

# Systems Biology for DOE Energy and Environmental Missions

## Bioenergy

### Biofuels > Bioenergy Research Centers

#### Joint Bioenergy Institute (JBEI)

#### Systematic Characterization of Glycosyltransferases Involved in Plant Cell Wall Biosynthesis

**Henrik Vibe Scheller**<sup>1\*</sup> (hscheller@lbl.gov), Ai Oikawaa, Lan Yin,<sup>1,2</sup> Eva Knoch,<sup>1,2</sup> Naomi Geshi,<sup>2</sup> Carsten Rautengarten,<sup>1</sup> Yuzuki Manabe,<sup>1</sup> and Chithra Manisseri<sup>1</sup>

<sup>1</sup>Feedstocks Division, Joint Bioenergy Institute, Emeryville, Calif.; and <sup>2</sup>Dept. of Plant Biology and Biotechnology, University of Copenhagen, Denmark  
<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

The goals of the project are to provide a detailed understanding of the enzymes that are responsible for biogenesis of the plant cell wall and develop a knowledgebase to enable generation of crop plants with improved properties as feedstocks for biofuel production.

Plant cell walls are composed mainly of polysaccharides and production of biofuels from biomass requires decomposition of the polymers. Many of the polymers are recalcitrant to degradation and they are composed of sugars that are not optimal for fermentation. Better understanding of the biosynthesis of the cell wall polysaccharides may enable development of crops with improved properties as biofuels feedstocks. Despite rather detailed information on the structure of the cell wall polysaccharides, little is known about their biosynthesis. The key enzymes are glycosyltransferases (GTs) and plants need a large number of GTs to synthesize the complex polysaccharides present in the walls. However, only a few GTs have had their activity demonstrated. In *Arabidopsis thaliana*, approximately 450 GT genes have been identified based on their sequence and deposited to

the CAZy database ([www.cazy.org](http://www.cazy.org)). We have cloned a large number of these GTs in Gateway vectors in order to heterologously express the GTs and characterize their activity. Systematic analysis of the GTs is in progress and results will be presented. *Agrobacterium*-mediated transient expression in *Nicotiana benthamiana* has a high success rate for expression of GTs.

An alternative way to elucidate the function of GTs and other biosynthetic enzymes is to study the effect of down-regulating or inactivating the corresponding genes. This approach is often hampered by the overlapping function of many GTs. Generation of mutants that are affected in several homologous genes can overcome this limitation. In other cases, GTs are functioning in complexes that contain several different polypeptide subunits. The CslD family of proteins provides examples of both redundancy in function and of protein complexes.

#### Analysis of Putative Feruloyltransferase Transcript Levels and Cell Wall Composition During Rice Development

**Laura E. Bartley**<sup>1,2\*</sup> (LEBartley@lbl.gov), Chithra Manisseri,<sup>2</sup> Brian Williams,<sup>1,2</sup> Dawn Chiniquy,<sup>1,2</sup> Jesper Harholt,<sup>2</sup> Henrik Vibe Scheller,<sup>2</sup> and Pamela C. Ronald<sup>1,2</sup>

<sup>1</sup>Dept. of Plant Pathology, University of California, Davis, Calif.; and <sup>2</sup>Joint BioEnergy Institute, Emeryville, Calif.  
<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

Identification of genes encoding cell wall modifying enzymes has applications in human and animal nutrition, plant defense against pathogens, and biofuel production efficiency. For plants with type II cell walls, such as rice and other grasses, glucuranoarabinoxylans (GAX) are a major component of the hemicellulose found in both primary and secondary walls. Ferulic acid, a phenolic compound, is added to the O-5 of arabinosyl units of GAX. Ferulic acid residues can covalently crosslink arabinosyl residues through dimer formation, or serve as attachment points between

GAX chains and lignin, providing rigidity to the cell wall matrix. Therefore, reducing the ferulic acid content of Type II cell walls may improve feedstock deconstruction for production of lignocellulosic biofuels. The genes that encode feruloyltransferases (FTs) have not yet been identified. However, Mitchell and colleagues (Plant Physiology 2007) have recently predicted that FTs may be encoded by a clade of genes from the CoA acyl transferase superfamily based on their greater EST abundance in grasses compared with dicots. Ratios of cell wall components, including GAX and ferulic acid, are dynamic, adjusting to allow for elongation while plants are young, and rigidity, as they mature. To correlate cell wall composition and gene expression during rice development, we have collected RNA and cell wall samples from various rice tissues throughout development. We are conducting quantitative RT-PCR to amplify transcripts from the putative FT genes identified by Mitchell et al. Cell wall sugar and phenolic composition will be examined by HPLC analysis. This study will determine how rice cell wall composition shifts to accommodate the changing needs of the plant, providing valuable data for further rice genomics analyses and evidence to test the hypothesis of Mitchell et al. regarding the putative FTs. Genes from the 20-member gene family that show correlation with ferulic acid accumulation will be targeted for functional studies toward improving the deconstruct-ability of grass cell walls for biofuels.

complete absence of many genes produced by such techniques results in a lethal phenotype as the gene is absolutely required for normal function of the plant. A more subtle approach utilizes genetic differences in naturally occurring variants to provide important information about gene function and genetic diversity. In collaboration with the Joint Genome Institute we have sequenced two *Arabidopsis* accessions (Bay-0 and Sha-0) that have previously been shown to have measurable differences in Ara-Rha ratios in cell wall extracts. This genetic information and the utilization of recombinant inbred lines (RILs) will be used to map QTLs identified in these and other *Arabidopsis* accessions to identify loci that contribute to functional differences in plant cell walls.

GTL

### Selection, Cloning and Functional Characterization of Rice-Diverged, Cell Wall-Related Glycosyltransferases

**Miguel Vega-Sánchez\*** (mevegasanchez@ucdavis.edu), Laura Bartley, Peijian Cao, and Pamela Ronald (pchronald@ucdavis.edu)

Dept. of Plant Pathology, University of California, Davis, Calif.; and Joint BioEnergy Institute, Feedstocks Division, Emeryville, Calif.

<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

Understanding plant cell wall biosynthesis is crucial for the development of the next generation of biofuels derived from lignocellulosic material. Current limitations in the harvest of fermentable sugars from cellulose derive from the inherent recalcitrance of plant cell walls. Basic knowledge of how the structure and composition of the cell wall can be modified to obtain biomass suitable for efficient and economically viable biofuel production is needed. Glycosyltransferases (GTs) are a large, multifamily class of enzymes that form glycosidic bonds between donor nucleotide sugars and acceptor substrates. Among the GTs are the enzymes responsible for the synthesis of important cell wall polysaccharides, including cellulose, hemicellulose and callose. However, the function, substrate specificity and biochemical activity of the majority of GTs are unknown. Many aspects of grass and other commelinoid monocot cell walls are distinct from that of better-studied dicots. As many of the preferred feedstocks for future bioenergy initiatives include grasses such as switchgrass and *Miscanthus*, the model species of choice to advance our understanding of monocot-specific aspects of the cell wall is rice (*Oryza sativa*). Rice was the first grass species to have its full genome sequenced and abundant

GTL

### The Utilization of *Arabidopsis* Genetic Variants to Understand Cell Wall Structure and Biosynthesis

**Joshua L. Heazlewood\*** (JLHeazlewood@lbl.gov), Katy M. Christiansen, Dominique Loque, and A. Michelle Smith

Feedstocks Division, Joint BioEnergy Institute, Emeryville, Calif.

<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

The process of plant cell wall biosynthesis involves a complex series of biochemical processes involving many hundreds of proteins. Determining the function of a specific gene through functional genomics techniques have proved problematic due to genetic redundancies and undetectable changes. Common techniques have used mutant collections in forward genetic screens or reverse genetics to directly target and disrupt genes of interest. Since such techniques are heavily reliant on some phenotypic discrimination to assess gene function, the absence of a measurable difference often results in little useful information. Furthermore the

functional genomic resources have accumulated in recent years. To identify potential targets involved in cell wall synthesis, we mined the recently created and publicly available rice GT phylogenomic database (<http://ricephylogenomics.ucdavis.edu/cellwalls/gt/>) to select a list of 33 rice-diverged GTs with high expression in above-ground, vegetative tissues. Cloning of these genes is underway, and will be followed by the creation of a variety of expression constructs for functional analysis in both transgenic plants and heterologous systems for protein expression, biochemical activity determination and protein-protein interaction network generation. We have ordered insertion and activation-tagged lines from worldwide repositories for cell wall composition/modification analyses as well as phenotypic characterization. Among this group of genes, we will study in detail a subset within GT family 2, the grass-specific Cellulose synthase like (Csl)-F and H sub-families, by generation of gene family knockdowns using artificial miRNAs. We will characterize other putative GTs from unknown families as well. We anticipate that by focusing on grass-specific GTs involved in cell wall synthesis and modification in plant tissues important for the production of lignocellulosic-derived biofuels, we can facilitate the improvement of genetic traits in second-generation bioenergy crops such as switchgrass.

GTL

## Monolignol Transporters and Cell Wall Oxidases Screens

Fan Yang\* (Fyang@lbl.gov), Nakako Shibagaki, and Dominique Loque

Joint BioEnergy Institute, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.

<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

Currently, biofuels, such as ethanol are produced largely from starch that comes from grains, but it represents only a little proportion of sugar polymer availability on Earth. Large quantities of sugar from polysaccharides that are not utilized thus far are cellulose and hemicellulose, which are the main constituent of plant cell walls. The energy efficiency of starch-based biofuels is however not optimal, while plant cell walls (lignocellulose) represent a huge resource for BioEnergy since plant cell walls are composed of 95-70% of sugar. The rest of the plant cell walls (5-30%) is mainly composed of lignin, which is a very strong phenolic polymer recalcitrant to degradation and inhibits efficient extraction of sugars from the cell wall thus prevents cost-effective lignocellulosic ethanol production. Unfortunately, lignin provides compressive resistance to plant cells and cannot simply be genetically removed without incurring deleterious

consequences on plant productivity. Lignin gives a strong structural support to the plant but also protects the plant against biotic (e.g. pathogens) and abiotic stress (e.g. UV, wind). Therefore, it is important to understand to which extent plant lignin content and composition can be modified in useful ways without deleterious consequences to plant growth and development. Since lignin cannot be removed, a better control of lignin deposition and cross-linking within the cell wall may increase sugar recovery from the cell wall polysaccharides. Another approach would be to replace the "hard bounds" (e.g. ether, carbon bonds) in the lignin polymer with easily cleavable ones (e.g., amide or ester bonds) and thus would facilitate removal of lignin with processes that are much more cost effective than current processes.

In order to be able to modify lignin deposition, polymerization and cell wall cross-linking, a better knowledge of the enzymes (oxidases) that participates in the polymerization of lignin and cross-linking to other cell wall components is required. Similarly, to be able to control monolignol export into the apoplast, monolignol transporters need to be identified and characterized. Therefore, we are currently developing a strategy using yeast complementation to try to identify protein involved in lignin deposition. Plant cDNA libraries, a complete *Arabidopsis* MDR transporters library and selected genes will be heterologous expressed in yeast and tested for their ability to detoxify phenolic compounds by export or polymerization.

GTL

## Starting Point for Enzymatic Hydrolysis for Cellulose: Enzyme Engineering of Glycoside Hydrolase-5 Endoglucanases

Blake A. Simmons<sup>1,2\*</sup> (basimmons@lbl.gov), Zhiwei Chen,<sup>1,2</sup> Supratim Datta,<sup>1,2</sup> Danielle T. Ercek,<sup>1</sup> Hanbin Liu,<sup>1,2</sup> Josh I. Park,<sup>1,2</sup> Huu Tran,<sup>1,2</sup> Swapnil R. Chhabra,<sup>1</sup> Masood Z. Hadi,<sup>1,2</sup> and Rajat Sapra<sup>1,2</sup>

<sup>1</sup>Joint BioEnergy Institute, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.;

and <sup>2</sup>Sandia National Laboratories, Livermore, Calif.

<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

Enzymes are catalysts that have a set of specific and well defined characteristics like substrate, pH, temperature and kinetics etc. The vast amount of available genomic data shows that orthologous genes code for proteins that vary in these characteristics and the changes in properties at the protein level can be traced to sometimes minor changes like substitutions in the sequence at the molecular level. This observation that changes in the sequence leads to changes in

characteristics can be used to design new and novel protein variants using two major approaches—directed evolution and rational design. We propose to use these techniques to engineer enzymes with improved activity and characteristics better suited to optimal conditions for cellulosic biomass deconstruction. Directed evolution of enzymes is a powerful technique that takes advantage of the Darwinian process of natural selection on a high throughput lab scale for the generation of protein mutants, called variants, which are then screened and selected for improved desirable traits as compared to the parent protein characteristics. At the core of the technique is the principle that incremental changes acquired either through mutagenesis or via recombination lead to a better variant when selection pressure is applied; the selection pressure can be any of the characteristics that are sought to be improved upon like activity, kinetics, pH or temperature stability and, in some cases, different substrates and novel reactions. The advantage for using the directed evolution approach is that there is no requirement for the availability of an extensive data set of orthologs as a starting point; only a gene sequence and a screening method for ‘evolving’ a protein function is needed. Therefore, this technique can be employed in all the enzymes currently known to be involved in cellulosic biomass deconstruction- cellulases for which there are a large number of gene sequences available, xylanases and ligninases etc. for which there is scant gene and enzyme data available in the database.

limit the enzymatic accessibility of the recovered polysaccharides for efficient saccharification into sugars. Various physical and chemical pretreatment methods are currently employed to break down the biomass recalcitrant structures, and increase their susceptibility to enzymes. Among these techniques, dilute acid pretreatment has been shown as a leading pretreatment process. However, dilute acid hydrolysis can lead to degradation products that are often inhibitory and significantly lower the overall sugar yields. Glucose and xylose degradation products that result from the pretreatment methods include hydroxymethylfurfural (HMF) and furfural, which produce levulinic and formic acids, respectively, which inhibit the subsequent fermentation of sugars to ethanol. Recently, ionic liquids have demonstrated great promise as efficient solvents for biomass with easy recovery of cellulose upon anti-solvent addition. However, to date, no comprehensive side-by-side comparative analysis has been conducted in order to evaluate the dilute acid and ionic liquid biomass pretreatment processes. In this study, we compare ionic liquid and dilute acid pretreatments acting on switchgrass with numerous analytical techniques to gain a better understanding of both techniques and the resultant saccharification yields of fermentable sugars.

