laboratory, [3]The Samuel Roberts Noble Foundation, [4]Earth Sciences Division

Project Goals: Our project, *Establishment to senescence: plant-microbe and microbe-microbe interactions mediate switchgrass sustainability,* aims to understand the bases of switchgrass productivity and potential biotic and abiotic effects in marginal soils by dissecting the key molecular mechanisms that differentiate soil organisms associated with superior switchgrass genotypes adapted to a range of resource limitations. Through our research we seek to provide a mechanistic understanding of the network of interactions that exists within the switchgrass root system, with the overall goal of understanding and synthesizing concepts that underlie soil food webs and their effects on ecosystem sustainability.

Successful cultivation of the biofuel crop switchgrass (*Panicum virgatum*) to produce high-yield biomass in marginal soils had been identified as an important goal to meet the US Department of Energy's goal to replace 30% of petroleum-based transportation fuels with biofuel by 2030. An important goal in this area is to identify ways to alleviate biotic and abiotic stresses while enhancing survival and growth during seedling establishment in marginal soils. We suggest that promoting a beneficial holobiome (consisting of bacteria, archaea, protists, fungi, nematodes and arthropods) is a logical and potentially sustainable way to support successful switchgrass establishment and sustainable cultivation. The network of interactions within this holobiome may provide key ecosystem services ranging from C sequestration and increased soil fertility. Yet, our understanding of the occurrence and regulation of mutualistic switchgrass-microbe interactions, carbon (C) and nutrient exchange between roots, soil food webs, rhizosphere C, nitrogen, phosphorus, and water cycles remains rudimentary. Over the next five years of this project, we aim to identify the holobiont of the switchgrass root system, and use and combination of 'omics and isotope techniques to understand the network interactions and nutrient dynamics.

We present an initial study of the switchgrass root system through characterization of the community of associated bacteria, archaea, and fungi from rhizosphere, rhizoplane, and bulk soil from 2 established field sites in Oklahoma. The soils from these sites represent nutrient limited areas in which future field experiments will be conducted. The results of the microbial community structure, including major microbial components, and soil characteristics from these sites will be discussed. The results of this study will inform our targets for designing a set of primers that will be used to amplify the holobiome of the switchgrass root-system, which will then be compared with deep sequencing of soil metagenomes.

We also present an overview of a set of experiments aimed to dissect plant-microbe and animal-microbe interactions in the context of nutrient solubilization and allocation. To study the interactions between switchgrass and microbes, we will collect rhizospheric soil as well as root

tissues in which to isolate bacterial and fungal endophytes. We have established high-throughput screens for many potential useful microbial traits, including N-fixation, ACC deaminase activity, and solubilization of inorganic phosphorus (CaPO4). Endophytic bacterial or fungal strains found to contain one or more will be candidates for stable isotope probing (SIP) tracer studies and simplified community modeling and analysis.

We will also test the nutritional exchange between two types of mycorrhizae that are found on the roots of native switchgrass plants, the sebacinoid (SM) and arbuscular (AM) mycorrhizal fungi. SIP will be used to quantify the movement of phosphorus and nitrogen from the bulk soil through these two beneficial fungi, and into establishing switchgrass plants. While AM are known to extract a steep carbon cost to the host plant, we predict that sebacinoid fungi, themselves competent saprobes and scavengers, will require significantly less photosynthetically derived carbon. Thus, we will in parallel track the movement of labeled carbon from the plant into the symbiotic fungi.

To study the interactions between animals and microbes in the context of the switchgrass root system, we have established the protocols for the study of the soil mesofauna and their associated gut microbial communities. These approaches include the separation of metazoan from soil samples using Burlese and Baermann funnels, followed by grouping into morphospecies. Individual specimens are used for high throughput DNA extraction and for the amplification and sequencing of eukaryotic and prokaryotic biomarkers (16S, 18S rRNA, and Cytochrome Oxidase I genes). These approaches are being combined with large-volume soil DNA extractions to further assess the composition and distribution of soil mesofauna under different experimental conditions.

We will combine analytical chemistry and spectroscopic (NMR, μXRF) methods to characterize and quantify the key phosphorus species from gut microbes. Similar to the endophytic microbe study described above, gut microbial isolates will be assessed for P solubilization activity by organic acid production, enzyme secretion and siderophores production on microbial growth media based on switchgrass root exudates and insoluble P (iron-, aluminum-phosphate, phytic acid) that we have developed. Effective P solubilizers will be sequenced and identified.

We will develop a novel framework for constructing directional networks for discerning the network interactions among different groups of microbes. As a proof of principle, we applied this new framework to *Saccharomyces cerevisiae* gene expression data, taking advantage of the availability of the information on the directions of gene interactions. Our results showed that among directed links predicted by this approach, 88.2% of the links are in the same direction to known genes in SGD (Saccharomyces Genome Database). These results demonstrate that the new method is capable of revealing biologically meaningful causal interactions in yeast. We predict that this method will be useful in discerning the causal network interactions in microbial communities when applied to metagenomes.

Funding
This research is based upon work supported by the U.S. Department of Energy Office of Science number DE-SC0014079 to UC Berkeley, Samuel Noble Foundation, University of Oklahoma, Lawrence Berkeley National Lab, and Lawrence Berkeley National Lab.

Connecting nitrogen transformations mediated by the rhizosphere microbiome to perennial cropping system productivity in marginal lands

Sarah Evans (evanssa6@msu.edu), Lisa Tiemann, Maren Friesen, James Cole

Michigan State University

Project Goals: We will perform a robust assessment of components of the rhizosphere microbiome-plant system by linking microbial and plant genomics and transcriptomics to N-cycling and C allocation and physiology across sites and over time (Fig. 1). We aim to understand how the rhizosphere microbiome associated with perennial biofuel crop systems (PBCS) in marginal lands:

1) **Mediates N transformations important to plant N availability;**
2) **Acquires C resources from plants versus SOM to fuel N-transformations;**
3) **Interacts with plant root exudation patterns and physiological pathways;**
4) **Varies by cropping system and environment.**

The demand for energy from biofuel production is increasing, prompting concerns about the environmental impact and long-term sustainability of bioenergy cropping systems. While many recent life cycle analyses of bioenergy sustainability focus on soil organic matter (SOM) accrual and overall carbon (C) budgets, there has been less attention paid to nitrogen (N) dynamics in these systems. N is the most commonly limiting nutrient for plants, especially in marginal lands that are unsuitable for food crops because of low productivity and vulnerability to environmental stress. The introduction of perennial bioenergy cropping systems (PBCS) in marginal lands can improve whole system N use efficiency and N retention. However, **little is known overall about N-cycling and associated microbial function in marginal land biofuel cropping systems**. It is particularly important to understand the mechanisms regulating nutrient acquisition by microbes and plants, as well as SOM accrual, if we are to maximize the productivity and C benefits of these systems. **We aim to understand how the rhizosphere microbiome in PBCS on marginal lands:** 1) mediates N transformations and availability; 2) gains C resources to fuel N-transformations; 3) affects PBCS productivity and SOM accrual through trade-offs between plant C allocation to support N-transformations versus biomass accrual and; 4) changes by cropping system and environment (Fig. 1).

