DOE/SC-0142

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

U.S. DEPARTMENT OF
ENERGY
Office of Science

2012

Genomic Science Awardee Meeting X

Bethesda, Maryland
February 26-29, 2012

[Revised: March 1, 2012]

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

http://genomicscience.energy.gov

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

Bioenergy Research Centers

Joint BioEnergy Institute (JBEI)

1
The Challenge of Enzyme Cost in the Production of Lignocellulosic Biofuels

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

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

http://www.jbei.org

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

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

2
Structural Comparison of Plant Glycosyltransferases

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

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

http://www.jbei.org

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

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

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

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

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

3

A Genome-Wide Survey of Switchgrass Genome Structure and Organization

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

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

http://www.jbei.org

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

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

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

4

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

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

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

http://www.jbei.org

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

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

5

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

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

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

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

http://www.jbei.org

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

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

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

6

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

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

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

http://www.jbei.org

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

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

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

7

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

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

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

http://www.jbei.org

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

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

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

8

Feedstock Agnostic Pretreatment Technology

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

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

http://www.jbei.org

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

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

9

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

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

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

http://www.jbei.org

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

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

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

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

10

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

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

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

http://www.jbei.org

Project Goals: Physical characterization of plant stems and leaves.

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

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

11

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

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

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.; [2]Sandia National Laboratories, Livermore, Calif. and Albuquerque, N.M.; [3]Department of Chemistry, Virginia Polytechnic Institute, Blacksburg; [4]Lujan Neutron Science Center, Los Alamos National Laboratory, Los Alamos, N.M.; [5]Spallation Neutron Source, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [6]Department of Cell Biology and Molecular Genetics, University of Maryland, College Park; and [7]Center for Integrated Nanotechnologies, Albuquerque, N.M.

http://www.jbei.org

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

Intensive efforts are underway to lower the cost of enzymatic hydrolysis of cellulose to sugars. There are three general types of cellulases: endoglucanses hydrolyze internal bonds and produce chain ends, exoglucanases hydrolyze from the chain ends and release cellobiose, and b-glucosidases convert cellobiose units to glucose. These generic activities are likely to depend upon substrate characteristics. While each component plays an individual role, they work synergistically for highly efficient cellulose degradation. The full nature of that synergism, and its dependence on substrate characteristics, is not fully understood. A fundamental understanding of enzyme synergy would greatly aid the design of enzyme cocktails. In this work we are using

quartz crystal microbalance (QCM) and neutron reflectivity (NR) measurements to reveal the actions of endoglucanases on amorphous cellulose. Amorphous cellulose is relevant to biomass pretreated with ionic liquids. QCD and NR are highly complementary and provide unprecedented detail into the effects of endoglucanses on film structure. Results are shown below for an enzyme cocktail from *T. viride* and from two endos that show qualitatively different behavior.

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

12

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

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

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

http://www.jbei.org

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

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

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

13

Tuning Cellulase Activity Using Carbohydrate Binding Modules

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

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

http://www.jbei.org

Project Goals: (see below)

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

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

14

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

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

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

http://www.jbei.org

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

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

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

15

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

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

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

http://www.jbei.org

Project Goals: The Microbial Communities group at JBEI aims to develop a fundamental understanding of how microbial communities degrade targeted biomass feedstocks, and to utilize a targeted, function-based screening approach to genomics and proteomics to identify, isolate, and characterize new enzymes that are capable of efficiently degrading lignocellulosic feedstocks. Focusing on a thermophilic switchgrass-adapted enrichment community yields an order of magnitude more useful enzyme sequences compared to our previous work on a more complex community, and the resulting enzymes are more likely to be well suited to our targeted feedstock, pretreatment, and processing conditions. Combining enzymatic assays, metagenomics, zymography, MS proteomics, and metabolic modeling provides a multidimensional view of the internal functioning of this highly active biomass degrading bacterial community.