GTL

GTL

## A Comparative Study of Dilute Acid and Ionic Liquid Pretreatment of Biomass and Model Lignocellulosics

Chenlin Li<sup>1,2\*</sup> (Cli@lbl.gov), Rohit Arora,<sup>1,2</sup> Chithra Manisseri,<sup>1</sup> Ken Vogel,<sup>3</sup> Blake Simmons,<sup>1,2</sup> and Seema Singh<sup>1,2</sup>

<sup>1</sup>Joint BioEnergy Institute, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Emeryville, Calif.; <sup>2</sup>Energy Systems Dept., Sandia National Laboratories, Livermore, Calif.; and <sup>3</sup>United States Department of Agriculture, Lincoln, Nebr.

<http://www.jbei.org>

**Project Goals: JBEI’s primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

Lignocellulosic biomass has great potential to serve as a low-cost and abundant feedstock for bioconversion into fermentable sugars, which can be subsequently fermented into biofuel. There are several obstacles that remain to be solved in this conversion process, specifically lignin content, crystalline cellulose structure and the presence of ester linkages between lignin and hemicellulose in the plant cell wall

## Understanding Ionic Liquid Pretreatment of Lignocellulosic Biomasses

Seema Singh<sup>1\*</sup> (SSingh@lbl.gov), Rohit Arora,<sup>1</sup> Chithra Manisseri,<sup>1</sup> Chenlin Li,<sup>1</sup> Henrik Vibe Scheller,<sup>1</sup> Ken Vogel,<sup>2</sup> and Blake Simmons<sup>1</sup>

<sup>1</sup>Joint BioEnergy Institute, Emeryville, Calif.; and <sup>2</sup>United States Department of Agriculture, Lincoln, Nebr.

<http://www.jbei.org>

**Project Goals: JBEI’s primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

Pretreatment of biomass is essential for breaking apart highly ordered and crystalline plant cell walls and loosening the lignin and hemicellulose conjugation to cellulose microfibrils, thereby facilitating enzyme accessibility and adsorption and reducing costs of downstream saccharification processes. Recent reports<sup>1,2</sup> have shown very high yields at very low enzyme loadings. However, pretreatment still remains one of the most costly steps in lignocellulosic biofuel production. Ionic liquids are novel solvents showing great promise for lignin and cellulose solubilization. Instant rejection of dissolved polysaccharides upon addition of anti-solvent shows promise for recyclability in addition to other desired attributes like low volatility, non-flammability and thermal stability. Although ionic liquids have been shown to be very effective in cellulose solubilization<sup>3,4</sup>, the disposition of hemicellulose and lignin are not fully understood. The

aim of our research is to develop a fundamental understanding of ionic liquid pretreatment by monitoring and analyzing process streams. To that end, we have employed HPAEC, XRD, FTIR, NIR, and SEM to study the impact of ionic liquid pretreatment on switchgrass and corn stover. We will present the results from these measurements in the context of developing and selecting optimized ionic liquid pretreatment conditions for selective depolymerization of either cellulose or lignin, whereby fractionation of different cellulosic and lignin components could be realized.

## References

1. Teymouri, F., Laureano-Perez, L., Alizadeh, H., & Dale, B. E. (2005). *Bioresource Technology*, 96, 2014–2018.
2. Kim, T. H., Kim, J. S., Sunwoo, C., & Lee, Y. Y. (2002). In *24th Symposium on Biotechnology for Fuels and Chemicals*.
3. Dadi et al, *Applied Biochemistry and Biotechnology*, 136–140, 407.
4. Singh, S. & Simmons, B, In-Situ Examination of Biomass Dissolution and Cellulose Regeneration Enabling Cellular level Insight of Ionic Liquid Pretreatment Process (manuscript in preparation).

GTL

## JBEI Microbial Communities Deconstruction Research Activities

**Terry C. Hazen\*** (TCHazen@lbl.gov), Phil Hugenholtz, Steve Singer, Jean VanderGheynst, Patrik M. D'haeseleer, Michael P. Thelen, Kristin DeAngelis, Amitha Reddy, Martin Allgaier, Julian Fortney, Gary Andersen, Todd DeSantis, Eoin Brodie, Cindy Wu, Dara Goodheart, Mary Firestone, Whendee Silver, and Blake Simmons

Joint BioEnergy Institute, Deconstruction Division, Microbial Communities Dept., Lawrence Berkeley National Laboratory, Berkeley, Calif.

<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

There exist several unexplored microbial communities that are promising sources of biocatalysts and metabolic pathways for hydrolysis of polysaccharides in plant cell walls (lignocellulose), such as those found on the tropical rain forest floor and within compost. The compost environment is nutrient rich, whereas the rain forest floor is nutrient deficient. Our approach is to examine both communities to increase the likelihood of discovering a variety of novel pathways for the biological deconstruction of lignocellulose. The tropical rain forest has developed metabolic pathways that quickly convert the available biomass into useful metabolites as evidenced by the scarcity of litter on

the forest floor despite high net primary productivity. These communities therefore have great potential to provide enzymes with very high specific activities and rapid kinetics of biomass degradation. In contrast, composting is an industrial process managed under thermophilic conditions. Microbial communities in compost are dynamic and vary based on the extent of decomposition of the compost, moisture, temperature, and oxygen concentration. Studying compost communities, in particular composts enriched in lignocellulosic biomass (greenwaste composts), will facilitate the discovery of novel enzymes and pathways amenable to certain high-temperature, high ionic strength, and non-neutral pH biomass pretreatment conditions, as well as insights into how such pathways are regulated for optimal lignocellulose degradation in their respective environments. Forests, and in particular tropical rain forests, have evolved to obtain the maximum nutrients from leaf litter. Thus, over many millions of years, the soil microbial communities in these environments have evolved to recycle all biomass that results from the episodic and changing forest environment. For example, the microbial communities present in the wet tropical forest soils of Puerto Rico have optimized the conversion of lignin and cellulose to methane and methanol through 40–60 million years of evolutionary selection for rapid and complete decomposition of plant material. Wet tropical forests typical of those in Puerto Rico have some of the highest known rates of plant productivity on Earth and concomitantly high microbial rates of biomass breakdown. Due to abundant rainfall and high carbon availability, these soils are frequently anaerobic leading to methane production during decomposition. Studies have confirmed that because of this frequent anaerobiosis, prokaryotes assume a primary role in macromolecular decomposition. Wet tropical soils will provide fertile ground for the discovery of novel lignin and cellulose degrading enzymes, and for new and unique consortia capable of effectively functioning under a wide range of environmental conditions. The aerobic-anaerobic cycling that naturally occurs over time and space in these wet forest soils has made selected bacterial and archaeal communities extremely effective in converting lignocellulose to methane. In composting prokaryotes are dominant in the early stages of decomposition, while both prokaryotes and eukaryotes play a role later in the process. Selecting incubation conditions that enrich both groups is an important part of our experimental approach. The Microbial Communities Department will identify and isolate key lignocellulolytic enzymes and metabolic pathways that will be delivered to the Deconstruction Division enzyme engineering group and the Fuels Synthesis Division, respectively of JBEI. During the past year this group has constructed and tested a new MycoChip microarray for rapid identification of more than 11,620 fungal taxa. Early in the year we placed bags of switch grass and lignin baited “bug traps” in the soil of the tropical rain forest to enrich for lignocellulose degraders. In only 1 week we observed increases in phenol oxidase and peroxidase. We have also observed a very high microbial diversity in these samples. Respirometer studies of switch grass with tropical rain forest soil revealed very high rates of production of CO<sub>2</sub>, CH<sub>4</sub> and H<sub>2</sub>S from the inoculated switchgrass and more than 156 different taxa from the soil alone. We also established optimal wetting conditions for

decomposition of corn stover and switchgrass inoculated with green waste compost and incubated under aerobic and thermophilic conditions. We determined that there were more than 2,602 bacterial taxa associated with switchgrass inoculated with compost. Preliminary results on Titanium 454 pyrosequencing of one tropical rain forest soil sample revealed more than 2,700 cellulases, hemicellulases, and ligninases. Similar data on a compost sample are currently being analyzed. We intend to characterize the enzyme repertoire present in the metagenomic samples, and correlate that with differences in activity and community composition between compost and tropical rain forest soil.

—  
GTL

## Discovery and Optimization of Lignocellulolytic Bacteria From Puerto Rican Rainforest Soils

**Kristen M. DeAngelis\*** (KDeAngelis@lbl.gov),  
Whendee L. Silver, Julian Fortney, and Terry C. Hazen

Joint BioEnergy Institute, Emeryville, Calif.

<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

Tropical soils in Puerto Rican rain forests have some of the highest decomposition rates recorded in the world, with almost total mass lost in decomposing plant material within one year. These soils are capable of deconstructing biofuel plant materials to basic components, like ethanol, methane or methanol. Rapidly fluctuating redox conditions are characteristic of the highly weathered soils of upland humid tropical forests, which are dominated by Fe-oxide mineralogy and have relatively low sulfate availability. The frequent episodes of anoxic conditions make it likely that these decomposing consortia are primarily bacteria, not fungi as are usually observed in temperate systems. Previous lab incubation under fluctuating redox conditions permitted simultaneous methanogenesis, N<sub>2</sub>O production, and iron reduction, all accompanied by steady CO<sub>2</sub> production. The objective of this research is to define field conditions that result in characteristic high methane production in Puerto Rico forest soils from decomposing plant materials, and determine whether different microbial communities break down different plant materials. Towards this end, we designed a field experiment and accompanying laboratory incubations that would allow us to investigate the rates, controls and mechanisms of switchgrass decomposition in tropical rainforest soils. In June of 2008, we buried litterbags filled with switchgrass in four different forest types in the Luquillo LTER, located at the El Yunque National Forest in Puerto Rico, USA. The four forest types vary from more aerobic soils, warmer temperatures and annual precipitation

on the order of 1,000 mm, to fluctuating redox soils, to mostly anaerobic soils, cooler soil temperatures and annual precipitation that can exceed 4,000 mm.

The experimental design included 4 field sites, 6 time points, and bags buried in pairs, one for. At each of 6 time point, litter bags and soil are collected from the field and assayed for microbial community analysis using 16S ribosomal DNA PhyloChip, potential enzyme activity ( $\beta$ -glucosidase, endoglucanase xylosidase, chitinase, phenol oxidase and peroxidase), and mass loss as indicators of decomposition. In the driest site, which we expect to also have the highest rates of decomposition, we also buried biosep beads baited with lignin (using unbaited beads as controls) as bug traps to identify and isolate microbes specifically able to decompose lignin. This site was also instrumented with oxygen sensors to measure oxygen levels in soil on an hourly basis over the course of this year-long incubation, and ultimately to correlate decomposition, enzyme activity and microbial community composition with oxygen availability at the end of the experiment. Concomitantly with the field experiment, we are using fresh soil to inoculate mini-reactors with dried ground switchgrass and incubate anaerobically to enrich for lignocellulose-degrading organisms. The initial inoculation of rain forest soil with switchgrass resulted in significant CO<sub>2</sub>, CH<sub>4</sub> and H<sub>2</sub>S production compared to uninoculated, anaerobic soil incubations, as well as a substantial change in microbial community composition. Switchgrass amendment resulted in significant change in 147 taxa compared to the 1847 detected in the soils. With switchgrass addition to soil, Archaea, methanogens, enteric bacteria, *Bacilli* and *Clostridia* were significantly increased, while *Acidobacteria*, *Burkholderia* and *Verrucomicrobial* were significantly reduced in the microbial community. Further passages of the soil microbial community with switchgrass as the sole carbon source has resulted in a low-richness, anaerobic microbial community capable of efficiently converting switchgrass to methane and carbon dioxide as well as depolymerizing cellulose, hemicellulose, and lignin in the process.

—  
GTL

## Metagenomic Characterization of Compost and Rain Forest Soil Microbial Communities

**Martin Allgaier\*** (MAllgaier@lbl.gov), Amitha Reddy,<sup>1,2</sup> Jean VanderGheynst,<sup>2</sup> Alex Copeland,<sup>3</sup> Victor Kunin,<sup>3</sup> Patrik D'haeseleer,<sup>4</sup> Kristen DeAngelis,<sup>1,5</sup> Julian Fortney,<sup>1,5</sup> Dylan Chivian,<sup>1,5</sup> Paramvir S. Dehal,<sup>5</sup> Blake Simmons,<sup>1,6</sup> Terry C. Hazen,<sup>1,5</sup> and Phil Hugenholtz<sup>1,3</sup> (PHugenholtz@lbl.gov)

<sup>1</sup>Joint BioEnergy Institute, Emeryville, Calif.; <sup>2</sup>Dept. of Biological and Agricultural Engineering, University of California, Davis, Calif.; <sup>3</sup>DOE Joint Genome Institute, Walnut Creek, Calif.; <sup>4</sup>Lawrence Livermore National Laboratory, Livermore, Calif.; <sup>5</sup>Lawrence Berkeley

National Laboratory, Berkeley, Calif.; and <sup>6</sup>Sandia National Laboratories, Livermore, Calif.