We will investigate experimental questions related to this goal using sites recently established as part of the DOE Great Lakes Bioenergy Research Center (GLBRC) in the upper Midwest (Fig. 2). On each site, six perennial crops have been established in 20 x 20 m plots replicated in four blocks together with an unplanted reference plot. We will take measurements on the reference plots, switchgrass, miscanthus, a 5-species native grass mix and a restored prairie consisting of an 18-species assemblage of native forbs, grasses, and legumes. In addition, we will sample fertilized sub-plots that will allow us to more fully explore N dynamics.

To address our objectives, we will use a systems approach that links microbial N cycling genes to N fluxes. We will first assess functional gene diversity via metatranscriptomes at each site over 3 years, followed by seasonal assessment via HTqPCR of the diversity of functional genes and their transcripts. To do this we will employ a new HTqPCR system (SmartChip Real-Time PCR Cycler, WaferGen Biosystems Inc., USA) followed by sequencing of the barcoded PCR products directly on the Illumina MiSeq platform in the MSU Core Genomics Facility. We will also measure gross N flux rates using standard biogeochemical assays and ^{15}N tracers to evaluate N transformations and plant N sources in conjunction with measures of gene diversity and expression. We will also perform two experiments alongside our field monitoring efforts: 1) $^{13}CO_2$ pulse-chase experiments in subplots of the field experiment to measure total root exudate production. 2) We will also determine the chemical composition of exudates by isolating and analyzing the root exudates of different bioenergy crops under sterile lab

conditions using LCMS and GCMS. Exudate chemicals we identify will be used to make synthetic exudate solutions that can be isotopically labeled and applied to soils in both the greenhouse and the field. Synthetic exudates will allow us to test the effects of different chemical compositions on microbial community composition, physiology, and N-cycling processes.

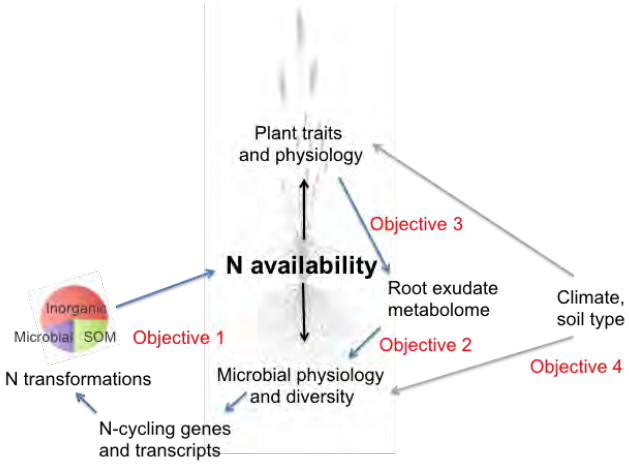


Figure 1. Conceptual illustration of project objectives

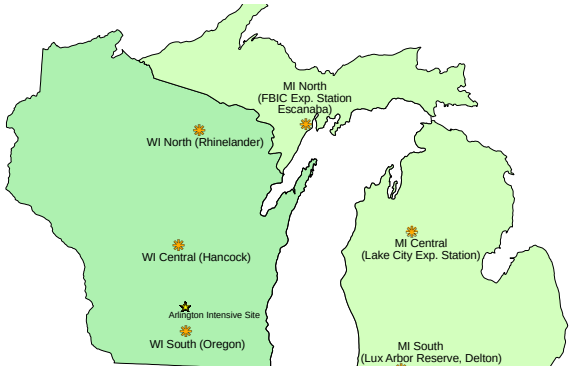


Figure 2. Marginal Land Experiment Sites, part of the DOE Great Lakes Bioenergy Research Center in the upper Midwest, USA

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

Plant-Microbe Interfaces: Linking diversity and function in fungal communities associated with *Populus trichocarpa*

Chongle Pan[1*](panc@ornl.gov), Zhou Li[1], Brian P. Looney[1,2], Sunny Liao[3], Jay Chen[1], Annegret Kohler[4], Francis Martin[4], Gerald A. Tuskan[1], **Jessy Labbé**[1] and **Rytas Vilgalys**[3]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]University of Tennessee, Knoxville, TN; [3]Duke University, Durham, NC; [4]INRA, Tree-Microbe Interactions, UMR1136, Champenoux, France

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

http://PMI.ornl.gov

This project aims to utilize novel sampling efforts and analyses to characterize the associated fungal community of *Populus trichocarpa* and link these fungi to their diverse functions. To this end, we focus on 1) sampling ectomycorrhizal fungi (EMF) from natural *P. trichocarpa* populations along a latitudinal gradient to create a culture collection and gather relevant material for identification and targeted –omics sequencing; 2) metatranscriptomic sequencing of bulk soil and root tips to analyze functional diversity within functional guilds of *Populus* with a focus on symbiotic and endophytic fungi; and 3) genomic sequencing of novel *Populus* associates to characterize the ectomycorrhizome of *Populus* and extend *Populus-Laccaria* findings to new model systems for future experimentation.