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

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

16

Synthetic Biology Design Automation Tool Suite

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

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

http://www.jbei.org

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

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

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

17

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

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

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

http://www.jbei.org

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

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

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

18

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

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

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

http://www.jbei.org

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

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

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

19

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

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

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

http://www.jbei.org

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

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

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

20

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

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

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

http://www.jbei.org

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

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

21

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

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

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

http://www.jbei.org

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

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

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

22

Engineering of Bacterial Methyl Ketone Synthesis for Biofuels

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

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

http://www.jbei.org

Project Goals: The Joint BioEnergy Institute (JBEI) aims to produce a chemically diverse suite of biofuels from lignocellulosic biomass. Many of the target fuels at JBEI rely on well-characterized metabolic pathways (such as the straight-chain fatty acid biosynthetic pathway) to provide precursors for synthesis of biofuel molecules. The goal of this project was to produce methyl ketones from fatty acids for use as blending agents or substitutes for diesel fuel.

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

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

23

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

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

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

http://www.jbei.org

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

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

24

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

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

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

http://www.jbei.org

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

Systems biology aims to provide a predictive and quantitative understanding of cell behaviour as the outcome of the interaction of its comprising parts. Metabolic flux profiles (i.e. the number of molecules traversing each biochemical reaction encoded in its genome per unit time) are not only a key phenotypic characteristic but also embody the essence of this complexity since they represent the final functional output of the interactions of all the molecular machinery studied by all the other "omics" fields. Two of the most

popular methods for studying metabolic fluxes are Flux Balance Analysis (FBA) and ^{13}C Metabolic Flux Analysis (^{13}C MFA), each of them displaying its own advantages and disadvantages. Here, we present a new method: Two scale ^{13}C Metabolic Flux Analysis (2S-^{13}CMFA), which combines the advantages of FBA and ^{13}C MFA. We showcase its applications and possibilities with data from the KEIO knockout collection.

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

25

Mutually Consistent Metabolic Flux and Metabolite Concentration Prediction

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

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

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

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

Great Lakes Bioenergy Research Center (GLBRC)

26

Development of Crucial Tools for Lignin Research

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

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

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

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

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

* Presenting author

27

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

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

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

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

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

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

28

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

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

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

http://glbrc.org

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

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

29

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

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

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

http://glbrc.msu.edu/

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

Cellulose, the world's most abundant biopolymer, is a major feedstock for bioenergy. Therefore, genetic manipulation of cellulose synthesis in specific cells/tissues/organs of economically important crops is one of the top-priorities in current plant biotechnology research. Yet, biotechnological manipulation of cellulose biosynthesis has been one of the most challenging tasks, mainly because of the fact that cellulose biosynthesis is carried out by multi-member cellulose synthase complexes at the plasma membrane. In the secondary walls of *Arabidopsis* plants, three CESAs (CESA4, CESA7 and CESA8) are involved in cellulose biosynthesis. The most prudent approach to increase cellulose biosynthesis will be to simultaneously upregulate the genes encoding for all of the necessary cellulose synthases in the presence

of increased substrate level. However, little is known about the transcription regulation of these CESA genes. In fact, no transcription factors that bind to the promoter of CESA gene has been reported yet. We have recently identified a direct regulator of CESAs, named Secondary wall Cellulose synthase Regulator 1 (SCR1), that can simultaneously upregulate the expression of all of the three secondary wall CESAs (U.S. Patent, in filing). Several lines of evidence support our hypothesis that SCR1 may function as one of the direct transcriptional regulators of secondary wall cellulose synthase genes CESA4, CESA7, and CESA8. Building upon this preliminary success, we are developing a novel approach, termed 'targeted regulation of cellulose biosynthesis,' which combines our current understanding of cellulose biosynthesis with the state-of-the-art biotechnology to materialize the concept of '*in planta*' or better yet '*in crop*' manipulation of cellulose biosynthesis.

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

30

Metabolic and Transcriptional Changes During Induced Senescence in Maize

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

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

http://www.glbrc.org/

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

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

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

31

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

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

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

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

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

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

32

Identification of Novel Biomass Deconstructing Enzymes via Phylogenetic and Structural Analyses

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

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

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

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

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

33

Strategies to Enable the Assembly of Soil Metagenomes

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

Michigan State University, East Lansing

http://www.glbrc.org

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

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

Our initial step is to reduce the size of the metagenomic dataset by normalizing the average coverage of the soil metagenome using an approach we term "digital normalization." We eliminate redundant, high-coverage short-reads

within a dataset using a single-pass, constant memory algorithm. This normalization reduces the metagenome dataset size so that it has an increasingly even distribution of read-coverage. Comparing assemblies before and after digital normalization for an *E. coli* genome (50x coverage normalized to 5x), we found that assemblies were >99% similar despite eliminating 90% of the reads. Similar results were observed for a subset of a soil metagenome.