<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

Microorganisms are a promising source of novel carbohydrate-active enzymes (cellulases, hemicellulases, and ligninases) since they are primarily responsible for plant biomass degradation in nature. However, most carbohydrate-active enzymes used by industry come from only a few model organisms. To identify new lignocellulolytic enzymes we shotgun sequenced genomic DNA from two sources: a sample of pristine Puerto Rican rain forest soil and a sample obtained from a solid state fermentation experiment in which switchgrass was incubated 30 days in a lab under mesophilic and thermophilic conditions after inoculating with green waste compost from an industrial facility. Both ecosystems display high rates of plant biomass degradation and are therefore prime targets for novel carbohydrate-active enzyme discovery.

454-Titanium pyrosequencing was used to generate metagenome data sets from the two samples resulting in a total of 1,412,492 reads (rain forest: 863,759; compost: 548,733) with an average read length of 424 bases. Reads were quality filtered and trimmed in preparation for comparative analyses with the metagenome analysis tools IMG/M and MicrobesOnline. The complexity of the rain forest soil metagenome precluded assembly, so sequence data were analyzed unassembled. However, for the compost sample, significant assembly occurred resulting in contigs up to 50 kb in length.

Preliminary comparative analysis of a fraction of the rain forest soil sequence data revealed more than 2,700 cellulases, hemicellulases, and ligninases including glycoside hydrolases as well as glycosyl transferases, representing ~1% of all predicted protein-coding genes (e.g. compared to 1.2% or 0.03% identified in metagenomic data sets from termite guts or silage surface soil, respectively). The enzyme repertoires present in the two metagenome data sets will be further analyzed to identify new deconstruction enzymes and compared to assess differences in activity and community composition between compost and rain forest soil.

## The Fuels Synthesis Division of the Joint BioEnergy Institute (JBEI)

**E. Baidoo**, H.R. Beller, R. Chan, S. Chhabra\* (SRChhabra@lbl.gov), H. Chou, R. Dahl, Z. Dmytriv, M. Dunlop, C. Fortman, D. Garcia, H. Garcia Martin, J. Gilmore, J. Gin, E.-B. Goh, J. Haliburton, T. Ham, C. Joshua, Y. Kang, R. Krupa, S.K. Lee, T.S. Lee, C. Liu, A. McKee, A. Mukhopadhyay, F. Nowroozi, M. Ouellet, P. Peralta-Yahya, N. Prasad, S. Rodriguez, B. Rutherford, E. Steen, and J.D. Keasling

Joint BioEnergy Institute, Emeryville, Calif.

<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

The Fuels Synthesis Division of JBEI aims to engineer microbial host platforms and pathways for the production of fuels from lignocellulose hydrolysates obtained from JBEI's Deconstruction Division. Host platforms being employed by JBEI include *E. coli*, *S. cerevisiae*, and *Sulfolobus spp.* Host engineering includes a variety of tasks, such as metabolic engineering that combines native and non-native pathways, assessment of bottlenecks in metabolic flux and toxic effects of metabolites, gene discovery, and rational and combinatorial strain evolution. Collaborative efforts with JBEI's Technology Division include systems biology analytical and computational tools (genomics, transcriptomics, proteomics, metabolomics, fluxomics), robotics, and imaging. Of the broad range of fuels of interest to JBEI, efforts to date have focused on the following: (1) short-chain alcohols (e.g., butanol), (2) isoprenoid-based fuels (e.g., isopentenol), and (3) fatty acid-based fuels. In this poster, we present some of our strategies and results.

## Building a *de novo* Synthetic Metabolic Pathway for Producing Branched-C5 Alcohols

**H.H. Chou**<sup>1,3,4\*</sup> (HHChou@lbl.gov) and J.D. Keasling<sup>1,2,3,4</sup>

<sup>1</sup>Dept. of Bioengineering and <sup>2</sup>Dept. of Chemical Engineering, University of California, Berkeley, Calif.;

<sup>3</sup>Joint Bioenergy Institute and <sup>4</sup>Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.

<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofu-**

**els – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

Many of the chemicals currently proposed as biofuels are derived from natural pathways, *i.e.* they already exist in nature. However, these natural chemicals were not designed to be fuels, so they have many costs associated with their use. The ability to build unnatural pathways to synthesize chemicals designed to have desirable fuel properties and be compatible with the current infrastructure would address many of the problems facing today's biofuels. This project demonstrates how pathways that don't exist in nature can be built, without the traditional approach of screening libraries of hundreds to thousands of enzymes, by taking advantage of the promiscuous nature of enzymes and enzyme superfamilies. Using this new approach, the project has constructed an unnatural pathway for synthesizing three different branched-C5 alcohols (3-methyl-3-butenol, 3-methyl-2-butenol, and 3-methyl-butanol) from the mevalonate pathway in *E. coli*. The three branched-C5 alcohols are promising biofuel candidates with many favorable fuel properties.

and heat shock. In this poster, we present some of these results as well as the unique challenge to engineer the cell to better cope with these stresses.

GTL

## Increasing Mevalonate Production by Engineering the Metabolism of *Escherichia Coli*

**Yisheng (Connie) Kang\*** (srhconnie@gmail.com), Hector Garcia Martin, Peter Benke, Mario Ouellet, Jennifer Gin, and Jay D. Keasling

Joint BioEnergy Institute, U.S. DOE, University of California and Lawrence Berkeley National Laboratory, Berkeley, Calif.

http://www.jbei.org

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

Synthesis the anti-malarial drug, artemisinin, precursor mevalonate in *Escherichia coli* branches from acetyl Co-A which is the entry point to TCA cycle. We have been using genetic and environmental manipulations to redirect carbon flux from the endogenous central metabolic pathways (CMPs) to the heterologous pathway precursors. Deletion the production pathways of acetate, which is an undesirable product of excessive glycolytic flux, result in excretion of pyruvate rather than help increase mevalonate production. Heterologous mevalonate pathway from *Enterococcus faecalis* is more effective to draw carbon flow than that of *Saccharomyces cerevisiae*. Providing limited amount of nitrogen source also efficiently cut carbon flux to biomass and redirect it to mevalonate production. We are also performing metabolic flux analysis using <sup>13</sup>C-labeled glucose. This information will help us determine how carbon flux through native metabolic pathways is affected by the presence of heterologous pathways, allowing us to identify and correct bottlenecks in the artemisinin production pathway. Successful completion of this project will provide insights into the metabolic status of living cells under different conditions and help us build a robust bacterium capable of producing high levels of artemisinin from cheap carbon sources.

GTL

## Transcriptomic Studies of the Response to Exogenous Exposure and Endogenous Production of Biofuel Candidates in *E. coli*

**R. Dahl<sup>1,3\*</sup>** (robdahl@berkeley.edu), B. Rutherford,<sup>1,3</sup> A. Mukhopadhyay,<sup>2,3</sup> and J.D. Keasling<sup>1,2,3,4\*</sup>

<sup>1</sup>Dept. of Chemical Engineering, University of California, Berkeley, Calif.; <sup>2</sup>Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; <sup>3</sup>Joint BioEnergy Institute, Emeryville, Calif.; and <sup>4</sup>Dept. of Bioengineering, University of California, Berkeley, Calif.

http://www.jbei.org

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

The production of biofuels in microbial systems presents a unique challenge to the host cell. Not only is the cell exposed to the solventogenic fuel molecule itself, the coordinated overexpression of an exogenous pathway presents a large burden on the cell's physiology, both in depleting nutrients and introducing foreign intermediates which could have toxic side-effects. At JBEI, we have focused on the impacts of short-chain alcohols (e.g. butanol) and isoprenoid-based fuels (e.g. isopentenol). Microarrays (in complementary studies with other system-wide "Omics" studies) were used to characterize *E. coli*'s response to these challenges. The general response has been one of a combination of oxidative, hyperosmotic (chaotropic/ desiccation)



GTL

## Microbial Production of Isoprenoid Biodiesel

**Taek Soon Lee\*** (tslee@lbl.gov), Nilu Prasad, Rossana Chan, Clem Fortman, and Jay Keasling

Joint BioEnergy Institute, Physical Bioscience Division,  
Lawrence Berkeley National Laboratory, Berkeley, Calif.

<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

Biodiesel, composed primarily of fatty acid methyl esters (FAMEs), has been one of the most successful renewable biofuels. They are derived from vegetable oil or animal fat. Limited supplies of feedstock, as well as competition for arable land, have been major drawbacks to expanding the use of biodiesel as alternative fuel. Microbial systems are used extensively for industrial small molecule production. Current techniques could be used to facilitate the production of biodiesel-like molecules in a microbial system. Furthermore, the agricultural industry generates a large and currently untapped supply of plant biomass (lignocellulose) that could be utilized as a renewable source of feedstock for these systems, resulting in carbon-neutral transportation fuels. In an effort to harness this potential, we have designed a new class of biodiesel fuels based on isoprenoids. We have chemically synthesized and tested a number of isoprenoid-based biodiesel candidates to determine which compounds are compatible with current engine technology. Biosynthetic pathways for the production of these validated isoprenoid-based fuel candidates have been designed in *E. coli*. Some of the pathways are constructed and tested for the production of target biofuels. The optimization of the production and the construction of more pathways genes are currently under investigation.

GTL

## Harnessing Genomic Recombination to Improve Microbial Metabolic Phenotypes

**Adrienne E. McKee\*** (AEMcKee@lbl.gov), John Haliburton, Veronica Fok, Mario Ouellette, Jay D. Keasling, and Swapnil Chhabra

Joint BioEnergy Institute, Physical Biosciences Division,  
Lawrence Berkeley National Laboratory, Berkeley, Calif.

<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in**

**plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

The microbial production of energy, pharmaceutical, and industrial compounds is a growing alternative to traditional, often costly, production processes. Many naturally occurring metabolic pathways of *Escherichia coli* and *Saccharomyces cerevisiae* have been enhanced for increased production of desired compounds. However, despite numerous advances, optimization of metabolic phenotypes faces many challenges as pathway improvement often requires both the redirection of intermediates and re-establishment of gene regulation – frequently requiring the modulation of many genes simultaneously. Furthermore, predicting the complement of genes that function cohesively for an organism to achieve a chosen metabolic phenotype may be exceedingly difficult, particularly if those gene products act at a distance from the pathway enzymes themselves. Genome shuffling (GS), a recently introduced strain improvement strategy, addresses these challenges through the use of genomic recombination to increase the genetic diversity of a population. When coupled with phenotypic screening and genome sequencing, GS holds the potential to discover genetic alterations that improve a phenotype as well as establish connections between gene products that may not otherwise be intuited from our current understanding of gene function or metabolic networks. Here, we present our recent efforts to develop protocols for protoplast fusion and genome shuffling in the industrial organisms *E. coli* and *S. cerevisiae*. We also discuss our current shuffling-based screens and selections that test the feasibility and effectiveness of GS in a directed application, namely increased production of the carotenoid 4,4'-diaplycopenone, a product of the isoprenoid pathway. Through deep sequencing and comparative genomics, we will assess the new genotypes of strains that arise from this approach. As all isoprenoids share common metabolic precursors, the strains and genomic knowledge generated through this research may be applicable to the biosynthesis of a wide number of valuable industrial, pharmaceutical, and energy-related compounds. The results of this study will deepen our understanding of metabolic networks and will increase our knowledge of the diverse genomic landscapes that may converge on a select phenotype.

GTL

## Optimizing Isoprenoid Biosynthesis

**Farnaz Nowroozi\*** (FNowroozi@lbl.gov), David E. Garcia, Edward E. Baidoo, and Jay D. Keasling

Dept. of Bioengineering, University of California,  
Berkeley, Calif.

<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in**

**plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

Isoprenoids are a diverse class of natural compounds that hold great commercial potential. The mevalonate pathway is one of the major pathways for isoprenoids biosynthesis. In the keasling lab, mevalonate pathway has been divided into two operons: the top pathway, and the lower pathway. We have shown that the lower mevalonate pathway is the limiting part of the mevalonate pathway and needs further optimization. We have shown that mevalonate kinase which is the first gene of the lower pathway is limiting. We are studying the kinetics of mevalonate kinase to determine the cause of this limitation. Also we are trying different ribosome binding sites with different strengths in front of the genes in the lower mevalonate pathway to optimize protein expression and therefore reach optimal metabolite production. For this purpose, it is desired to be able to measure pathway intermediates to determine which steps of the pathway are limiting, causing the intermediate metabolites to accumulate. We are trying to understand how the pathway expression is regulated when certain toxic intermediates accumulate.