Previous sampling efforts focused on assembling a unique culture collection, representative of root endophytic and rhizospheric fungi. To expand this collection to mycorrhizal fungi, we focused on the collection, isolation, and characterization of major EMF fungal associates from a variety of *P. trichocarpa* genotypes from across major watersheds in the Pacific Northwest. EMF fungi are essential mutualistic partners with *Populus* that have shown to significantly expand nutrient uptake and acquisition of the plant host while also providing protection against antagonistic parasites. We recently conducted sampling of several watersheds to increase the representation of EMF in the collection and to enable detailed studies of functional diversity within this ecological guild. This effort consisted of five core watersheds on the Squamish (BC), Snohomish (WA), Puyallup (WA), Columbia (OR and WA), and Willamette (OR) rivers. This resulted in over 100 EMF sporocarp collections and sampling of bulk soil from 8 different *Populus* sites. All sporocarp collections were plated on modified Melin-Norkrans medium,

photographed, spore printed, and dried for identification and accession into a fungal herbarium. A consortium of taxonomic experts assessing morphological features and using ITS sequencing is currently identifying species. Taxonomic groups that were found associated with *Populus* include Russulaceae (18 coll.), *Inocybe* (12 coll.), *Boletus* (4 coll.), *Laccaria* (4 coll.), *Hebeloma* (3 coll.), *Amanita* (3 coll.), *Tricholoma* (2 coll.), *Scleroderma* (1 coll.), and *Clavulina* (1 coll.). Russulaceae comprised the highest taxonomic diversity of any EMF lineage sampled with at least 10 distinct species confirmed as *Populus* associates. Three of these members are host-restricted to *Populus*. Some species of EMF, such as *Inocybe geophylla*, were found at all sample sites, indicating that they may be ubiquitous associates of *P. trichocarpa*. We were also able to confirm *Laccaria bicolor* as a natural associate of *Populus* in a monodominant stand. Sterile tissue from the center of sporocarps was removed and frozen for transcriptome and genome sequencing. Representative isolates are being prepared for full genomic and transcriptomic sequencing at the DOE-JGI.

Many plant-associated fungi host symbiotic endobacteria with reduced genomes. While endobacteria play important roles in plant/fungal/endobacterial systems, the active physiology of fungal endobacteria has not been elucidated. We used integrated proteomics and metabolomics to characterize the endobacterium *Candidatus Glomeribacter* sp. and its host, the root endophytic fungus *Mortierella elongata*. In a nitrogen-poor condition, *M. elongata* had constrained growth, but hosted a large and growing endobacterial population. The active endobacterium likely extracted malate from the fungal host as the primary carbon substrate for energy production and biosynthesis of phospho sugars, nucleobases, peptidoglycan, and several amino acids. The endobacterium obtained nitrogen by importing a variety of nitrogen-containing compounds. Furthermore, nitrogen limitation significantly perturbed the carbon and nitrogen flows in the fungal metabolic network. *M. elongata* regulated many pathways by concordant changes on enzyme abundances, post-translational modifications, reactant concentrations, and allosteric effectors. Such multimodal regulations may be a general mechanism for metabolic modulation.

With this additional sampling and the development of integrated –‘omics analyses, we are positioned to address the fundamental hypothesis that molecular signaling pathways are conserved among different microbial and plant cohorts. The mycorrhizal lifestyle is thought to have arisen multiple times on the evolutionary tree, and we will determine whether similar mechanisms have as well. We aim to investigate the specific genomic determinants in fungi and plants that contribute to the symbiosis through comparative genomics, transcriptomics, proteomics, and metabolomics analyses, and, ultimately, characterize key components (e.g., nutrient distribution and C/N dynamics) of the mutualistic lifestyle.

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

Plant-Microbe Interfaces: High-Throughput Comparative Genomics for *Populus*-associated Microbes

Visanu Wanchai, Se-Ran Jun, Intawat Nookaew, Miriam Land, Preecha Patumcharoinpol, Michael Leuze, and **David W. Ussery*** (usserydw@ornl.gov)

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

http://PMI.ornl.gov

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

We have characterized 179 *Populus*-associated bacterial genomes. Based on full-length protein alignments, we find a core of only 11 protein families conserved among all 179 genomes, and a total of 294,668 unique protein families. Based on a larger set of more than 70,000 available genome sequences, we find that, within a given bacterial species, as much as 80% of the proteins in any one genome belong to the core. The core-genomes for *Populus*-associated bacterial genomes can be used to estimate basic metabolic pathways for a given species, which can be contrasted with strain-specific additional metabolic capabilities. To further explore functional space, we found a total of 7,343 unique protein functional domains in the 179 proteomes, with 47,075 architectures. In the larger set of 70,502 bacterial genomes, we find 12,496 total unique PfamA domains (12,496 out of 16,230 total domains in the database, or 77% of all domains). By storing all of the PfamA domain information in a graph database, we can quickly identify (within a few seconds) transcription factors unique to a specific taxonomic group, as well as from PMI genomes associated with a specific environment (e.g., endosphere, rhizosphere, tree). For example, we find that for the 179 PMI proteomes, 1,047 of the 18,225 architectures (5.7%) contain transcription factor domains, and a slightly larger number was found for all 70,502 genomes (14,111 out of 195,778, 7.2%). This large fraction of transcription factor architectures implies an evolutionary selection for diversity of regulatory proteins. Finally, we also examine sigma factors across all of the genomes.

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

Plant-Microbe Interfaces: The *Populus* Microbiome Atlas Project - Dissecting the microbiome landscape of trees from soil to canopy.

M. Cregger[1], A. Veach[1], W.N. Cude[1,3], M. Crouch[1], Z. Yang[1], T-Y. Lu[1], P. Busby[4], K. Hameed[4], R. Vilgalys[4], T. Rials[5], C. Pan[1], I. Nookaew[1], D. Ussery[1], M. Podar[1], S. Tringe[6], G. Tuskan[1], D. Pelletier[1], and **C. Schadt**[1]* (schadtcw@ornl.gov)

[1]Oak Ridge National Laboratory, Oak Ridge, TN. [2]Colorado State University, Fort Collins, CO [3]Novozymes Corporation, Raleigh, NC, [4]Duke University, Durham, NC, [5]University of Tennessee Institute of Agriculture, Knoxville, TN, [6]Joint Genome Institute, Walnut Creek, CA.

http://PMI.ornl.gov

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

Populus spp. (Poplar, Cottonwood or Aspen) are broadly distributed in temperate North American habitats, making them ideal ecological model species. Additionally, their highly developed genetic toolsets, rapid growth rates, association with ecto- and endomycorrhizal fungi, and ability to grow on land not suitable for food production make them good candidates for bioenergy production scenarios. *Populus* spp. are also now emerging as model systems for understanding the role of the plant microbiome. Most of our and other's work to date on tree microbiomes has focused on comparisons of individual habitats within and across tree species or a few habitat types such as root endosphere, rhizosphere, mycorrhizae, or leaf endosphere versus phyllosphere communities. Surprisingly, a comprehensive comparison of the overall phytobiome of such woody tree species across tissue and habitat types from the soil to the canopy has been lacking. Additionally, high host DNA background levels have limited the ability of microbial ecologists to apply shotgun metagenomic techniques to the sequencing of host endophytic habitat types.