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

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

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

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

34

Developing a Toolbox for Soil Metagenomes

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

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

http://www.glbrc.org

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

Metagenomic de novo assembly with digital normalization and Khmer

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

HMMgs: gene-targeted assembly from metagenomic datasets

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

* Presenting author

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

Phylogenetic profile comparisons between pyrosequencing and shotgun sequencing

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

Analysis of genomes detected by mapping raw reads to known genomes

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

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

35

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

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

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

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

The expansion of land under agricultural management has significantly increased atmospheric forcing from greenhouse gases—carbon dioxide, methane and nitrous oxide. Agricultural soils now contribute approximately half the anthropogenically derived N_2O in the atmosphere[1], and soils that are typically a sink for methane consume little to no methane when converted[2]. Microbial communities are almost entirely responsible for catalyzing nitrous oxide flux, reducing nitrate to nitrous oxide and nitrogen in a process known as denitrification[3]. Although present at low abundances, methanotrophs in soil oxidize methane, with higher diversity communities consuming up to 8 g CH_4-C per hectare per day[4]. In an effort to assess changes in the taxonomic composition and metabolic potential of microbial communities, metagenomic and targeted genomic

surveys across a gradient of land uses at the KBS LTER and Michigan GLBRC sites were conducted. We have found that microbial communities take approximately 20 years to recover, both taxonomically and functionally, from agricultural management, concomitant with soil biogeochemistry, and that denitrifier communities are particularly affected by agricultural practices. Under agriculture, the denitrifier community increases from approximately 10% of the community to 33%. Additionally the community composition of denitrifiers changes to an increased proportion of ammonia oxidizing bacteria that lack the capacity to reduce N_2O to N_2. It is likely that fertilization is having a primary affect on this shift[5] and further studies on a switchgrass fertilization gradient are being conducted. Our data suggests that both the abundance of denitrifiers in a community, and their structure, determine the rate of nitrous oxide production in soils and that an understanding of denitrifier communities could lead to solutions for mediation or more accurate models of terrestrial nitrous oxide fluxes.

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

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

References

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

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

36

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

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

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

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

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

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

For biogeochemical cycling, the soil microbial community presents a promising target as both a major driver of observed agroecosystem processes and a sensitive integrator of climatic, edaphic, and agronomic effects. However, the size and complexity of the soil microbial community pose a serious technical challenge to meaningfully operationalizing composition metrics. Shotgun metagenomic sequencing has greatly advanced our understanding of the genetic composition microbial communities, particularly since the advent

of Next-Gen sequencing and improvements in the computational approaches to handle such data. Despite these advances, metagenomic sequencing remains a nontrivial technical challenge, making it ill-suited for widespread diagnostic use or studies with high biological replication. Narrower approaches to community characterization, such as targeted sequencing and microarrays, provide more tractable datasets, but are limited in their scope to a predefined universe of genes and sequences. In this project, we are combining the open-ended exploratory benefits of shotgun metagenomic sequencing with the throughput and turnaround time of a microarray.

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

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

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

37

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

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

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

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

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

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

38

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

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

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

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

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

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

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

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

39

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

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

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

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

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

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

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

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

Reference

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

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

40

Improved Enzyme Mixtures for Biomass Deconstruction

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

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

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

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

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

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

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

References

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

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

41

Three-Dimensional Structures of Enzymes Involved in Cellulose Degradation

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

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

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

42

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

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

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

http://www.glbrc.org/

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

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

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

43

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

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

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

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

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

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

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

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

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

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

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

References

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

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

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

44

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

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

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

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

Project Goals:

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

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

Publications

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

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

45

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

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

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

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

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

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

Publications

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

46

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

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

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

The GLRBC plant oil project aims at increasing the content of high energy compounds in vegetative tissues of plants. The concept of enhancement of energy density in plant biomass is synergistic with other GLBRC efforts to develop lignocellulosic feed stocks for biofuel. Enhancing the energy yield of plant biomass can be achieved by accumulating energy-dense compounds such as triacylglycerol (TAGs).