We hope that this work would help us in further optimizing the isoprenoids biosynthetic pathway and constructing new pathways.

isozymes from a number of different organisms (*S. cerevisiae*, *Escherichia coli*, *Clostridium beijerinckii*, and *Ralstonia eutropha*) were substituted for the Clostridial enzymes and their effect on n-butanol production was compared. By choosing the appropriate isozymes, we were able to improve production of n-butanol ten-fold to 2.5 mg/L. The most productive strains harbored the *C. beijerinckii* 3-hydroxybutyryl-CoA dehydrogenase, which uses NADH as a co-factor, rather than the *R. eutropha* isozyme, which uses NADPH, and the acetoacetyl-CoA transferase from *S. cerevisiae* or *E. coli* rather than that from *R. eutropha*. Surprisingly, expression of the genes encoding the butyryl-CoA dehydrogenase from *C. beijerinckii* (*bcd* and *etfAB*) did not improve butanol production significantly as previously reported in *E. coli*. Using metabolite analysis, we were able to determine which steps in the n-butanol biosynthetic pathway were the most problematic and ripe for future improvement.

GTL

## Omics Research at the Joint BioEnergy Institute (JBEI)

**Paul D. Adams**, Tanveer S. Batth, Peter I. Benke, Rob Dahl, Jay D. Keasling, Aindrila Mukhopadhyay\* (amukhopadhyay@lbl.gov), Mario Ouellet, Alyssa M. Redding, Becky J. Rutherford, Chris J. Petzold, and Pramila Tamrakar

Joint BioEnergy Institute, Emeryville, Calif.

<http://www.jbei.org>

## Metabolic Engineering of *Saccharomyces cerevisiae* for the Production of n-Butanol

**Eric J. Steen**<sup>1,2\*</sup> (EJSteen@lbl.gov), Rossana Chan,<sup>1,3</sup> Nilu Prasad,<sup>1,3</sup> Samuel Myers,<sup>1,3</sup> Christopher J. Petzold,<sup>1,3</sup> Alyssa Redding,<sup>1,3</sup> Mario Ouellet,<sup>1,3</sup> and Jay D. Keasling<sup>1,2,3,4</sup>

<sup>1</sup>Joint BioEnergy Institute, Emeryville, Calif.; <sup>2</sup>Dept. of Bioengineering, University of California, Berkeley, Calif.; <sup>3</sup>Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; and <sup>4</sup>Dept. of Chemical Engineering, University of California, Berkeley, Calif.

<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

**Background** - Increasing energy costs and environmental concerns have motivated engineering microbes for the production of “second generation” biofuels that have better properties than ethanol.

**Results & Conclusions** - *Saccharomyces cerevisiae* was engineered with an n-butanol biosynthetic pathway, in which

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

Omics research at JBEI is part of the Technologies division and provides enabling tools for a variety of cell wide and analytical measurements required for a Biofuels research program. For the conversion of sugars derived from deconstructed lingo-cellulosic biomass to fuel compounds, an important area of research is on engineered organisms that contain combinations of native and non-native biochemical pathways for the production of a target metabolite and also efforts to understand the causes of toxicity/stress during such applications. For example, the incorporation of exogenous biochemical pathways into a host organism places unregulated strain on the cell by consuming metabolites, energy and critical cofactors creating an imbalance that will trigger a variety of stress response systems. While these stress responses may be useful to the cell in such an environment, they are unfavorable in an engineering context and reduce product yield or viability during production culturing. Systems biology, built on the foundations of omics studies (genomics, proteomics, metabolomics and fluxomics), enables a comprehensive view of the impact of an exogenous pathway on the host within the context

of its full metabolism. To match the requirement of such high-throughput profiling of particular classes of cellular components, genes, proteins and metabolites, our capabilities now include microarray analysis and high resolution mass spectrometry (combined with LC, GC, and CE for shotgun proteomics, targeted protein studies, primary and secondary metabolite analysis and glycomics). In collaboration with the computational core we also have powerful data analysis and integration tools. These functional genomics workflows are now being applied to gather data for the effect of (1) exposure of deconstruction conditions (Ionic liquids, post saccharification mix from simple and complex cellulose sources); (2) accumulation of endogenous and exogenous target metabolites; and (3) impact of different growth condition and expression of the biosynthetic pathway proteins in our model host microbes (*E. coli*, *S. cerevisiae*). Specifically the use of targeted proteomics using the MRM workflow has proved valuable in gauging the presence of a complete engineered pathway. Finally, the functional genomics and analytical tools described above are also being extensively used by the Feedstock division for a molecular characterization of cell wall and by the Deconstruction division in meta-genomic studies.

GTL

## JBEI Computational Biology Core

**Dylan Chivian**<sup>1,2,3\*</sup> (DCChivian@lbl.gov), Paramvir S. Dehal,<sup>2,3</sup> Marcin P. Joachimiak,<sup>2,3</sup> Keith Keller,<sup>2,4</sup> Morgan N. Price,<sup>2,3</sup> Jason K. Baumohl,<sup>2,3</sup> John T. Bates,<sup>1,3</sup> and Adam P. Arkin<sup>1,2,3,4</sup>

<sup>1</sup>DOE Joint BioEnergy Institute, Emeryville, Calif.;

<sup>2</sup>Virtual Institute for Microbial Stress and Survival, Berkeley, Calif.; <sup>3</sup>Lawrence Berkeley National Laboratory, Berkeley, Calif.; and <sup>4</sup>University of California, Berkeley, Calif.

http://www.jbei.org

**Project Goals:** JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).

**Background:** The Computational Biology Core group in the Technology Division of the Joint BioEnergy Institute (JBEI) is responsible for data integration and comparative, evolutionary, and functional genomic analysis for the purpose of enabling metabolic engineering for biofuel production. Leveraging the VIMSS MicrobesOnline website and database (http://www.microbesonline.org) for comparative and evolutionary genomics and analysis of microarray, proteomic, and metabolomic data sets, we are extending and integrating these capabilities to allow for pursuit of questions specific to biofuels challenges. These new tools will support the research of the Plant Feedstocks Division, the Deconstruction Division, and the Biofuels Synthesis

Division of JBEI, and will be made available to the wider research community through the MicrobesOnline website.

**Data Integration:** The tools developed by the Computational Biology Core will interface with the functional data captured by a laboratory information management system (LIMS), including datasets from biomass production, growth curves, protein structures and imaging data, mass spec data, phenotype microarray data, regulatory motif and ChIP-chip data, transcriptomic, proteomic, metabolomic, metabolic flux, and genomic sequence data, as well as links to literature and other external databases. We are extending the abilities of MicrobesOnline to allow for integrated analyses of even more of the above data types, as such combined analyses of orthogonal data types allows for more complex questions to be asked and yields more robust results.

**The MicrobesOnline Database:** We have extended the MicrobesOnline database to include eukaryotic microbes that may be useful for understanding the process of biologically-mediated degradation of plant cell walls. Building a more complete picture of the enzymes nature employs for breaking down plant biomass is essential for developing industrial processes for biofuel production, and as such we are developing computational tools for metagenomics to permit searching for and comparative analysis of such enzymes in natural microbial communities. We are also working to combine computational structural biology with evolutionary analysis to grasp the mechanistic details of such “deconstruction” enzymes. This will permit prediction and engineering of novel enzymes with enhanced activities and custom specificities with an eye toward building a library of parts for metabolic engineering. Our efforts also include the study of the regulation of the expression and activity of cellulases and other deconstruction enzymes. Finally, we are extending the visualization and analysis tools in MicrobesOnline to provide a pathway-based view of systems to permit integrated analysis of systems biology data to facilitate metabolic engineering.

GTL

## High Throughput Technologies to Break the Biological Barriers to Cellulosic Fuels

**Masood Hadi**<sup>1,3\*</sup> (MZhadi@sandia.gov), H. Tran,<sup>1,3</sup> S. Morrison,<sup>1,3</sup> H.C. Chu,<sup>1,3</sup> A. DeGiovanni,<sup>2,3</sup> and P. Adams<sup>2,3</sup>

<sup>1</sup>Sandia National Laboratories, Livermore, Calif.;

<sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, Calif.; and <sup>3</sup>Joint BioEnergy Institute, Emeryville, Calif.

http://www.jbei.org

**Project Goals:** JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).

A recent paradigm shift in biological science research is from characterizing single genes or proteins (over an investigator's career) to studying whole genomes, proteomes or pathways in single "experiments" in a few months. It has been known for some time that while it is quite straightforward to clone and characterize a single gene. However, it is a completely different matter to perform detailed functional and structural studies in parallel for a few hundred genes, metabolic pathway or whole genome.

The ability to produce proteins is currently a major biological, physical, and computational challenge in protein research. Given a standard set of conditions, less than 30% of any given genome is expressible in a recombinant host. Protein expression requires complex, lengthy procedures, and specific proteins commonly require individual strategies for optimal expression. Standard bench-level procedures for protein production (expression and purification) do not exist. This lack of validated processes leads to a lengthy search for correct vector, host, expression and purification conditions to yield protein in milligram amounts. This problem is further compounded during metabolic engineering experiments where not only proteins have to be expressed, in addition genetic and regulatory processes have to be optimized for successful production of a product.

To this end the technologies division has been developing comprehensive suite of technologies in a consolidated facility. These successful methods and workflows are aimed at directed cloning methods for generating large numbers of expression constructs for protein expression and purification, screening of libraries for enzyme engineering and metabolic engineering. These processes will improve technical performance, productivity and reduce costs to allow affordability and timely progress towards our goals. We discuss some of these process development efforts and present initial results.

Mass spectrometry's ability to efficiently generate intact biomolecular ions in the gas phase has led to a wide range of biological applications and is recently being applied for global metabolite profiling ('metabolomics') primarily through liquid chromatography coupled to electrospray mass spectrometry. However the complexity and relatively low throughput of this approach has limited application for high throughput enzymatic assays. To overcome this, we have developed the Nanostructure-Initiator Mass Spectrometry enzymatic (Nimzyme1) assay where enzyme substrates are immobilized on the mass spectrometry surface using fluorophilic phase interactions. This 'soft' immobilization allows efficient desorption/ionization while also allowing surface washing to reduce signal suppression from complex biological samples as a result of the preferential retention of the tagged products and reactants. We have also shown that Nimzyme can detect multiple and competing enzymatic activities and screen for optimal pH, temperature, and enzyme inhibition from crude cell lysates and a hot springs microbial community. This approach is being implemented at the DOE Joint BioEnergy Institute for high throughput functional characterization of both enzyme libraries and environmental samples. Specifically, we are constructing a complete set of glucose polysaccharides (cellobiose to cellichexose) for screening glucosyltransferase and glucotransferase activities and a p-coumaric alcohol substrate for characterization of laccase activity. Together these assays will help to identify and optimize the conversion of lignocellulose into biofuels.

#### Reference

1. Northen, Trent R, Lee J-C, Hoang L., Raymond J., Hwang D-R, Yannone S.M., Wong C-H, Siuzdak G. A Nanostructure-Initiator Mass Spectrometry Based Enzyme Activity Assay, *PNAS*, (2008) 105(10), 3678-3683.

GTL

GTL

## High Throughput Mass Spectrometry Based Enzymatic Assays for Biofuels Development

Trent R. Northen<sup>1\*</sup> (trnorthen@lbl.gov), Jinq-Chyi Lee,<sup>3</sup> Chris Petzold,<sup>2</sup> Blake Simmons,<sup>2</sup> Paul Adams,<sup>2</sup> and Anup Singh<sup>2</sup>

<sup>1</sup>Life Sciences Division and <sup>2</sup>Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, Calif.; and <sup>3</sup>Division of Biotechnology and Pharmaceutical Research, National Health Sciences Institute, Taipei, Taiwan

<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

## Multi-Mode Spectroscopic High Throughput Screening (HTS) of Phenols and Monolignols

Komandoor Achyuthan\* (kachyut@sandia.gov), Blake Simmons, Paul Adams, and Anup Singh

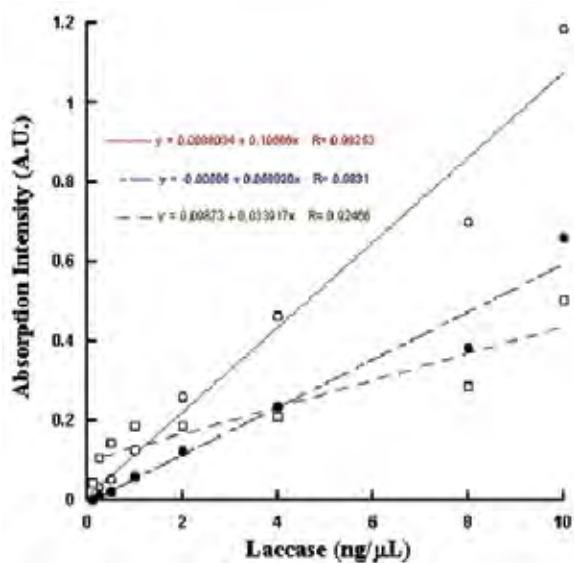
Joint BioEnergy Institute (JBEI), Emeryville, Calif.  
<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

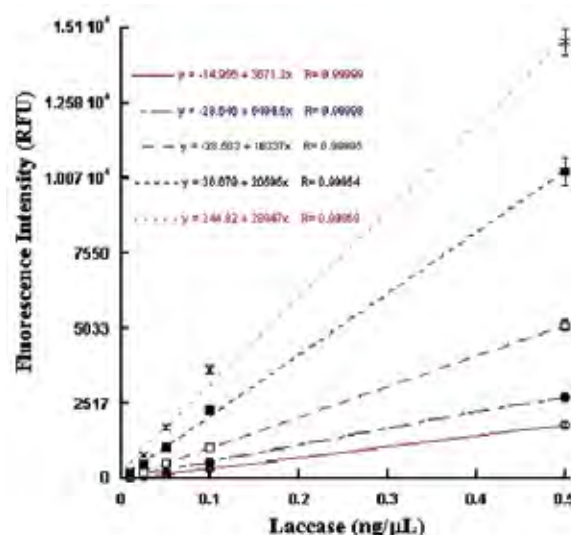
Phenolic structure is a common motif among the monolignols, coniferyl alcohol, sinapyl alcohol and p-coumaric alcohol. We are developing sensitive, rapid, multiplexed and high throughput screening (HTS)-compatible UV-Vis and fluorescence spectroscopic assays for phenols and

monolignols within 96- and 384-well microplates. We used p-cresol as a model for the phenols and coniferyl alcohol as a prototype monolignol. We employed the fungal (*Trametes versicolor*) laccase enzyme to oxidize p-cresol and coniferyl alcohol, and thereby expanded the spectroscopic properties of these molecules. Laccases are involved in lignin degradation. Our choice of laccase is especially relevant, since the enzyme was purified from white rot basidiomycetes that are efficient degraders of lignins and widely studied for biofuels.