In our ongoing work, we are examining 30 different plant tissue/habitat types across five *Populus deltoides* and five *P. trichocarpa x deltoides (*TxD*)* hybrids (replicated identical genetic clones) collected from destructive whole tree harvests in East Tennessee in August of 2014. Microbiomes of these tissue-level habitats are being analyzed by 16S rRNA bacterial/archaeal amplicons and fungal ITS2-rRNA amplicons for community comparisons across 300 total samples. These samples encompass multiple belowground tissue types (e.g., fine and coarse roots, rhizosphere, soil - from deep and shallow soil locations), aboveground tissue types (e.g.,

wood, live xylem, bark/phloem/cambium tissues – from large structural roots and each of the 3 main stem age segments), as well as leaf endosphere, petiole and phyllosphere samples from developing and mature leaves. For select belowground tissues and habitats, we are also applying new differential centrifugation methods to enrich the microbial cells from root tissues prior to DNA extraction and Nextera-based metagenomic sequencing in order to avoid host DNA background contamination. These metagenomes have enabled us to contrast the fine root endosphere and rhizosphere communities as well as bulk soils in the shallow rooting zone of each *P. deltoides* and TxD hybrid tree (30 metagenomes).

To date we have completed the bacterial amplicon sequencing for all samples and the majority of metagenome samples (28/30). Across each tree type, bacterial community structure from rRNA gene amplicons varied significantly across leaf, stem, roots and soil/rhizosphere tissue/habitat types ($p < 0.01$). Leaf and stem habitat types had significantly lower OTU richness compared to root and stem habitats and featured decreasing abundance of Proteobacteria from leaf, to stem, to root and finally to soil. The oldest woody stem tissue (3-year-old heartwood) was also distinguishable from the younger 1st and 2nd year tissues, and featured enrichment of Firmicutes, consistent with potentially anaerobic/fermentative taxa in this habitat. Within the leaf, stem, root, and soil habitats, bacterial community structure in *P. deltoides* samples could also be differentiated from TxD hybrid samples ($p < 0.01$). Belowground rhizosphere and bulk soil habitats could be differentiated by soil depth ($p < 0.01$). Fungal ITS2 amplicon sequencing of these same samples has been optimized to incorporate improved primer designs for phylogenetic representatives among fungi. In addition, we have incorporated a newly designed ITS2 PNA blocker that reduces host nuclear contamination from >90% in many tissue samples to <1%. Using these optimized protocols, we were able to reduce host contamination to <10% of metagenome reads in 9/10 root endosphere metagenomes samples. Comparative metagenomics-based analyses between tissue/habitat types (e.g., soil, rhizosphere, root endosphere) and with amplicon-based datasets are ongoing.

When complete, these studies will allow us to comprehensively dissect the plant microbiome both phylogenetically and functionally. The results from this work should greatly enhance our understanding of plant microbiomes in general, and the important model species *Populus* in particular.

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

Plant Microbe Interfaces: Emerging Analytical Techniques for Controlling and Monitoring Structural Changes in Developing Multi-Kingdom Systems

Jayde Aufrecht,[1,3] Julie Hoang,[1] Jessie Uehling,[2] Larry Millet,[1] Ryan Hansen,[5] Collin Timm,[1] David Allison,[3] **Scott T. Retterer**[1,3,4]* (retterert@ornl.gov), Jessy Labbé,[1,3] Rytas Vygalys,[2] Jennifer Morrell-Falvey,[1,3] Dale Pelletier,[1] and Mitchel J. Doktycz[1,3,4]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Biology Department, Duke University, Durham, NC.; [3]University of Tennessee, Knoxville, TN; [4]Center for Nanophase Materials Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; Chemical Engineering, Kansas State University, Manhattan, KS

http://PMI.ornl.gov

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

Plant growth and the structure of its associated microbial community are mediated by complex physical cues and chemical signals exchanged between the different organisms. These interactions mediate the flow of chemical information, raw materials, and energy resources, shaping and being shaped by the physical architecture of the system in a continuous feedback loop. While molecular genetics and systems biology approaches reveal the genetic content and molecular signals that underpin these interactions, new methods that address the importance of spatial organization in modulating such interactions are essential. These emerging methods can help elucidate the role of environmental heterogeneity, niche size, connectivity, and solute transport on fluctuating microbial populations as well as architectural and mechanical changes in root structure. Here we describe efforts to examine interkingdom signaling between plants, bacteria, and fungi using a combination of atomic force microscopy, advanced optical imaging, biopatterning, and nanostructured fluidic environments with the goal of recreating the complex and emergent behaviors found in natural systems. The development of model fluidic environments combined with quantitative imaging and analysis techniques for capturing temporal information about microbe localization, fungal hyphae growth, and changes in root structure will be described.

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

Plant-Microbe Interfaces: Constructed communities of *Populus* and bacterial isolates to study microbiome function

Collin Timm[1], Alyssa Carrell[1], Sara Jawdy[1], Lee Gunter[1], Zamin Yang[1], Jayde Aufrecht[1,] Shen Lu[1], Nancy Engle[1], **Timothy Tschaplinski**[1], **Dale Pelletier**[1*](pelletierda@ornl.gov), **Dave Weston**[1], Gerald A Tuskan[1], Mitchel Doktycz[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

http://PMI.ornl.gov

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

Functions of a plant-microbiome system are the result of complex interactions between microbiome members and the host plant. Direct experimentation on the microbiome is difficult due to the inability to culture the microbiome in the laboratory. Our solution is to study constructed communities designed to represent the microbiome, which consists of diverse, cultivable members isolated from natural systems. Our isolate collection consists of >2700 bacterial strains, of which ~200 have been genome-sequenced to date through a JGI CSP. Using genomic content, strain functional data, and microbiome community data as a guide, we selected a community of diverse microbial isolates to colonize axenic *Populus* cuttings in microbiome replacement experiments.

Our first goal was to identify the contribution of individual members to host phenotype in a community environment. A *Pseudomonas* representative from our isolate collection was chosen based on several factors including 1) its ability to produce the plant hormone indole-3-acetic acid; 2) the ability to grow using plant metabolites sucrose and 3-hydroxybenzoate as sole carbon sources, and 3) the enhancement of root hair production in *Arabidopsis*. The second bacterial isolate, from the *Burkholderia* genera, has predicted enzymes for growing on multiple carbohydrate sources and it colonized *Populus* cuttings at a high density (10^8 CFU/g root). When inoculated on *Populus* cuttings, these strains increased root growth relative to uninoculated controls. The enhanced root growth in the dual-inoculated samples can be explained by the

combination of the two individual effects. Transcriptome and metabolome data showed responses that were unique to individual bacterial treatments, and the expression of genes and metabolites in the mixed conditions was consistent with the combination of the effects of the individual strains.