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

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

47

Improving the Sustainability of Bioenergy Crops Through Arbuscular Mycorrhization

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

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

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

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

Availability of water and nutrients is a major constraint for crop productivity and sustainable agriculture. Over the last decades, there has been an excessive dependence on chemical fertilizers with major economic, ecological and heath consequences. Taking better advantage of plant-microbe symbioses like arbuscular mycorrhization (AM) seems a reasonable alternative to improve crop yields and the sustainability of our agricultural systems. All the major bioenergy crops can form associations with AM fungi that improve their acquisition of water and nutrients (especially phosphorus). AM is the most efficient symbiosis between soil microbes and bioenergy crops. On a global level, AM symbiosis contributes significantly to phosphate, nitrogen and carbon cycling. Therefore, improving the efficiency and the development of AM associations especially under sub-optimal conditions has a tremendous potential for improving the sustainability of biofuel production. Our goal is to characterize the signals and the genes controlling the establishment of AM in energy crops. We developed a high-throughput screening of maize mutants affected in AM symbiosis. We screened more than 4000 lines of mutagenized population of B73 and identified six mutants which are unable to establish AM. The absence of AM in these mutants was confirmed by microscopy. These mutants have been self-fertilized to produce M3 progenies for further phenotypic characterization and crossed to a polymorphic parent for positional cloning. The screening of additional maize lines is ongoing. In order to characterize the signals produced by AM fungi, we developed an easy procedure to collect such diffusible signals in germinating spore exudates (GSE); these stimulate plant growth in monocots (maize, rice) and eudicots (alfalfa). They also induce expression of AM-specific genes through several genetic pathways and this induction is negatively regulated by the stress hormone, ethylene. In order to analyze the GSE-induced early responses in maize at the transcript level and their regulation by ethylene, microarray experiments were performed. Data analysis indicates that these signals induce gene expression in maize and ethylene inhibits this gene expression.

48

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

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

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

http://www.glbrc.org

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

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

References

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

49

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

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

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

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

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

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

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

50

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

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

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

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

Rhodobacter sphaeroides is a metabolically diverse purple nonsulfur photosynthetic bacterium. During anaerobic photoheterotrophic growth (with light as an energy source and an organic substrate as a carbon and electron source),

R. sphaeroides can produce hydrogen gas (H_2) via its nitrogenase enzyme. The wild-type strain 2.4.1 is capable of channeling up to 40% of the electrons it consumes from organic substrates into H_2, depending on the organic substrate.

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

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

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

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

BioEnergy Science Center (BESC)

51

BioEnergy Science Center Education and Outreach—Overview

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

[1]The Creative Discovery Museum, Chattanooga, Tenn.; [2]University of Georgia, Athens; [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [4]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). Education and outreach to the general public is critical in the acceptance and deployment of bioenergy. In addition to leveraging successful education and training programs already in place at our partner institutions, BESC has developed educational lessons and activities that target elementary and middle school children.

Our Center has taken a novel approach to education and outreach in that our education efforts begin with 5th graders. This is in addition to our efforts to prepare a new generation of scientists for the emerging fields of bioenergy through the interdisciplinary training of graduate students and postdocs. We have developed lesson plans aimed at 4th, 5th and 6th grades to educate and inform students about the basics of energy production and utilization. They include basic concepts such as the carbon cycle, lignocellulosic biomass as substrate for the production of biofuels as well as technical and economic obstacles to a bio-based fuel economy. These lessons were piloted in schools in Georgia and Tennessee and were made available to schools nationwide in the spring of 2010. We have piloted a series of "science night" programs offered to students and the general public through local schools, museums and community centers and have reached more than 35,000 students, teachers and parents. In addition, we have developed educational programs in the form of games that teach strategies for energy use. Students are allowed to design their own cars and select types of fuel to travel to familiar destinations. The games, exhibited in kiosks to be placed in schools, museums and other educational venues, incorporate lessons to explore fuel efficiency, fuel availability (for example electric and E85 cars) as well as environmental impact. A prototype kiosk has been built and was exhibited at a recent ASTC meeting attracting interest