We supply a menu of spectroscopic options for the HTS of laccase oxidation of p-cresol through multiple modes of detection. Laccase activity was monitored kinetically at pH 4.5 by absorption changes at 250nm, 274nm or 297 nm, and in endpoint mode by the bathochromic shift in absorption to 326nm. Laccase oxidation of p-cresol was also detected by product fluorescence at 425nm after excitation at 262nm or 322nm. We optimized the kinetic parameters for p-cresol oxidation (pH optimum 4.5-5.1; 37°C;  $K_m = 2.2\text{mM}$ ) resulting in laccase limits of detection and quantization (LOD, LOQ) of 25pg/ $\mu\text{L}$  and 75pg/ $\mu\text{L}$ , respectively (~360pM; 25ppb). The p-cresol LOD was 8 $\mu\text{M}$  with a potential for further improvements in sensitivity. A key advantage of our assay is that laccase catalysis could be interrogated using multi-mode spectroscopy under acidic or basic conditions, in real time or endpoint modes.

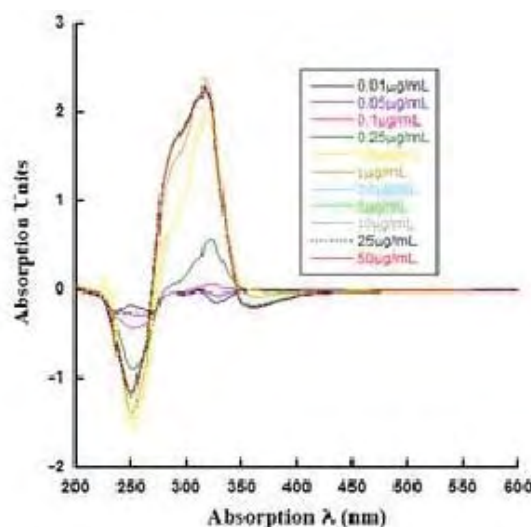


**Figure 1.** Absorption changes of laccase-catalyzed p-cresol oxidation. Laccase oxidation of 1mM p-cresol was in pH 4.5 buffer for 10 min. Absorption changes at 250nm (open circles), 297nm (closed circles) and 326nm (open squares) were measured as described above.

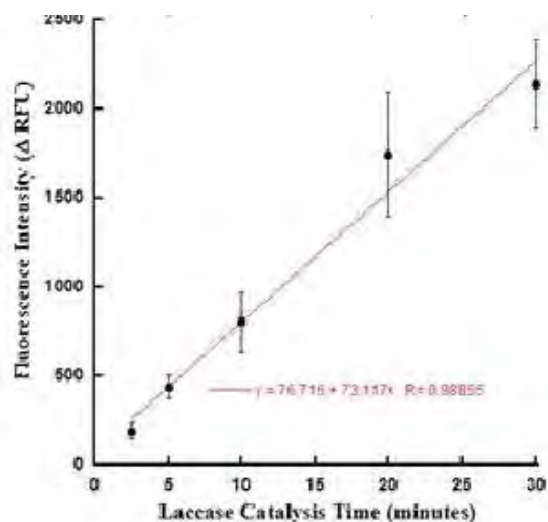


**Figure 2.** Laccase dose-response curves from 425nm fluorescence emission (excitation = 322nm). Reaction conditions were as described for Figure-1. Fluorescence was monitored for 2.5 min. (open circles), 5 min. (closed circles), 10 min. (open squares), 20 min. (closed squares) and 30 min. (crosses). All reactions were linear ( $r^2 > 0.99$ ).

We similarly characterized the spectroscopic properties of coniferyl alcohol in seven different solvents. Three isosbestic wavelengths were identified at 240nm, 242nm and 262nm between NaOH and the six solvents. A S/B of ~50 with 500 $\mu\text{M}$  coniferyl alcohol indicated assay sensitivity. The excitation spectrum was broad (270 – 335nm) and overlapped with absorption spectrum, as expected. Fluorescence emission was between 360 – 500nm with peak at 416 – 420nm. Fluorescence spectroscopy gave 1 $\mu\text{M}$  of coniferyl alcohol detection sensitivity. Unlike p-cresol, a fluorescence quench was observed following laccase oxidation of coniferyl alcohol.



**Figure 3.** UV-Vis Difference Spectra of laccase-catalyzed oxidation of Coniferyl alcohol. Increasing concentrations of laccase were reacted in pH 4.5 assay buffer for 60 minutes with 1mM Coniferyl alcohol. The absorption changes taking place over the wavelength of 200 – 600nm are shown along with the concentrations of laccase used as inset.



**Figure 4.** Laccase reaction kinetics from 416nm fluorescence emission (excitation =  $\lambda$ 310 or 320nm). Reaction conditions were as described for Figure-1. Laccase activity was detected rapidly, in as little as 2.5 minutes towards the Coniferyl alcohol substrate.

In conclusion, we demonstrated sensitive, rapid, HTS-compatible fluorescence “turn on” and “turn off” spectroscopic assays for phenols and monolignols. Orthogonal interrogation and ratiometric analysis are key features of our assay, enabling high specificity and minimizing interferences during compound library screening. A portion of this work has been accepted for publication (below). We plan to expand our investigations to include the remaining two monolignols: sinapyl and p-coumaryl alcohols.

*Publication:* K. E. Achyuthan, et al, Comb. Chem. HTS, 2009 (Accepted; In Press).

*Acknowledgements:*

This work was funded by the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy. Sandia is a multi-program laboratory operated by Sandia Corporation, a Lockheed Martin Company for the United States Department of Energy's National Nuclear Security Administration under contract # DE-AC04-94AL85000.

GTL

## Microfluidics for Protein Expression, Purification, and Screening

**Rajiv Bharadwaj\*** ([rbharadwaj@lbl.gov](mailto:rbharadwaj@lbl.gov)), Jim Brennan, and Anup K. Singh

Technology Division, Joint BioEnergy Institute, Emeryville, Calif.

<http://www.jbei.org>

**Project Goals:** JBEI's primary scientific mission is to advance the development of the next generation of biofu-

els – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).

High-throughput (HTP) technologies for protein expression and screening are critical to the rapid development of cost-effective bioconversion processes. For example, methods are required for functional analysis of tens of thousands of wild-type and engineered lignocellulose degrading enzymes, as well as for full characterization of the plant cell wall synthesis machinery with a throughput of thousands of gene clones per year. We are developing novel integrated microfluidic platforms for small-scale (sub- $\mu$ g levels) high-throughput protein expression, purification, and screening. The promising enzyme candidates can thus be rapidly identified and this paves the path for scaled-up production of enzymes.

Our microfluidic platform seeks to integrate fractionation of cell lysate, purification, characterization, and screening. A key to facile integration of these steps, which often features incompatibility of scale and modes of operation, is the replacement of centrifugation with liquid-liquid extraction for fractionation of cell lysates. We are also developing microfluidic cell-free-protein synthesis technology for simplified HTP protein expression. In addition, we are integrating on-chip protein characterization steps such as reversed-phase chromatography, SDS-PAGE, and native PAGE. Finally, we are developing colorimetric and fluorescence-based enzyme assays for screening lignocellulases and glycosyltransferases.

GTL

## Electron Microscopic Imaging at JBEI

**Bernhard Knierim\*** ([BKnierim@lbl.gov](mailto:BKnierim@lbl.gov)), Lina Prak, Seema Singh, Danielle Jorgens, Marcin Zemla, Kristen DeAngelis, Amitha Reddy, Jean Vanderghenst, Terry C. Hazen, Brad Holmes, Rajat Sapra, Blake Simmons, Paul Adams, and Manfred Auer

Joint BioEnergy Institute, Emeryville, Calif.

<http://www.jbei.org>

**Project Goals:** JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).

Advanced imaging approaches at JBEI include sophisticated sample preparation such as high-pressure freezing/freeze substitution, resin-embedded transmission electron microscopy (TEM), cryo-EM and electron tomography, as well as scanning electron microscopy (SEM). In addition we have advanced optical imaging techniques including confocal fluorescence microscopy, Raman microscopy, as well as atomic force microscopy. These techniques are applied to

characterize feedstock cell walls, to determine subcellular protein localization in feedstocks, to monitor at high resolution the consequences on cell wall properties of ionic liquid pretreatment of biomass, to visualize microbial communities, and to analyze in detail the macromolecular lignocellulose degradation strategies of selected candidate microbes.

Specifically, we have examined the cell wall of *Brachypodium*, *Miscanthus*, *Equisetum* and found in 2D TEM projection views differences in density texture, most likely reflecting real differences in cell wall architecture.

Second, we have characterized the effect of pretreatment on cell walls as a function of length of exposure of ionic liquid pretreatment on Switchgrass biomass using high-resolution wide-field electron microscopy. We are planning to subject specimen from selected time points to electron tomographic 3D analysis.

Third, we have studied Puerto Rico rain forest and compost microbial communities and found an abundance of bacterial shapes and sizes, as well a variety of interesting extracellular features likely to be involved in lignocellulose degradation. Our images will complement phylogenetic profiling of such samples and may allow a spatial mapping of the respective position of community members and their interaction.

Fourth, we have studied *Sulfolobus* samples incubated in the presence of a variety of different substrates. Only in the presence of cellulose did we find an organelle-like feature that appears to be a novel membrane-bound cellulose-degrading extracellular specialization. We are currently in the process of characterizing this novel feature in more detail.

—  
GTL

## An MRM-based Mass Spectrometry Method for Optimization of Protein Expression to Increase Biofuel Production

**Christopher J. Petzold**<sup>1\*</sup> (CJPetzold@lbl.gov), Alyssa M. Redding,<sup>1</sup> Sung Kuk Lee,<sup>1</sup> Farnaz F. Nowroozi,<sup>2</sup> Tanveer S. Batth,<sup>1</sup> Aindrila Mukhopadhyay,<sup>1</sup> Paul D. Adams,<sup>1</sup> and Jay D. Keasling<sup>1,2</sup>

<sup>1</sup>Joint BioEnergy Institute, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; and <sup>2</sup>Dept. of Chemical Engineering, University of California, Berkeley, Calif.

<http://www.jbei.org>

**Project Goals: JBEI's primary scientific mission is to advance the development of the next generation of biofuels – liquid fuels derived from the solar energy stored in plant biomass. JBEI is one of three new U.S. Department of Energy (DOE) Bioenergy Research Centers (BRCs).**

Optimization of biofuel pathway protein levels is crucial to balancing the energetic and carbon utilization of a microbe for efficient biofuel production. However, identification and quantification of specific proteins in complex mixtures is a difficult task. Since the physical attributes of proteins (e.g., MW, pI) are quite similar extensive separation or high specificity are needed to correctly identify a particular protein from a cell lysate. Western blots simplify analyses due to their high selectivity towards the target protein and tagging the protein of interest offer a means by which selective enrichment is possible. Yet, Western blots and tagging have limitations that make alternate methods attractive. One method, multiple-reaction monitoring, is capable of rapidly changing the target protein, something not possible without an antibody for the new protein, and detecting multiple target proteins in the same sample, something not possible without multiple tagging strategies and different enrichment steps. Multiple-reaction monitoring (MRM) is a mass spectrometric technique that has been used for small molecule DMPK studies for many years and has recently been adapted to peptides. Coupled to liquid chromatography, MRM-based analysis offers high selectivity and sensitivity. This method utilizes two points of selection (a peptide mass and a specific fragment mass generated by MS/MS) to eliminate background signal and noise even in very complex mixtures. Since the entire mass range is not scanned and only specific MRM transitions (combinations of peptide and fragment masses) are detected a significant increase in sensitivity is typically observed. Careful selection and optimization of MRM transitions permits detection of 5-10 specific proteins per LC-MS analysis. We are currently developing and optimizing MRM transitions to target proteins of interest for producing high titers of biofuel molecules. Our initial efforts are directed at optimizing the mevalonate pathway, the foundation for producing isoprenoid-based biofuels. The mevalonate pathway also serves as a good model for butanol-producing microbes. With these methods we will characterize a variety of protein expression parameters (promoter; ribosome binding site; plasmid) to determine optimal metabolite (e.g., mevalonate, isoprenoids, butanol) production.