In ongoing work we have begun studying two large communities, each consisting of 10 genome-sequenced bacterial isolates from *P. deltoides* or *P. trichocarpa,* respectively. The 10 strains from each constructed community represent abundant and diverse orders of natural *Populus* microbiomes identified in previous studies, including α-, β- and γ- *Proteobacteria*, *Bacilli,* and *Actinobacteria.* Both constructed communities increased root growth when inoculated on axenic host plants and were dominated by *Burkholderia* and *Pantoea* genera, despite isolation from different hosts and different genomic content. Finally, we observed positive and negative correlations between community members in replicate samples, an example of emergent behavior that would not be observed in one-on-one studies or genera level sequencing studies. These promising results provide a foundation for future constructed community 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.

Plant-Microbe Interfaces: Microbial transformation of *Populus* secondary metabolites by microbiome isolates

Jennifer Morrell-Falvey[1*] (morrelljl1@ornl.gov), **Dale Pelletier**[1], **Greg Hurst**[2], Amber Bible[1], Karuna Chourey[2], Maddie Denney[3], Kitty Kertesz[1], Allen Zinkle[1], Charles Dokytcz[1], Collin Timm[1], Shen Lu[1], Nancy Engle[1], **Timothy Tschaplinski**[1], Mitchel Doktycz[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [3]Genome Science and Technology Graduate program, ORNL/University of Tennessee, Knoxville, TN

http://PMI.ornl.gov

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

Populus is a widely studied model woody plant species and a potential cellulosic feedstock for biofuels. These trees host a wide variety of microbial associations within their roots and rhizosphere and thus serve as a powerful model to study interactions between plants and microorganisms. One of the defining characteristics of *Populus* is the production of secondary metabolites known as higher-order salicylates (HOS), which are involved in host defense and signaling mechanisms. We are interested in determining how the presence and production of these HOS influence the host-microbiome composition and physiology. We have isolated and characterized a number of bacteria from genera (*Pseudomonas*, *Sphigobium*, *Rhizobium*, *Bacillus*, *Pantoea*, *Duganella*, *Burkholderia*, *Caulobacter*, *Streptomyces*, *Bradyrhizobium*, *Rahnella* and *Varivorax)* prevalent in the endosphere and rhizosphere compartments of *Populus* roots and obtained whole genome sequence data. Here we describe physiological characterization of these bacterial isolates with the goals of determining the ability of these strains to transform salicin and other HOS extracted from *Populus* tissues, and measuring the sensitivity and resistance of these bacterial strains to HOS compounds and their degradation products. We present proteomics, metabolomics and comparative genomics data from select strains that have led to identification of potential microbial mechanisms for transformation of the

HOS. We hypothesize that the complex nature of these HOS metabolites may lead to microbial cross feeding and this has been borne out in bacterial co-cultivation experiments. This analysis provides initial insight into the prevalence and diversity of *Populus* microbiota capable of transforming HOS, potential transformation mechanisms, and interactions between microbiome community members.

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

Plant-Microbe Interfaces: Fungal and plant small secreted effector proteins play a key role in the development of the *Populus*-*Laccaria* ectomycorrhizal symbiosis

Jessy Labbé[1]*(labbejj@ornl.gov), Yohann Daguerre[2], Sebastian Wittulsky[2], Annegret Kohler[2], Ritesh Melawal[1], Sara Jawdy[1], Francis Martin[2], Gerald A. Tuskan[1], **Claire Veneault-Fourrey**[2], **Xiaohan Yang**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]INRA, Tree-Microbe Interactions, UMR1136, Champenoux, France

http://PMI.ornl.gov

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

Mycorrhizal symbiosis between perennial plants and fungal associates has critical implications for diverse phenomena including global carbon, water and nutrient cycling. As such, characterizing the molecular genetics underlying such interactions holds tremendous potential in engineering biological systems for enhanced carbon sequestration and sustainable biomass production.

The crosstalk between the two partners is fundamental for the timing, establishment and maintenance of beneficial relationships. Many fungal lineages within the pathogenic/mutualistic continuum have evolved elaborate protein-based signals to manipulate their hosts to foster their needs during symbiosis. These signals, called effectors, which are typically fungal strain- or species-specific, contain a secretion signal motif and are less than 250 amino acids (aa) in size. We have previously described protein-based effectors in the ectomycorrhizal fungus *Laccaria bicolor*. Indeed, we showed that *L. bicolor* relies on Mycorrhizal-induced Small Secreted Proteins (MiSSP) to establish the mutualistic interaction with *Populus*. In particular, MiSSP7 interacts with the jasmonic acid (JA) co-receptor PtJAZ6 of *P. trichocarpa*, blocking JA signaling and promoting mutualism. We showed that PtJAZ6 is able to interact with *Populus* NINJA and TOPLESS proteins as well as the bHLH transcriptional factor, as was found in *Arabidopsis* leaves. Our results emphasize that the JA-mediated signaling pathway is a hub that plant-interacting microbes have to control in order to colonize plant tissues.

In addition, plant-encoded small proteins have also been implicated in affecting symbiotic associations. Recently, deep RNA-seq data and re-annotation of the *Arabidopsis* and *Populus*

genomes showed evidence for new small proteins in *Populus*. We identified 417 plant-based putative small secreted proteins (SSPs) that were significantly regulated during *Populus-Laccaria* interactions. This SSP set was over-represented by proteins predicted to be localized in the nucleus. Of the 417 SSPs regulated during *P. trichocarpa-L. bicolor* mutualism, only 3% were also significantly up- or down-regulated in the pathogenic interaction between *Populus* and *Melampsora laricini-populina*. Furthermore, several *Populus* SSPs were shown to enter the *L. bicolor* nucleus and affect the growth and/or branching of the fungal hyphae.

Our results suggest that both the fungal and plant partners could use small proteins as signals to mediate this mutualistic association. Our ultimate goal aims at identifying the protein-coded signaling circuitry that regulates the early steps of this symbiotic association, in both *L. bicolor* and *Populus*,

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

Plan-Microbe Interfaces: Pleiotropic and Epistatic Network-Based Discovery of Plant Functions Involved in Microbial Interactions: Integrated SNP Correlation, Co-expression and Genome-Wide Association Networks for *Populus trichocarpa*

Deborah Weighill[1,2], Carissa Bleker[1,2], Gerald Tuskan[1], Wellington Muchero[1], Timothy Tschaplinski[1], and **Daniel Jacobson**[1,2] (jacobsonda@ornl.gov).