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

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

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

52

BESC Knowledgebase, Tools, and Laboratory Information Management System

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

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

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

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

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

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

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

The microbial core KB for microbes consists of 37 microbes, including biomass degraders, fuel producers, endophytes and model organisms. Data collected on these organisms include genome annotation, biochemical data including enzymes, ligands and pathway annotation generated using the Pathway Tools software and provided in KEGG, carbohydrate active enzyme data generated by CAT and provided by CAZy, operon predictions from BeoCyc and DOOR, protein functional domain predictions from CDD, Pfam, COGS, TigrFam, SMART. Pair-wise predictions of orthologs and inparalogs for the genomes generated using bi-directional BLAST hits and Inparanoid software, BLAST hits against NCBI NR database. Trans membrane protein predictions using TMHMM and signaling proteins predictions using SignalP. We also have omics data on effect of alcohol on cell, fermentation time course study, growth on model substrates found in lignocelluloses, growth on simple and complex sugars, biomass deconstruction, and many other experiments from several *Clostridium* and other bioenergy relevant microbes. Recent addition to microbial core includes a microbial resequencing database (ReSeqDB); we have made available resequencing data for an ethanol-adapted *Clostridium thermocellum* strain. We provide not only

high confidence differences, but also processed results from several different tools (Genome Resequencing Toolkit) and comprehensive annotations of those sequence changes, position of sequence change with respect to gene and intergenic region, changed protein sequence, and probable change in gene function or regulation. SNPs indels are also visualized and made available from the genome browser. The goal of the resequencing projects is to discover genomic modifications underlying the specific phenotype of the mutant. The Genome Resequencing Toolkit reveals mutational changes in the mutant strains to understand the biological effects of each change on the mutant phenotype.

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

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

53

Computational Biology to Target Plant Pathways and Experimental Data

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

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

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). Addressing the roadblock of biomass recalcitrance will require a greater understanding of plant cell walls from synthesis to deconstruction. This understanding would generate models, theories and finally processes that will be used to understand and overcome biomass recalcitrance. This grand challenge calls for an integrated research approach, which is illustrated by the project described below. BESC research is multidisciplinary by design and multi-institutional in composition.

The computational biology team of BESC is working with experimental biologists on a number of collaborative projects to study plant cell wall recalcitrance. Four examples are highlighted.

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

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

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

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

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

Reference

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

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

54

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

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

[1]National Renewable Energy Laboratory, Golden, Colo.; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]University of Tennessee, Knoxville; and [4]University of Georgia, Athens

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC biomass formation and modification research involves working directly with two potential bioenergy crops (switchgrass and *Populus*) to develop varieties that are easier to break down into fermentable sugars. We are using both testing and generation of large numbers of natural and modified plant samples as well as developing genomic tools for detailed studies into poorly understood cell wall biosynthesis pathways.

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

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

55

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

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

[1]The University of Georgia, Athens and [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC biomass formation and modification research involves working directly with two potential bioenergy crops (switchgrass and *Populus*) to develop varieties that are easier to break down into fermentable sugars. We are using both testing and generation of large number of natural and modified plant samples as well as developing genomics tools for detailed studies into poorly understood cell wall biosynthesis pathways. BESC researchers in characterization, modeling, and data management areas are engaged in 1) applying advanced technologies to analyze chemical and structural changes within biomass and 2) storing, tracking, analyzing and integrating data and understanding across the Center.

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

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

56

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

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

[1]School of Chemistry and Biochemistry, Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta; [2]BioEnergy Science Center, National Renewable Energy Laboratory, Golden, Colo.; [3]Department of Chemical and Environmental Engineering, Center for Environmental Research and Technology, University of California, Riverside; and [4]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC researchers in characterization, modeling, and data management areas are engaged in 1) applying advanced technologies to analyze chemical and structural changes within biomass and 2) storing, tracking, analyzing and integrating data and understanding across the Center.

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

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

57

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

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

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

http://www.bioenergycenter.org

Project Goals: BESC researchers in characterization, modeling, and data management areas are engaged in 1) applying advanced technologies to analyze chemical and structural changes within biomass and 2) storing, tracking, analyzing and integrating data and understanding across the Center. BESC characterizes samples in order to identify low recalcitrance lines of switchgrass and *Populus*. This collaboration with plant biologists provides understanding how and why plants low recalcitrance lines are easier to process.