## Great Lakes Bioenergy Research Center (GLBRC)

GTL

## Streamlined Method for Biomass Whole-Cell-Wall Structural Profiling

John Ralph<sup>1\*</sup> (jralph@wisc.edu), Fachuang Lu,<sup>2</sup> Hoon Kim,<sup>2</sup> Xiaowen Fang,<sup>2</sup> Mark Norton,<sup>2</sup> James J. Ellinger,<sup>2</sup> John L. Markley,<sup>2</sup> Daniel Yelle,<sup>3</sup> Kenneth E. Hammel,<sup>4</sup> Sally A. Ralph,<sup>4</sup> Ronald D. Hatfield,<sup>5</sup> Paul F. Schatz,<sup>5</sup> Mattias Hedenström,<sup>6</sup> and Björn Sundberg<sup>7</sup>

<sup>1</sup>Dept. of Biochemistry (Enzyme Institute) and Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, Wis.; <sup>2</sup>Dept. of Biochemistry, University of Wisconsin, Madison, Wis.; <sup>3</sup>USDA Forest Products Laboratory and Dept. of Forestry, University of Wisconsin, Madison, Wis.; <sup>4</sup>USDA Forest Products Laboratory, Madison, Wis.; <sup>5</sup>U.S. Dairy Forage Research Center, Madison, Wis.; <sup>6</sup>Dept. of Chemistry and <sup>7</sup>Umeå Plant Science Centre, Umeå University, Umeå, Sweden  
<http://www.greatlakesbioenergy.org>

**Project Goals: The aim is to provide the plant cell wall and biomass research communities with improved methods for polysaccharide and lignin structural profiling, based on complete cell wall solubilization and NMR. The intent was to develop and streamline procedures to allow 20-30 samples per day to be profiled.**

## Introduction

In wide-ranging research aimed at altering plant cell wall characteristics, by conventional breeding or modern genetic methods, one of the biggest problems is in delineating the effects on the cell wall. Plant cell walls are a complex conglomerate of a variety of polysaccharides and lignin. Although other methods have their place, and can be more rapid (e.g. NIR), the difficulty in interpretation of some spectral methods, or the destruction of structure by chemical methods, assures that key features of cell walls benefiting, for example, biomass production and conversion are lost. A promising recent approach is the dissolution of the whole cell wall and high-resolution solution-state NMR analysis. These methods are providing promising approaches to detailing the compositional and (chemical) structural characteristics of the complex cell wall polymers that are the basis of biomass conversion efforts. A coordination of the developments from cell wall dissolution, high-resolution 2D NMR profiling, and chemometrics approaches to relate the NMR profile to other properties, seems poised to allow medium-throughput profiling of the wall, an area in which there are no currently comparable methods. The dissolution, NMR, and associated chemometrics methods are anticipated to impact research on optimal feedstock selection, aid in predicting biomass conversion efficiencies, and help process optimization.

## Objectives

The aim was to provide the plant cell wall and biomass research communities with improved methods for polysaccharide and lignin structural profiling, based on complete cell wall solubilization and NMR. The intent was to develop and streamline procedures to allow 20-30 samples per day to be profiled.

## Background to CW-dissolution/NMR/Chemometrics, and Progress

Traditionally, structural analysis of cell wall polymers has been via destructive analytical methods or via NMR on fractionated isolates. The problem with the latter is that isolation of components from the complex and interconnected wall polymers delivers them altered, in only low yields, and partitioned in ill-defined ways. Some years ago, Fachuang Lu developed two solvent systems that allowed full dissolution of finely-ground cell walls.<sup>1</sup> Acetylation of cell wall material dissolved by one of the solvents (DMSO-NMI) delivered acetylated walls that were soluble in common (organic) NMR solvents. The dispersive power of 2D (and even 3D<sup>2</sup>) NMR produced the most comprehensive structural profiling of all the polymers in the cell wall, without fractionation. Already it is finding utility in the elucidation of changes in polysaccharides and lignins in transgenic plants,<sup>3-5</sup> is overturning accepted theories on fungal rotting mechanisms [even extensively degraded material is soluble in these solvents],<sup>6</sup> has recently revealed new cell wall cross-linking mechanisms (Kim unpublished), and is beginning to be applied to other recalcitrant polymers.

In a logical development, Daniel Yelle prepared perdeutero-NMI allowing CW dissolution directly into DMSO-d<sub>6</sub>/NMI-d<sub>6</sub> right in the NMR tube, without requiring the tedious isolation steps following acetylation.<sup>7</sup> Some structures resolved better, some worse, than when acetylated. A major advantage was that natural lignin and polysaccharide acetylation could be detected – acetylation is typically a detriment to efficient saccharification and therefore needs to be assessed. The drawback is that NMI-d<sub>6</sub> is expensive.

In attempts to find cheaper and simpler systems, Hoon Kim discovered that poorer quality but nevertheless impressive and informative spectra could be acquired by simply swelling the ball-milled cell walls in the readily accessible NMR solvent, DMSO-d<sub>6</sub>.<sup>8</sup> More recently (Kim, manuscript in preparation), the addition of commercially available pyridine-d<sub>5</sub> was shown to improve the quality of the spectra to near DMSO-d<sub>6</sub>/NMI-d<sub>6</sub> levels. Additionally, with selection of superior 2D <sup>13</sup>C-<sup>1</sup>H correlation NMR pulse sequences and optimization of parameters, NMR throughput at the rate of 20-50 samples per day (compared to the original rate of one sample per day!) is now attainable;<sup>8</sup> longer acquisition times continue to result in superior spectra with better detail for minor components, so will still be required for some studies. With the use of modern cryogenically-cooled NMR probes, whole-cell-wall 2D NMR spectra can readily be acquired with just 10-30 mg of material.

Beyond the utility of these methods to analyze cell wall structure (identifying the effects of gene manipulation, for example) is the potential to utilize the 2D NMR cell wall



profile itself directly in multivariate analysis. As this had not previously been accomplished with 2D NMR data of this type, we collaborated with colleagues at Umeå University (Mattias Hedenström, Björn Sundberg) to develop the methods necessary to convert the data and apply a range of chemometrics methods to them. Initial test studies using tension wood vs normal wood (poplar) and pectin methyl-esterase misregulated transgenics are beginning to attest to the power of multivariate methods on the NMR profile.<sup>9-11</sup>

Work continues on improving all aspects of the procedures. Although we have not yet succeeded, it is hoped to substantially improve the step that is now the bottleneck, ball-milling the cell wall, via a microfuge tube system. NMR methods continue to be improved, and a significant effort is on improved assignments, especially of the polysaccharide components. Chemometrics methods continue to evolve. The methodology is being applied to a wider array of samples. The big remaining aspect, beyond the initial objectives, is to provide automated assignments and quantification of components, and to more effectively database and search the increasingly massive NMR databases via interactions with the NMR Facility at Madison.

## References

1. F. Lu and J. Ralph. Non-degradative dissolution and acetylation of ball-milled plant cell walls; high-resolution solution-state NMR. *The Plant Journal*, 2003, **35**, 535-544.
2. J. Ralph and F. Lu. Cryoprobe 3D NMR of acetylated ball-milled pine cell walls. *Organic and Biomolecular Chemistry*, 2004, **2**, 2714-2715.
3. J. Ralph, H. Kim, F. Lu, J. H. Grabber, W. Boerjan, J.-C. Leplé, J. Berrio Sierra, M. Mir Derikvand, L. Jouanin and C. Lapierre. Identification of the structure and origin of a thioacidolysis marker compound for ferulic acid incorporation into angiosperm lignins (and an indicator for cinnamoyl-CoA reductase deficiency). *The Plant Journal*, 2008, **53**, 368-379.
4. A. Wagner, J. Ralph, T. Akiyama, H. Flint, L. Phillips, K. M. Torr, B. Nanayakkara and L. Te Kiri. Modifying lignin in conifers: The role of HCT during tracheary element formation in *Pinus radiata* *Proceedings of the National Academy of Sciences, USA*, 2007, **104**, 11856-11861.
5. J.-C. Leplé, R. Dauwe, *et al.* Downregulation of cinnamoyl coenzyme A reductase in poplar; multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure. *Plant Cell*, 2007, **19**, 3669-3691.
6. D. J. Yelle, J. Ralph, F. Lu and K. E. Hammel. Evidence for cleavage of lignin by a brown rot basidiomycete. *Environmental Microbiology*, 2008, **10**, 1844-1849.
7. D. J. Yelle, J. Ralph and C. R. Frihart. Characterization of non-derivatized plant cell walls using high-resolution solution-state NMR spectroscopy. *Magnetic Resonance in Chemistry*, 2008, **46**, 508-517.
8. H. Kim, J. Ralph and T. Akiyama. Solution-state 2D NMR of Ball-milled Plant Cell Wall Gels in DMSO-d<sub>6</sub>. *BioEnergy Research*, 2008, **1**, 56-66.
9. M. Hedenström, S. Wiklund, J. Ralph, U. Edlund and B. Sundberg. A novel approach for analysis of poplar wood using two-dimensional NMR spectroscopy on dissolved whole cell walls in combination with chemometrics *Physiologia Plantarum*, 2007, **130**, XIth Cell Wall Meeting, Abstract 22.
10. M. Hedenström, S. Wiklund, F. Lu, L. Gerbner, P. F. Schatz, B. Sundberg and J. Ralph. Analysis of plant cell walls using 2D NMR spectroscopy and multivariate analysis; Application to tension wood and PME-mis-regulated plants. *Plant Journal*, 2009, in preparation.
11. J. Ralph. High-Resolution Solution-state NMR of Unfractionated Plant Cell Walls; Potential for Biomass Selection and Process Optimization. in *2008 Pittsburgh Conference*, New Orleans, LA, Paper 1730-1735.

This research was supported by the Office of Science (BER), U.S. Department of Energy, and USDA-CSREES, Interagency Agreement No. DE-AI02-06ER64299.

GTL

## Quantification of Whole Plant Cell Wall and Plant Metabolites Using Advanced 2D <sup>1</sup>H-<sup>13</sup>C HMQC NMR Techniques

Xiaowen Fang,<sup>1\*</sup> James Ellinger,<sup>1</sup> Mark Norton,<sup>1</sup> William M. Westler,<sup>1</sup> Michael Casler,<sup>2</sup> Shawn M. Kaeppler,<sup>3</sup> and John M. Markley<sup>1</sup> (markley@nmrfam.wisc.edu)

<sup>1</sup>National Magnetic Resonance Facility at Madison, Dept. of Biochemistry, University of Wisconsin, Madison, Wis.; <sup>2</sup>Dairy Forage Research Center, USDA-ARS, Madison, Wis.; and <sup>3</sup>Dept. of Agronomy, University of Wisconsin, Madison, Wis.

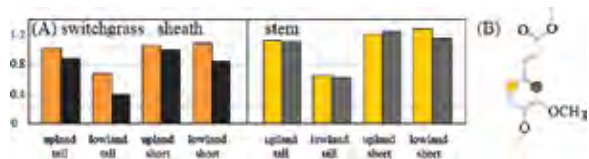
<http://www.greatlakesbioenergy.org>

<http://www.nmrfam.wisc.edu>

**Project Goals: Quantify the relative abundance of cell wall components such as specific lignin, cellulose and hemicellulose in various biomass sources.**

In the conversion of lignocellulosic plant material to liquid bioethanol, the linkages between lignin and hemicellulose/cellulose represent an important research focus, because the presence of lignin severely blocks access of cellulases to the hemicellulose/cellulose energy components. Thus it is important to be able to quantify the relative abundance of cell wall components such as ferulate in various biomass sources. Multi-dimensional solution NMR spectroscopy of unfractionated cell wall material<sup>1</sup> dispersed in organic solvents<sup>2</sup> provides a powerful approach for screening for variations in wall structure. 2D <sup>1</sup>H-<sup>13</sup>C heteronuclear single quantum correlation (HSQC) spectroscopy has been applied widely in studies of biomolecules and natural products including cell walls.<sup>1</sup> However, the internal dynamics of solubilized cell wall polymers approach the rigid limit, and their shorter spin-spin relaxation times ( $T_2$ ) and longer spin-lattice relaxation times ( $T_1$ ) reduce signal sensitivity. We demonstrate here that these problems can be reduced through the use of sensitivity enhanced heteronuclear multiple quantum correlation (HMQC) spectroscopy. The use of gradient enhanced HMQC improves the uniformity of signal intensities across the spectral width and improves spectral dynamic range. This leads to improved quantification of signals from acetyl groups, whose presence correlates

with inhibition of saccharification. The improvements result in part from the longer characteristic  $T_2$  of multiple quantum terms and the fewer number of radio frequency pulses required by the NMR experiment.<sup>3</sup> An additional benefit is that HMQC is more easily incorporated into pulse sequences designed for more detailed structural identification. In studies of switchgrass and gene mutated corn, have found that 2D  $^1\text{H}$ - $^{13}\text{C}$  HMQC data collection in combination with added internal standards allows accurate quantification of cell wall constituents. Ten samples per day can be analyzed in this way even without an automated NMR sample changer. Figure 1 illustrates the quantification of ferulate in switchgrass tissue from different species.



**Figure 1.** (A) Quantification of ferulate on the basis of 2D  $^1\text{H}$ - $^{13}\text{C}$  HMQC data in sheath and stem samples from upland tall and short and lowland of tall or short switchgrass. The orange and black bars represent data from two different CH groups of ferulate as shown in (B).

## References

1. H. Kim, J. Ralph and T. Akiyama. Solution-state 2D NMR of ball-milled plant cell wall gels in DMSO- $d_6$ . *BioEnergy Research*. 2008, **1**, 56-66.
2. D. J. Yelle, J. Ralph and C. R. Frihart. Characterization of nonderivatized plant cell walls using high-resolution solution-state nmr spectroscopy. *Magn Reson Chem*, 2008, **46**, 508-517
3. P. K. Mandal and A. Majumdar. A Comprehensive Discussion of HSQC and HMQC Pulse Sequences. *Concepts in Magnetic Resonance A*, 2004, **20A(1)** 1-23.