[1] Oak Ridge National Laboratory, Oak Ridge, Tennessee
[2] University of Tennessee, Knoxville

http://PMI.ornl.gov

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

Biological organisms are complex systems that are composed of pleiotropic functional networks of interacting molecules and macro-molecules. Complex phenotypes are the result of orchestrated, hierarchal, heterogeneous collections of expressed genomic variants regulated by and related to biotic and abiotic signals. However, the effects of these variants are the result of historic selective pressure and current environmental as well as epigenetic interactions, and, as such, their co-occurrence can be seen as genome-wide associations in a number of different manners. In this context, a plant's association with its microbiome is a complex set of interactions involving many genes and metabolites. We are using data derived from the re-sequenced genomes from over 1000 alternate *Populus trichocarpa* genotypes in combination with transcriptomics, metabolomics and phenomics data across this population in order to better understand the molecular interactions involved in plant-microbe interfaces. The resulting Genome-Wide Association Study networks, integrated with SNP correlations and co-expression networks, are proving to be a powerful approach to determine the pleiotropic and epistatic relationships underlying cellular functions and, as such, some of the molecular underpinnings for plant-microbiome associations.

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

Plant-Microbe Interfaces: Probing the Molecular Mechanisms of Plant-Bacterial Interactions

Amy L. Schaefer[1] (amyschae@uw.edu), Yasuhiro Oda[1], Bruna Goncalves Coutinho[1], Justin Weiburg[1], Dale Pelletier[2], Jennifer Morrell-Falvey[2], Vittorio Venturi[3], **E. Peter Greenberg**[1], **Caroline S. Harwood**[1], and Mitchel Doktycz[2]

[1]University of Washington, Seattle WA, [2]Biosciences Division, Oak Ridge National Laboratory, Knoxville TN, [3]ICGEB, Trieste Italy

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

http://pmi.ornl.gov

We have shown that quorum-sensing (QS) genes are prevalent in Proteobacteria isolated from roots of the Eastern cottonwood tree, *Populus deltoides*. Many of these isolates encode an orphan LuxR homolog, which is closely related to OryR from the rice pathogen *Xanthomonas oryzae*. OryR does not respond to acyl-homoserine lactone (AHL) QS signals, instead it detects an unknown plant compound. We discovered an OryR homolog, named PipR, in the endophyte *Pseudomonas* sp. GM79. We examined the genomic region surrounding *pipR* and found genes annotated as a peptide transporter and peptidases. We purified the peptidases and found they are most active against compounds with terminal proline and alanine residues. A reporter responsive to the GM79 PipR homolog was used to show that, similar to *X. oryzae* OryR, its activity increased in the presence of plant leaf macerates, but it was not influenced by AHLs. Because of the abundance of flanking peptide metabolism genes, we hypothesized the PipR signal may be peptide-like in nature. We found that protein hydrolysates (peptone) activated the reporter in a PipR-dependent manner and a specific tripeptide showed moderate inducer activity. Strains with peptidase gene mutations showed increased responses to plant leaf macerate, peptone and the tripeptide signal(s), relative to the wildtype. A mutation in the putative ABC-type peptide transporter locus blocked the response to plant leaf macerate, peptone, and the tripeptide signal(s). We hypothesize that the plant/peptone/tripeptide signal(s) enters the bacterial cells by active transport and that the peptidases affect the signal, likely by enzymatic degradation, in a negative feedback loop. We have also determined that the periplasmic binding protein component of the ABC-type transporter binds the plant and peptone signal(s) tightly and we are using this as a tool to purify and characterize the signal(s). Our analysis of the PipR system in a *Populus*-associated strain opens the door to studies of a specific *Populus*-bacterial interaction that is unexplored. We believe that a better understanding of these PipR-type plant signal

receptors and their plant signals is of general importance as they occur in dozens of bacterial species that are associated with economically important plants.

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

Plant-Microbe Interfaces: Transient expression assays, stable transgenics, and a genome editing system for studying *Populus*-microbe interactions

Jay Chen[1*] (chenj@ornl.gov), Wellington Muchero,[1] Jessy Labbé,[1] David Weston,[1] Xiaohan Yang,[1] Udaya Kalluri,[1] Sara Jawdy,[1] Lee Gunter, [1] and **Gerald A. Tuskan**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

http://PMI.ornl.gov

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

The symbiosis between perennial plants and microbial associates has critical implications for diverse phenomena including global carbon, water and nutrient cycling as well as biomass production on marginal croplands. As such, characterizing the molecular genetics underlying such interactions holds tremendous potential in engineering biological systems for enhanced carbon sequestration and sustainable biomass production.

Over the last six years of PMI SFA research, a number of genetic loci regulating the interactions between *Populus* and soil microbes have been identified, mainly through genome-wide association studies (GWAS) and quantitative trait locus (QTL) mapping. While these natural variants and QTL pedigrees continue to serve as valuable resources for evaluating host genotype influences on microbial community composition, diversity and function, genetic materials with a more defined background (i.e., only with alternation in expression of a single gene) can provide key insights to pinpoint the function of these genetic regulators at the molecular and biochemical levels. Previously, we have established a *Populus* mesophyll protoplast transient expression system which has been used for the studies of protein subcellular localization, protein-protein interaction, protein degradation and gene activation and repression. Recently, this system has also been successfully applied to chromatin immunoprecipitation (ChIP) and ChIP-sequencing (ChIP-seq).

We are also developing a number of additional systems including the generation of transgenic lines in *Populus deltoides* and *Populus trichocarpa* backgrounds. Transformation efficiency is determined by both genetic and non-genetic factors. Both *Populus* species and within-species genotypes vary greatly at the gene, allelic and nucleotide levels. Because the genotype-specific variation is large, there is no universal transformation protocol suitable for all *Populus* genotypes

at this time. The 1084 *P. trichocarpa* GWAS population provides a rich source of raw materials for identifying superior genotypes for efficient transformation based on callus induction efficiency. From this population, we have identified a number of genotypes that are transformable. Non-genetic factors including tissue type, callus-induction medium composition, shoot induction medium composition, *Agrobacterium* strain, culture conditions and vector type will be further optimized for improving transformation efficiency. In addition to traditional tissue culture and an *agrobacterium*-based transformation method, we will apply the CRISPR/Cas9 genome-editing system to specifically disrupt the function of selected target genes in *Populus*. Our initial test will be on the disruption of a gene encoding a lectin receptor-like kinase that has been shown to play a key role in *Populus-Laccaria bicolor* interactions.

Because the time required to generate stably transformed *Populus* plants and to propagate plant materials is lengthy (~9 to 15 months), we are also establishing a hairy root transformation system using *Agrobacterium rhizogenes* in *Populus*. This system can be used to rapidly assess gene expression and function, and can also be used as a screening method to select genes for the generation of stable *Populus* transgenic lines.