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

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

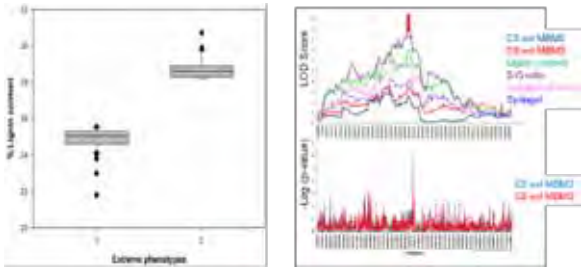

Figure 1: Data from a study of 800 naturally occurring *Populus* from the Pacific Northwest. Left extreme lignin phenotypes, right genetic mapping finds results from both the cell wall chemistry and the sugar release assay coincide.

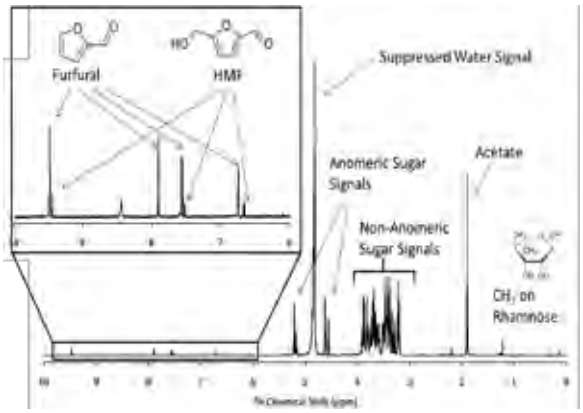

Figure 2: Example NMR spectrum from HTP sugar analysis.

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

References

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

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

58
High Resolution Characterization of Biomass

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

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

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC researchers in characterization, modeling, and data management areas are engaged in 1) applying advanced technologies to analyze chemical and structural changes within biomass and 2) storing, tracking, analyzing and integrating data and understanding across the Center.

BESC has developed a method to provide chemical information about biomass at the scale of tens of nanometers. When employing absorption spectroscopy as a means for obtaining chemical information from samples in bulk form, one is faced with an inherent limitation on the spatial resolution. This lack of resolution, which is caused by the clas-

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

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

59

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

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

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

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC research in biomass deconstruction and conversion targets CBP by studying model organisms and thermophilic anaerobes to understand novel strategies and enzyme complexes for biomass deconstruction.

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

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

60

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

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

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

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC researchers in characterization, modeling, and data management areas are engaged in 1) applying advanced technologies to analyze chemical and structural changes within biomass and 2) storing, tracking, analyzing and integrating data and understanding across the Center.

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

References (recent publications supported by this project)

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

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

61

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

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

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

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC biomass formation and modification research involves working directly with two potential bioenergy crops (switchgrass and *Populus*) to develop varieties that are easier to break down into fermentable sugars. We are using both testing and generation of large numbers of natural and modified plant samples as well as developing genomic tools for detailed studies into poorly understood cell wall biosynthesis pathways.

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

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

grant DE-PS02-06ER64304.
The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

62

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

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

[1]West Virginia University, Morgantown, W.Va.; [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [3]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, Calif.

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC biomass formation and modification research involves working directly with two potential bioenergy crops (switchgrass and *Populus*) to develop varieties that are easier to break down into fermentable sugars. We are using both testing and generation of large number of natural and modified plant samples as well as developing genomic tools for detailed studies into poorly understood cell wall biosynthesis pathways.

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

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

63

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

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

[1]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge Tenn.; [2]The University of Georgia, Athens; [3]Great Lakes Bioenergy Research Center, East Lansing, Mich.; and [4]University of California, Riverside

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). Addressing the roadblock of biomass recalcitrance will require a greater understanding of plant cell walls from synthesis to deconstruction. This understanding would generate models, theories and finally processes that will be used to understand and overcome biomass recalcitrance. This grand challenge calls for an integrated research approach, which is illustrated by