## Acknowledgements

This work was carried out at the Great Lakes Bioenergy Research Center with support from the U. S. Department of Energy. We thank John Ralph and Hoon Kim for valuable advice of sample preparation.

GTL

## Endogenous Variation for Biofuel Quantity and Quality Traits in Maize and Switchgrass

**Shawn Kaeppler\*** (smkaeppl@wisc.edu), Natalia de Leon (ndelegatti@wisc.edu), Mike Casler (mdcasler@wisc.edu), Heidi Kaeppler (hfkaeppl@wisc.edu), Rajan Sekhon, James Johnson, and Candy Hansey

University of Wisconsin, Madison, Wis.

<http://www.greatlakesbioenergy.org>

Ethanol potential per unit of land is a function of biomass yield and biomass quality. Maize and switchgrass have substantial phenotypic variation among diverse cultivars,

populations, and genetic collections for these traits. The goal of this research is to discover genes and alleles underlying this variation both to enhance the improvement of maize for biofuel production, as well as to use information from maize as a model for the closely related biofuel crop, switchgrass. Biomass yield and quality are being evaluated on genetic mapping populations and diverse inbred lines. In maize, we are studying nearly 6000 genotypes from 28 recombinant inbred line populations. These populations have a useful combination of balanced allele frequencies, known allelic composition, and have large phenotypic variation for biomass quantity and quality. These populations are used to identify quantitative trait loci underlying biofuel traits. To further define the specific genes and alleles underlying the quantitative trait loci, we are developing an association mapping population composed of more than 500 lines that will mature in our environment. These materials are being genotyped with anonymous SNP markers to allow us to account for population structure caused by pedigree relationships and geographic isolation. Sequence polymorphisms for candidate genes will then be scored and associated with observed phenotypic variation to identify causal genes and polymorphisms. Pathways that are identified in this process will be subject to further perturbation and analysis using molecular approaches such as proteomics and metabolomics and genetic approaches including mutagenesis and transgenics. Genes identified in maize as important in controlling phenotypic variation for biofuel traits will then be evaluated in switchgrass using similar genetic materials and approaches. The first traits that we have measured are developmental traits related to biomass quality and quantity. Approaches to best evaluate quality are under investigation and currently include NIR prediction and wet laboratory analysis using methods developed to assess digestibility in ruminants as well as using high throughput assays to measure sugars, convertibility, and direct ethanol production.

## Discovery of Genes that Mediate and Regulate Hemicellulose Biosynthesis

**Kenneth Keegstra\*** (Keegstra@msu.edu), Markus Pauly (Paulymar@msu.edu), Curtis Wilkerson (Wilker13@msu.edu), David Cavalier, Jean-Christoph Cocuron, Jacob Jensen, Nick Thrower, and Yan Wang

Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, Mich.

<http://www.greatlakesbioenergy.org>

Current knowledge regarding hemicellulose biosynthesis in plants is fragmented and incomplete. While some proteins, and the genes that encode them, have been identified, many gaps remain and little is known about how hemicellulose biosynthesis is regulated. One attractive strategy for identifying the required genes is to perform expression profiling during periods of rapid hemicellulose deposition. Many plants produce large quantities of specific polysaccharides

as storage polymers in developing seeds. Studies of these seed systems has been used to identify proteins involved in mannan and xyloglucan biosynthesis (1,2). We are extending these earlier studies by using 454 sequencing technology to perform deep EST sequencing at various stages of seed development during and just before rapid synthesis of mannan (Fenugreek-*Trigonella foenum-graecum*), xyloglucan (Nasturtium-*Tropaeolum majus*), or arabinoxylan (Psyllium-*Plantago ovata*). Analysis of the sequences obtained has confirmed the expression of genes known to be involved in the biosynthesis of these polysaccharides. In addition, a number of other genes have emerged as strong candidates for involvement in the production of these polysaccharides or in regulation of these pathways. Among the candidates that have been identified are putative sugar nucleotide biosynthetic enzymes, putative sugar nucleotide transporters, putative and known glycosyltransferases and glycan synthases of unknown specificity, proteins of unknown function, and transcription factors. Many of the genes have homologs that are expressed in developing wood or in other plant tissues where wall synthesis is occurring rapidly, providing support for the hypothesis that the same genes are involved in depositing these polymers in secondary cell walls. Promising examples from each class of candidate genes have been selected for detailed functional analysis. Selected examples from each polysaccharide will be presented on the poster.

#### References

1. Dhugga KS, Barreiro R, Whitten B, Stecca K, Hazebroek J, Randhawa GS, Dolan M, Kinney AJ, Tomes D, Nichols S, Anderson P (2004) Guar seed  $\beta$ -mannan synthase is a member of the cellulose synthase super gene family. *Science* 303: 363-366.
2. Cocuron J-C, Lerouxel O, Drakakaki G, Alonso AP, Liepman AH, Keegstra K, Raikhel N, and Wilkerson CG (2007). A gene from the cellulose synthase-like C family encodes a  $\beta$ -1,4 glucan synthase. *Proc Natl Acad Sci USA* 104: 8550-8555.

GTL

### Biomass Trait Screening in a *Brachypodium* Mutant Population

Megan Galvin,<sup>1</sup> Sonali Nath,<sup>1</sup> Colleen Stephens,<sup>1</sup> Florian Kraemer,<sup>2</sup> Markus Pauly,<sup>2</sup> John Vogel,<sup>3</sup> and **John Sedbrook**<sup>1\*</sup> (jcsedbr@ilstu.edu)

<sup>1</sup>School of Biological Sciences, Illinois State University, Normal, Ill.; <sup>2</sup>MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, Mich.; and <sup>3</sup>Western Regional Research Center, U.S. Dept. of Agriculture, Albany, Calif.

<http://www.greatlakesbioenergy.org>

The grass *Brachypodium distachyon* is emerging as an important model system for bioenergy crop grasses such as switchgrass and *Miscanthus* owing to its small genome size (~300Mbp), small stature, short generation time,

transformability, and self-fertilization. The DOE Joint Genome Initiative (JGI) has recently sequenced the entire *Brachypodium* genome and is in the process of sequencing 300,000 *Brachypodium* Expressed Sequence Tags (ESTs). While others within the Great Lakes Bioenergy Research Center (GLBRC) are surveying natural variation of biomass traits within a collection of *Brachypodium* accessions collected worldwide, our group is focusing on identifying *Brachypodium* EMS mutagenized lines with differences in either enzymatic digestibility or stem morphology/growth characteristics. To date, using an HPLC based digestibility assay, we have screened through over 1,000 EMS lines and identified over 10 putative mutants with increased release of glucose and/or xylose upon hydrolytic enzyme digestion. We will discuss these results along with more detailed GC/MS-based cell wall analyses results. Our long-term goal is to identify genes and gene isoforms that can be introduced or bred into bioenergy crop grasses and corn thereby improving biomass digestibility/fermentability as well as biomass density and yields.

GTL

### Understanding the Transcriptional Regulation of Secondary Wall Biosynthesis: A Step Toward Optimizing Lignocellulosic Feedstock for Biofuel Productivity and Processing

Jae-Heung Ko, Won-Chan Kim, Zhongnan Zhang, and **Kyung-Hwan Han**\* (hanky@msu.edu)

Dept. of Forestry, Michigan State University, East Lansing, Mich.

<http://www.greatlakesbioenergy.org>

Secondary cell wall of plant biomass is a major source of lignocellulosic material, which is the most abundant natural compound on earth and a promising source of sugars for liquid biofuel production. Better understanding of the molecular mechanisms underlying its biosynthesis will help us develop biotechnological means to genetically control key pathways that determine the quantity and quality of the biomass. In an effort to identify candidate genes involved in transcriptional regulation of secondary wall biosynthesis, we developed an inducible secondary wall thickening system in *Arabidopsis thaliana*. Using this system, a battery of differentially expressed genes was identified. Most of the secondary wall biosynthetic genes (e.g., cellulose, hemicellulose and lignin genes) were induced within 6-hr of secondary wall thickening-induction treatment. We then identified several transcriptional regulators whose expression is coincided or preceded with the induction of secondary wall biosynthetic genes. The candidate genes, including C3H, LIM, NAC, MYB, and PXY transcription factors, are being tested in transgenic plants. Based on the transcriptome analysis, we constructed a tentative hierarchical transcriptional regulatory network leading to the biosynthesis of secondary wall components. In order to confirm the relationship between

transcription factors and their target genes, we are using both transient activation assay and electrophoretic mobility shift assay (EMSA). This poster presentation will describe (1) our genomics approach for identifying a transcriptional regulatory network that control secondary wall biosynthesis and; (2) functional characterization of selected candidate genes in the network.

GTL

## Use of Proteomics Technologies for the Characterization of Proteins, Microbes and Microbial Communities Important for Bioenergy Production

Mary S. Lipton<sup>1\*</sup> (mary.lipton@pnl.gov), Stephen J. Callister,<sup>1</sup> Kim K. Hixson,<sup>2</sup> Samuel O. Purvine,<sup>2</sup> Angela D. Norbeck,<sup>1</sup> Ronald J. Moore,<sup>1</sup> Therese R. Clauss,<sup>1</sup> and Carrie D. Nicora<sup>1</sup>

<sup>1</sup>Biological Sciences Division and <sup>2</sup>Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Wash.

*GLBRC Collaborators:* Michigan State University: Shishir Chundawat, Dahai Gao, Venkatesh Balan, and Bruce Dale; University of Wisconsin, Madison: Robert Landick, Patricia Kiley, Tim Donohue, Eva Ziegelhoffer, Safak Yildaz, Daniel Noguera, Cameron R. Currie, and Jarrod J. Scott

<http://www.greatlakesbioenergy.org>

Microbial processing for substrates to biofuels, whether the conversion of ligno-cellulosic material to ethanol or microbial biorefineries to produce hydrogen or electricity is a central part of the GLBRC mission. Inherent in the use of microbes for these purposes is the characterization of the fundamental machines of the cell, the proteins, and how these proteins dictate microbial function. The quantitative determination of protein expression patterns and how these patterns change with changing cell state is critical for the GLBRC to remove bottlenecks in the bio fuels pipeline. Additionally, accurate measurements of protein levels and modifications will provide more extensive insights into both the plants and the microbes in the bioenergy pipeline. To this end, microbial researchers of the GLBRC will utilize state of the art proteomics facilities resident at the Pacific Northwest National Laboratory that allow for rapid global determination of protein expression patterns of cells or organelles. The proteomics facility at PNNL is one of these Genomics:GTL programs and as such will continue to characterize the protein complement of microbes and enzymes and is poised to make significant contributions in the characterization of microbial communities and plants.

Advances in proteomic technologies at PNNL have enabled characterization of unique features of microbial systems. Ranging from protein preparations purified from fungal and bacterial sources that demonstrate the ability to degrade

lignocellulosic material to the quantitative proteomic profiling of microbes and microbial communities, the proteomics capability at PNNL is employed at producing data to further the understanding of systems important to the GLBRC. The proteomic analyses on all systems can be expanded to include temporal profiling through the analysis of time course studies, characterization of post-translational modifications, and determination of sub cellular localization of proteins. Extension of proteomic capabilities to community profiling will enable deeper understandings of how microbes interact with each other in environmental settings.

In the past year, the proteomics facility has supported the GLBRC in four aspects. The characterization of cellulolytic and hemicellulolytic rich enzyme cocktails found that the dominant cellulases were CBH I, Xyloglucanase, CBH II, EG I, EG II, EG III,  $\beta$ -glucosidase. The dominant hemicellulases found in most enzyme cocktails were Endoxylanase (GH 11),  $\beta$ -xylosidase, arabinofuranosidase (GH 62 & 54) & Glucuronidase (GH 67). This data will help determine critical classes of cellulases and hemicellulases necessary for hydrolyzing lignocellulosic biomass and are currently absent in commercially available mixtures. Supplementation of a minimalist and optimal enzyme set for hydrolyzing ammonia fiber expansion (AFEX) treated biomass (i.e. corn stover) will help reduce the total number and amount (mg protein/gm substrate) of enzymes required for hydrolysis.

The quantitative comparison of *E. coli* grown aerobically and anaerobically serves the preliminary basis for genetically engineering the organism to ferment ethanol. Analysis of aerobic and anaerobic cell cultures yielded a combined total of 1697 proteins identified from the two cultures using strict cross correlation and cutoff values for the peptides and the requirement of two peptides per proteins. Of the 1697 proteins identified, 46 proteins were found in the anaerobic cultures only, 30 peptides were found in the aerobic cultures only, and 1621 proteins were identified in both cultures. Quantitative analysis of 1254 proteins between both of the samples showed 86 proteins showed at least a two-fold increase in abundance in the aerobic sample and 56 showed at least a two-fold increase in abundance in the anaerobic sample. The qualitative and quantitative characterization of *Rhodobacter sphaeroides* proteome is being used to report on protein expression changes associated with solar- and feedstock-powered hydrogen accumulation.

Also, preliminary characterization of microbial communities isolated from Panamanian leaf-cutting ant colonies, which are known to degrade lingo-cellulose to organic carbon, show differences in microbial community protein expression between different sample areas of the nests.

GTL

## Protein Expression Approaches to Cellulose Destruction

Nathaniel L. Elsen, Abolfazl Arabshahi, Allison Riederer, and **Brian G. Fox\*** (bgfox@biochem.wisc.edu)

Dept. of Biochemistry, Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, Wis.