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.

Title: Intron-retained splice variants of the VND6 and SND1 transcription factors are dominant negatives that cross-regulate VND6 and SND1 members in *Populus trichocarpa*.

Ying-Chung Lin[1,*] (ylin14@ncsu.edu), Hao Chen,[1] Quanzi Li,[1,2] Rui Shi,[1] Sermsawat Tunlaya-Anukit,[1] Peng Shuai,[1,3] Wei Li,[1,4] Huiyu Li,[1,4] Ying-Hsuan Sun,[5] **Ronald R. Sederoff**,[1] and **Vincent L. Chiang**[1]

[1] Forest Biotechnology Group, Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, North Carolina 27695

[2] State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing 100091, China

[3] College of Forestry, Fujian Agriculture and Forestry University, Fuzhou 350002, China

[4] State Key Laboratory of Tree Genetics and Breeding, Northeast Forestry University, Harbin, China

[5] Department of Forestry, National Chung Hsing University, Taichung 40227, Taiwan

Vascular-Related NAC-Domain 6 (VND6) is a key transcription factor (TF) involved in xylem and secondary cell wall differentiation. We discovered a splice variant of PtrVND6, called PtrVND6-C1^{IR}, which is a dominant negative regulator of full-size PtrVND6 members. PtrVND6-C1^{IR} lacks a transactivation domain and DNA binding ability, and can be translocated from the cytosol into the nucleus as a heterodimeric partner with any full-size PtrVND6 member. The formation of heterodimers between PtrVND6-C1^{IR} and the full-size PtrVND6 disrupts the function of the full-size PtrVND6, thereby repressing transcription of PtrVND6 direct targets in its network. Secondary Wall-Associated NAC Domain 1 (SND1) also affects secondary cell wall biosynthesis. We previously demonstrated that the splice variant of PtrSND1-A2, PtrSND1-A2^{IR}, can inhibit the PtrSND1 transcription network through the same mechanism. Using laser capture microdissection, we found that PtrVND-6C1^{IR}, PtrSND1-A2^{IR}, and all full-size PtrVND6 and PtrSND1 are expressed in both fiber and vessel cells. We further discovered that either PtrVND6-C1^{IR} or PtrSND1-A2^{IR} can inhibit both PtrVND6 and PtrSND1 transcription by the same mechanism. The cross-regulation between the PtrSND1 and PtrVND6 families through their splice variants suggests a general mechanism for the function of xylem specific NAC TFs controlling wood formation.

Inducible Extreme Expression of Cellulases in Poplar

Charleson Poovaiah, Yao Xiao and **Heather Coleman*** (hcoleman@syr.edu)

Syracuse University, Syracuse, New York

Project Goals: The overall goals of the project is to verify in poplar the In Plant Activation (INPACT) technology, which enables inducible expression of genes and accumulation of proteins at very high levels *in planta* and to evaluate the ability of cellulases produced through this technology to hydrolyze cellulose to simple sugars for fermentation.

Cost of cellulolytic enzymes is a significant constraint in biofuel production from lignocellulosic biomass. *In planta* production of these enzymes could potentially reduce the cost associated with bioethanol production. This project aims to produce high levels of cellulases in poplar upon induction with an exogenous chemical. Production of enzymes *in planta* would decrease the amount of additional enzymes necessary for hydrolysis of cellulose. In Plant Activation (INPACT) technology allows for very high inducible expression of recombinant proteins in planta. INPACT uses the rolling circle replication of Gemini virus to produce high levels of gene amplification and protein production. In this project initially, we will verify the adaptability of this technology in poplar for its ability to accumulate recombinant proteins at very high levels. We will then use INPACT to express cellulases from three major groups of enzymes, endoglucanases, exoglucanases and β-glucosidases in poplar with constitutive and tissue specific promoters. Cellulases from thermophilic organisms have been plant codon optimized and synthesized. These cellulases have been assessed for correct splicing in tobacco and protein production is currently being assessed in yeast. The constructs harboring the alcohol inducible replication initiation protein (Rep) which allows for the induction of the INPACT system has been successfully transformed into poplar and a mother line selected based on Rep/RepA gene expression and plant growth before and after alcohol treatment. The transgenic poplar plants with the alcohol inducible Rep and the GUS gene in the split orientation are being multiplied to produce clonal lines and will then be evaluated for expression in leaf and developing xylem using the GUS reporter system. The transgenic mother plants with thermostable cellulases in the split orientation are being regenerated, as are the positive and negative control lines.

Funding statement: This project is funded by the U.S. Department of Energy Office of Biological and Environmental Research (DE-SC0010411).

Natural Variation of Abiotic Stress Tolerance for Biomass Production in a C4 Model Grass

Malia A. Gehan[1], Todd C. Mockler[1*](tmockler@danforthcenter.org), Tracy Ficor[1], Stephanie Turnipseed[1], Erica Agnew[1], Andy Lin[1], Noah Fahlgren[1], Pu Huang[1], Maximilian Feldman[1], Ivan Baxter[1,2] (ibaxter@danforthcenter.org), and **Thomas P. Brutnell[1]**

[1] Donald Danforth Plant Science Center, St Louis, Missouri and [2] USDA-ARS, St Louis, Missouri

http://plantcv.danforthcenter.org/

Project Goals: Genetically tractable model systems closely related to bioenergy grasses need to be developed to drive the crop improvement required for large scale, ecologically sustainable bioenergy production. Setaria viridis is an ideal candidate C4 panacoid grass. The objectives of this project are to utilize genomic, computational and engineering tools to begin the genetic dissection of drought and density response in S. viridis. This will be achieved through: 1) Quantitative trait and Association genetics; 2) novel controlled environment and field phenotyping combined with molecular and chemical profiling; 3) development of metabolic and gene networks; 4) development of transformation technologies; 5) reverse genetic testing of candidate genes.

To tackle the daunting challenge of producing more fuel with fewer inputs, a variety of strategies to improve and sustain crop yields are necessary. Plans for crop improvement may include mining natural variation of wild crop relatives to breed crops that require less water or increasing crop tolerance to temperature extremes to expand the geographical range in which they grow. We are examining natural variation in the low-temperature tolerance of *Setaria viridis* accessions collected throughout North America using low-cost phenotyping technologies. Improving the low-temperature tolerance of a plant could not only increase its geographical growth range,but accessions that germinate earlier in the growing season could avoid late season drought and heat stress. Further, with overlapping stress response mechanisms, improving the low-temperature tolerance of a plant may also increase drought tolerance. To capture emergence and growth phenotypes under low-temperature we are using low-cost Raspberry Pi microcomputers and cameras. To quantify growth traits, we use Plant Computer Vision (PlantCV), which we developed to be an open-source, platform independent quantitative image analysis community resource.