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

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

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

64

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

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

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

http://bioenergycenter.org

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

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

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

Though much work thus far has investigated biomass deconstruction, recent efforts in our lab have focused on uncovering the mechanisms of biomass formation. In this regard, we have developed an experimental approach that combines an enhanced cell lysis and proteome solubilization protocol with state-of-the-art LTQ-Velos MS technology

to achieve the deepest proteome coverage of *Populus* to date. In response to the enhanced genomic complexity inherent to eukaryotic species, especially plants, we have optimized our post-data acquisition informatics to resolve ambiguous protein identifications resulting from the substantive genetic redundancy exhibited by *Populus* spp. Clustering redundant proteins prior to mapping peptides to the proteome allowed us to more thoroughly characterize *Populus* organ-specific protein information, focusing on comparing and contrasting the proteomes of leaf (young and old), stem, and root samples. Taken together, these MS-based proteome approaches provide remarkable insight into the fundamental mechanisms of biomass formation by plants as well degradation by cellulolytic microbes as a step to enhance biofuel production.

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

65

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

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

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

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC biomass formation and modification research involves working directly with two potential bioenergy crops (switchgrass and *Populus*) to develop varieties that are easier to break down into fermentable sugars. We are using both testing and generation of large number of natural and modified plant samples as well as developing genomic tools for detailed studies into poorly understood cell wall biosynthesis pathways.

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

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

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

66

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

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

[1]The Samuel Roberts Noble Foundation, Ardmore, Okla.; [2]University of Tennessee, Knoxville; [3]National Renewable Energy Laboratory, Golden, Colo.; and [4]BioEnergy

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

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). Addressing the roadblock of biomass recalcitrance will require a greater understanding of plant cell walls from synthesis to deconstruction. This understanding would generate models, theories and finally processes that will be used to understand and overcome biomass recalcitrance. This grand challenge calls for an integrated research approach, which is illustrated by the project described below. BESC research is multidisciplinary by design and multi-institutional in composition.

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

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

67

Consolidated Bioprocessing Conversion of Genetically Modified Switchgrass

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

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

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). Addressing the roadblock of biomass recalcitrance will require a greater understanding of plant cell walls from synthesis to deconstruction. BESC research in biomass deconstruction and conversion targets CBP by studying model organisms and thermophilic anaerobes to understand novel strategies. This grand challenge calls for an integrated research approach, which is illustrated by the project described below.

Consolidated bioprocessing (CBP) conversion comprises biomass fermentative microorganisms that produce some or all of the needed biomass hydrolytic enzymes, mitigating the cost of added enzymes. It has been shown previously that down regulation of the caffeic acid O-methyltransferase (COMT) gene in switchgrass (*Panicum virgatum*) reduced lignin content, reduced S/G ratio, increased sugar release, and improved the yeast-based SSF conversion of the modified switchgrass line compared to the wild-type switchgrass (Fu, et al. 2011). As expected, after either dilute acid or hot water pretreatment of switchgrass provided by the Noble Foundation, yeast-based SSF conversion showed the COMT transgenic switchgrass yielded at least 25% more ethanol than the wild-type regardless of the pretreatment process. To follow-up this work, we examined the fermentation of both wild-type and COMT transgenic switchgrass lines using anaerobic bacteria including *Thermoanaerobacterium saccharolyticum* and very active consolidated bioprocessing (CBP) microorganisms *Clostridium thermocellum*, *Caldicellulosiruptor obsidiansis*, and *Caldicellulosiruptor bescii.* Comparison of the wild-type switchgrass to a modified COMT plant line after either dilute acid or hot water pretreatment showed that the transgenic switchgrass yielded over 20% greater total fermentation products on a (g/g) substrate basis when fermented by *Thermoanaerobacterium saccharolyticum* after addition of enzymes ala the SSF mode. Biomass fermentation by *C. thermocellum*, *C. obsidiansis*, or *C. bescii* requires no added industrial enzymes but switchgrass fermentation showed differential inhibition with wild-