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

Use of KitBase to Facilitate Forward and Reverse Genetics Research in Rice

Guotian Li[1, 2*](guotianli@lbl.gov), Mawsheng Chern[1, 2], Rashmi Jain[1, 2], Joel A. Martin[3], Wendy S. Schackwitz[3], Liangrong Jiang[1], Miguel E. Vega-Sánchez[2], Anna M. Lipzen[3], Kerrie W. Barry[3], Jeremy Schmutz[3, 4], and **Pamela C. Ronald**[1, 2]

[1]Department of Plant Pathology and the Genome Center, University of California, Davis, CA 95616, USA; [2]Feedstocks Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA; [3]U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA, [4]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA.

http://www.jbei.org

Project Goals: We are using whole-genome sequencing approach to establish a fully sequenced mutant collection in a model rice variety that is being used to identify genes involved in grass cell wall biosynthesis and modification.

To achieve cost efficient conversion of lignocellulosic biomass into biofuels, basic knowledge on the genes that control cell wall biosynthesis and modification is needed in grasses. To facilitate cell wall research, we have generated a mutant population using fast-neutron irradiation in the model rice cultivar Kitaake, a tractable model grass species. Kitaake is an early flowering, short-statured, short life cycle rice that is easy to transform, compared with other Japonica and Indica rice varieties. In collaboration with the Joint Genome Institute, we are sequencing 4,000 mutants. Genomic analysis of more than 1,000 mutants has been done, revealing that 13,469 genes are affected. Mutation types include single base substitutions, deletions, insertions, inversions, translocations, tandem duplications, and complex events. Single base substitutions predominate, but deletions affect the greatest number of genes, accounting for 57.2% of all affected genes. To make the genetic resource publicly available, we established KitBase, a comprehensive repository for rice mutant information. KitBase integrates JBrowse and BLAST to facilitate identification of mutations and searching of the database. KitBase includes genomic data, phenotypic data, and seed information for each of the mutant lines. Dozens of glycosyltransferase (GT) and glycohydrolase genes have been mutated in this mutant collection, and the function of multiple GTs in rice cell wall biosynthesis is being studied.

This work is supported by the Office of Science of the U.S. Department of Energy Contract No. DE-AC02-05CH11231 to the Joint Genome Institute and a National Science Foundation grant (IOS-1237975) to PCR.

Dissecting the role(s) of host genotype and phytobiome composition and function in the successful establishment of switchgrass (*Panicum virgatum* L) on marginal soils.

KD Craven* (kdcraven@noble.org), *[1], MC Saha[1], WR Scheible[1], M Udvardi[1], and **MK Firestone[2]**

[1]The Samuel Roberts Noble Foundation, Ardmore OK, 73401; [2]University of California, Berkeley, CA, 94720; * Presenting author

Project Goals: (1) Identify high- and low-performing switchgrass (SG) genotypes in marginal soils and determine the functional succession of SG-associated microbial communities during successful SG establishment of each. (2) Characterize plant–microbe and microbe–microbe interactions within and between SG and its microbiome, particularly when challenged by water or nutrient stress. (3) Determine how low-input SG production in marginal soils may enhance ecosystem sustainability metrics such as: C storage, nutrient availability, and soil food webs. (4) Integrate and synthesize experimental data to reveal plant-microbe interactions and the underlying mechanisms critical to SG effects on ecosystem sustainability.

To jump start research aimed at achieving the ambitious goals listed above, *PI Saha* has identified high and low biomass genotypes from fifteen families of a switchgrass nested association mapping (NAM) population (see figure 1). The NAM population was derived by crossing fifteen diverse parents, selected from a diversity panel pre-screened for certain useful agronomic traits like early vigor and yield, to a recurrent parent AP13 (draft genome, microarray and other genomic resources available). Subsequently, 10 F_1s from each of 15 families were chain crossed to develop a NAM population of 2,000 genotypes (see figure 2). The population has been evaluated in field experiments at two locations for two years. Soil quality for these trials had adequate levels of nitrogen, phosphorus and organic matter. Based on overall performance across environments, 80 high yielding and 20 low yielding genotypes were identified. Clonal ramets are being generated for 4-6 high-yielding genotypes in the greenhouse for initial testing on soils poor in nitrogen, phosphorus, or both. Based on these results, a single high-yielding genotype (on these marginal soils) will be chosen for subsequent analyses.

To dissect the role of the microbiome in facilitating switchgrass growth on marginal soils, seedlings will be planted at 2-3 Oklahoma sites with soils characterized as nutritionally depleted of organic matter. *PI Craven* will collect bulk soil for baseline microbial metagenomic analysis, and subsequently follow the succession of switchgrass rhizosphere microbial and microfaunal communities associated with establishment-phase plants that are high and low performers. Rhizospheric soil and root tissues will also be collected from each for the isolation of bacterial and fungal endophytes. We have established high-throughput screens for many potential useful microbial traits, including N-fixation, ACC deaminase activity, and solubilization of inorganic phosphorus (CaPO4). Bacterial or fungal strains found to contain one or more will be candidates for SIP tracer studies and simplified community modeling and analysis.

To elucidate differences in nutrient use efficiency (NUE), flux and utilization, *PIs Scheible* and *Udvardi* will carry out physiological, transcriptomic, metabolomics and fluxomic studies on plants grown under optimal or limiting nutrient conditions. Switchgrass growth experiments using plants grown from Alamo seed have been initiated using sand to define nutritional

parameters for subsequent studies. Transcriptomic and metabolomic studies will reveal genes and processes that are involved in acclimation to nutrient limitation, and which are conserved or not between switchgrass genotypes or plant species more broadly. This will help to clarify how comparable switchgrass is to other species, and will also help to identify molecular and genetic markers for subsequent analyses. A qPCR platform will be developed to determine expression of SG homologs to known P and N-signaling, -metabolism and -status indicator genes, including microRNAs. Transcriptomic and metabolomic studies of plants devoid of microbes will serve as a base-line for comparison of plants associated with microbes, with the objective of identifying nutritional and other services provided by microbes to plants, and vice versa.

Funding
This research is based upon work supported by the U.S. Department of Energy Office of Science number DE-SC0014079 to UC Berkeley, Samuel Noble Foundation, University of Oklahoma, Lawrence Berkeley National Lab, and Lawrence Berkeley National Lab.

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z