type and COMT transgenic switchgrass. Specifically, wild-type and COMT transgenic switchgrass were pretreated with dilute acid (0.5% H_2SO_4, 180°C, 7.5 min.) followed by water washing of the solids. Bioconversion of the solids using *C. thermocellum* showed partial inhibition of fermentation with efficient hydrolysis liberating free sugars, but only partial conversion of these sugars to fermentation products. The COMT transgenic feedstock showing greater inhibition compared to the wild-type switchgrass. However, full fermentation capabilities for *C. thermocellum* were restored after an extra hot water extraction of soluble components from the wild-type and transgenic switchgrass and the COMT line produced about 25% more total products than from the wild-type switchgrass, as seen with yeast and *Thermoanaerobacterium saccharolyticum* SSF conversion. However, using the same biomass sources processed identically with extensive water extraction, both *Caldicellulosiruptor* species failed to ferment the transgenic switchgrass (ca. 7-10% of expected) while successfully fermenting the wild-type switchgrass. This suggests *Caldicellulosiruptor* can be used as an indicator for fermentation inhibition with different biomass sources.

Reference

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

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

68

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

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

[1]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Georgia Institute of Technology, Atlanta; [3]The Samuel Roberts Noble Foundation, Ardmore, Okla.; [4]University of Tennessee, Knoxville; and [5]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). Addressing the roadblock of biomass recalcitrance will require a greater understanding of plant cell walls from synthesis to deconstruction. This understanding would generate models, theories and finally processes that will be used to understand and overcome biomass recalcitrance. This grand challenge calls for an integrated research approach, which is illustrated by the project described below. BESC research is multidisciplinary by design and multi-institutional in composition.

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

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

69

Flowthrough Pretreatment to Characterize Biomass Deconstruction

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

[1]University of California, Riverside; [2]Georgia Institute of Technology, Atlanta; [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [4]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC researchers in characterization, modeling, and data management areas are engaged in 1) applying advanced technologies to analyze chemical and structural changes within biomass and 2) storing, tracking, analyzing and integrating data and understanding across the Center.

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

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

70

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

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

[1]North Carolina State University, Raleigh; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]University of Georgia, Athens; and [4]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC research in biomass deconstruction and conversion targets CBP by studying model organisms and thermophilic anaerobes to understand novel strategies and enzyme complexes for biomass deconstruction.

Consolidated bioprocessing (CBP) of second-generation biofuels from lignocellulose will require that the microbe(s) be able to deconstruct complex polysaccharides from plant biomass. The extremely thermophilic, non-cellulosomal genus *Caldicellulosiruptor* have come under renewed interest due to their plant biomass-degrading abilities at high

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

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

71

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

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

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

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC research in biomass deconstruction and conversion targets CBP by studying model organisms and thermophilic anaerobes to understand novel strategies and enzyme complexes for biomass deconstruction.

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

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

72

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

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

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

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC research in biomass deconstruction and conversion targets CBP by studying model organisms and thermophilic anaerobes to understand novel strategies and enzyme complexes for biomass deconstruction.

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

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

73

Improving Ethanol Production in *Clostridium thermocellum*

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

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

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC research in biomass deconstruction and conversion targets CBP by studying model organisms and thermophilic anaerobes to understand novel strategies and enzyme complexes for biomass deconstruction.

Clostridium thermocellum has been considered as a candidate for biofuel production due to its ability to rapidly solubilize crystalline cellulose and produce ethanol. Low ethanol yield and titer are currently the primary obstacles preventing *C. thermocellum* from further consideration as a consolidated bioprocessing organism. Recent development of a system for genetic modification of this organism has resulted in the creation of more than 30 mutants, some of which exhibit dramatic improvements in ethanol yield. Analysis of these strains using a metabolic flux analysis framework has allowed us to understand the role of electron carrier flexibility in determining ethanol yield.

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

74

Thermophilic Isobutanol Production in *Geobacillus thermoglucosidasius*

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

[1]University of California, Los Angeles and [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC research in biomass deconstruction and conversion targets CBP by studying model organisms and thermophilic anaerobes to understand novel strategies and enzyme complexes for biomass deconstruction.

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

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

75

Engineering Improved Cellulases for CBP Microbes

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

[1]National Renewable Energy Laboratory, Golden, Colo.; [2]Cornell University, Ithaca, N.Y.; and [3]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves 1) designing plant cell walls for rapid deconstruction and 2) developing multitalented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC research in biomass deconstruction and conversion targets CBP by studying model organisms and thermophilic anaerobes to understand novel strategies and enzyme complexes for biomass deconstruction.

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

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

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