<http://www.greatlakesbioenergy.org>

**Project Goals: The goal of this project is to provide efficient methods for the discovery of new combinations of enzymes for improved destruction of cellulosic biomass.**

Our long-term focus will be to provide a new combinatorial paradigm for evaluation of novel enzymes from new environmental sources as well as synthesized genes and engineered enzymes. The GLBRC bioenergy platform derives from work at the NIH Protein Structure Initiative-funded Center for Eukaryotic Structural Genomics, where over 10,000 genes from various eukaryotic organisms have been cloned, tested for expression, and in the best performing cases, purified and subjected to structure determinations. In this other project, over 1000 proteins have been purified, and over 100 protein structures have been determined. An adaptation of the modular design of this platform provides the basis for this new effort on genes and proteins contributing to cellulose destruction. Vector design principles will be discussed, and a catalog of vectors available will be presented. Some of these vectors are already available in the public domain at the NIH Protein Structure Initiative Material Repository, <http://www.hip.harvard.edu/PSIMR>. Results of the application of our developed methods to characterization of the reactivity of various treated and untreated biomass materials will be presented.

This work is funded by DOE DE-FC02-07ER64494 as a sub-grant from the Great Lakes Bioenergy Research Center to B.G.F.

GTL

## Engineering Cellulases with Improved Stability

**George N. Phillips Jr.\*** (phillips@biochem.wisc.edu), Thomas J. Rutkoski, Robert W. Smith, and He Zhang

Great Lakes Bioenergy Research Center and Dept. of Biochemistry, University of Wisconsin, Madison, Wis.

<http://www.greatlakesbioenergy.org>

**Project Goals: To improve the efficiency or lower the cost of enzymes used to degrade cellulose to fermentable sugars for biofuel production.**

Although significant progress has been made, the enzymatic hydrolysis of lignocellulosic feedstocks continues to be a significant factor affecting the economical production of cellulosic ethanol. Enzymes account for \$0.10–0.25/gal of cellulosic ethanol produced. In its 2006 publication, “Breaking Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda”, the DOE has targeted a 10-fold reduction in cellulase cost as an R&D milestone. The goal of 1–2 cents for cellulase per gal of ethanol is comparable to the cost of amylase used in the production of ethanol from corn grain. Protein engineering efforts, incorporating both rational and directed evolution strategies, will be essential for reaching this ambitious goal. Specifically, enhancing the thermostability of these industrial enzymes would allow for higher specific activity, reduce the amount of enzyme loading during hydrolysis, and allow greater flexibility in process configurations. Here, we describe preliminary efforts to further enhance the conformational stability of cellulases.

Supported by the Department of Energy (DE-FC02-07ER64494).

GTL

## Simple Chemical Transformation of Lignocellulosic Biomass into Fuels and Chemicals

Joseph B. Binder<sup>1,2,\*</sup> and **Ronald T. Raines<sup>1,2,3</sup>** (rtraines@wisc.edu)

<sup>1</sup>Great Lakes Bioenergy Research Center, <sup>2</sup>Dept. of Chemistry, and <sup>3</sup>Dept. of Biochemistry, University of Wisconsin, Madison, Wis.

<http://www.greatlakesbioenergy.org>  
<http://www2.biochem.wisc.edu/raines>

**Project Goals: Overcome biomass recalcitrance and transformation challenges through solvents tailored to biomass and innovative catalysis for carbohydrate conversion.**

Lignocellulosic biomass is a plentiful, renewable, and largely untapped resource for fuels and chemicals. Despite this potential, nearly all renewable fuels and chemicals are now produced from edible resources, such as starch, sugars, and oils; the challenges imposed by notoriously recalcitrant and heterogeneous lignocellulosic feedstocks have made their production from non-food biomass inefficient and uneconomical.

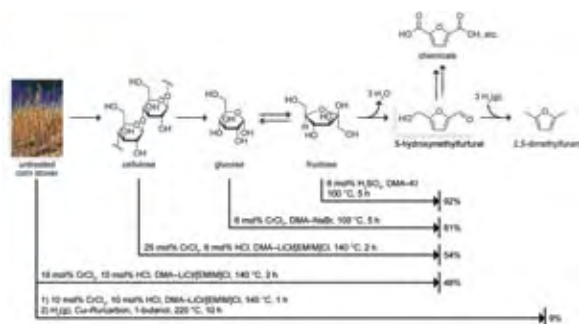
Here, we report that *N,N*-dimethylacetamide (DMA) containing lithium chloride (LiCl) is a privileged solvent that enables the synthesis of the renewable platform chemical 5-hydroxymethylfurfural (HMF) in a single step and unprecedented yield from untreated lignocellulosic biomass, as well as from fructose, glucose, and cellulose (see Figure). Mechanistic analyses reveal that loosely ion-paired halide ions in DMA–LiCl are critical for the remarkable rapidity

(1–5 h) and yield (up to 92%) of this low-temperature ( $\leq 140$  °C) process.

We also show that, in a second step, the HMF product can be converted into a promising liquid fuel: 2,5-dimethylfuran (DMF). DMF has an energy-content similar to that of gasoline and 40% greater than that of ethanol. Moreover, DMF is less volatile than ethanol and is immiscible with water. These attributes bode well for the use of DMF as an alternative fuel that is compatible with extant infrastructure.

Finally, we have demonstrated that our chemical approach enables efficient biomass saccharification to deliver sugars for fermentation processes.

Thus, a simple chemical transformation of lignocellulose can complement extant bioprocesses, providing a new paradigm for the use of biomass as a raw material for renewable energy and chemical industries.



**Figure.** Halide salts in DMA enable previously elusive yields of bio-based chemicals from a variety of carbohydrates. Conditions are optimized for the conversion of carbohydrates into HMF (1 step) and DMF (2 steps). (Photograph courtesy of DOE/NREL.)

solidated bioprocessor *E. coli* strains, capable of the complete conversion of lignocellulose to ethanol.

The conversion of *E. coli* to a consolidated processor requires the introduction of heterologous genes responsible for cellulose degradation, as well as a secretory system for their transport from the cell. We are employing two parallel approaches to solve the secretion problem. First, we are introducing inducible promoters to activate the cryptic Type II secretory apparatus within *E. coli*. Second, we will engineer *E. coli* to express genes encoding the Type II secretory apparatus from closely related bacteria. We are also engineering *E. coli* to more efficiently produce and tolerate ethanol. To improve the ability of *E. coli* to produce ethanol from the C5 and C6 sugars generated from lignocellulose degradation, we are using candidate gene approaches, as well as random mutagenesis. Furthermore, we are employing metabolic modeling to identify novel combinations of mutations that link growth rate to the amount of ethanol production. Strains containing these combinations of mutations will then be subjected to directed evolution to identify variants with improved growth rates, with the expectation that such strains will also show improved rates of ethanologenesis. The mutants will then be subjected to resequencing, as well as a combination of metabolomics, proteomics, and transcriptomics, aimed at understanding the molecular mechanism behind the improved ethanologenesis. Collectively, these approaches will allow for the isolation of lead organisms that can then be subjected to further rounds of directed evolution. These studies will also produce an improved molecular understanding of the current limitations of ethanologenesis, and allow for the development of novel flexible approaches useful in diverse ethanologenic microorganisms.

GTL

GTL

## Construction of a Consolidated Bioprocessor Derived from *Escherichia coli*

David H. Keating\* (dkeating@glbrc.wisc.edu), Mary Tremaine, and Robert Landick

Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, Wis.

<http://www.greatlakesbioenergy.org>

**Project Goals:** The goal of GLBRC is to remove obstacles to economical microbial cellulosic ethanologenesis by generating new fundamental understanding of the cellular processes that underlie these bottlenecks.

Research at the Great Lakes Bioenergy Research Center aims to generate an improved understanding of the bottlenecks associated with the conversion of lignocellulose to ethanol. Our studies focus on the bacterium *Escherichia coli*, due to its sophisticated genetics, well-understood physiology, and use as an industrial microbe. We aim to construct con-

## Molecular, Genetic and Genomic Approaches to Alleviate Bottlenecks in Cellulosic Ethanol Production by Yeast

Trey Sato<sup>1\*</sup> (tksato@glbrc.wisc.edu), Jennifer Van Vleet,<sup>1,2,3</sup> Thomas Jeffries,<sup>2,3</sup> Jeffrey Lewis,<sup>1,3</sup> and Audrey Gasch<sup>1,3</sup>

<sup>1</sup>Great Lakes Bioenergy Research Center, Madison, Wis.;

<sup>2</sup>USDA Forest Products Laboratory, Madison, Wis.; and

<sup>3</sup>University of Wisconsin, Madison, Wis.

<http://www.greatlakesbioenergy.org>

**Project Goals:** The mission of the GLBRC is to perform the basic research needed to generate the technology that realizes the potential of cellulosic biomass to produce ethanol and other advanced biofuels.

Although progress has been made in engineering the yeast, *Saccharomyces cerevisiae*, to ferment cellulosic feedstocks, significant bottlenecks remain that limit the yield and efficiency of this process. These bottlenecks include

1) inhibition of fermentation by toxins generated in the pretreatment of cellulosic feedstocks and the fermentation process, 2) incapacity to hydrolyze cellulose and cellobiose into usable monomeric sugars and 3) an inability to ferment xylose, the second most prevalent monomeric sugar found in cellulose after glucose. The goal of our project is to utilize novel molecular, genetic and genomic approaches on the yeasts, *S. cerevisiae* and the xylose-fermenting *Pichia stipitis*, to develop a consolidated bioprocessing organism that can produce higher yields of cellulosic ethanol more efficiently than the current yeast standards. Our approach differs from other yeast engineering approaches in two key ways: 1) we will perform high-throughput screening of wild yeast strains to identify genetic backgrounds ideal for cellulosic fermentation, and 2) we will simultaneously engineer those strains for cellulase secretion, xylose fermentation, and multi-stress resistance, rather than optimize traits individually. This approach will advance the development of a single yeast strain with superior cellulosic fermentation properties, whereas optimization of single traits (e.g., xylose metabolism) often occurs at the expense of other desired phenotypes (such as stress tolerance). Here, we present our current strategies and ongoing research to engineer yeast at the Great Lakes Bioenergy Research Center for the production of ethanol from cellulosic feedstocks.

GTL

## Catalytic Processing of Carbohydrates for the Production of Liquid Fuels

Drew J. Braden<sup>1\*</sup> (braden@wisc.edu), Elif I. Gurbuz,<sup>1</sup> Edward L. Kunkes,<sup>1</sup> Dante A. Simonetti,<sup>2</sup> Ryan M. West,<sup>1</sup> Juan Carlos Serrano-Ruiz,<sup>1</sup> Christian A. Gartner,<sup>1</sup> and **James A. Dumesic<sup>1</sup>**

<sup>1</sup>University of Wisconsin, Madison, Wis.; and <sup>2</sup>University of California, Berkeley, Calif.

<http://www.greatlakesbioenergy.org>

**Project Goals:** Our goal is to develop a catalytic processing approach for the conversion of concentrated aqueous carbohydrate solutions into liquid fuels. The catalytic conversion approach involves the use of a combination of flow reactors operating in cascade mode where the effluent of one reactor is simply flown into the next. We have had success starting with feeds consisting of C<sub>5</sub> and C<sub>6</sub> sugars. Our goal now is to use feeds derived directly from lignocellulose. Specifically, we will examine the effects of solubilized lignin compounds on the catalyst surface chemistry. We will also investigate potential reactions for upgrading the solubilized lignin into liquid fuel components.

Biofuels can be produced in large volume from a wide variety of renewable sources through potentially carbon neutral processes [1]. In this respect, we have recently developed a catalytic approach for the conversion of aqueous C<sub>6</sub>-sugar solutions, such as glucose and sorbitol, into gasoline, jet and

diesel fuels [2]. The approach, shown in Figure 1, involves a combination of flow reactors operating in series, where the effluent of one reactor is simply fed into the next reactor.

Concentrated aqueous carbohydrate solutions (40–60 wt%) are initially reacted in a flow reactor over a Pt–Re catalyst. The carbohydrate species adsorbed on the catalyst surface undergo successive dehydration and hydrogenation reactions that effectively deoxygenate the molecule, increasing its hydrophobicity, as depicted in the bottom portion of Figure 1. In addition to the exothermic deoxygenation reactions occurring, part of the carbohydrate feed is converted to H<sub>2</sub> and CO<sub>2</sub> via endothermic reforming reactions. The *in situ* generation of H<sub>2</sub> eliminates the need for an external source of H<sub>2</sub> and balances the overall thermochemistry of the reaction, such that the overall conversion is mildly exothermic. The resulting reactor effluent consists of three phases: a gas phase, an organic phase and an aqueous phase.

The organic layer from the first catalytic processing step consists of monofunctional hydrocarbons in the form of alcohols, ketones, carboxylic acids and alkanes. This organic layer can be upgraded to components that are currently used in transportation fuels. For example, the monofunctional hydrocarbons can be combined via C–C coupling in aldol condensation reactions on CuMg<sub>10</sub>Al<sub>7</sub>O<sub>x</sub> to produce the C<sub>8</sub> to C<sub>12</sub> linear alkanes used in diesel and jet fuel. Similarly, aromatic compounds used in gasoline can be produced from the organic species through aromatization reactions over an H-ZSM-5 catalyst. One advantage of this approach is that following the first processing step, more than 80% of the oxygen contained in the carbohydrate has been removed, allowing the subsequent upgrading processes to operate at reduced capacity and increased efficiency.

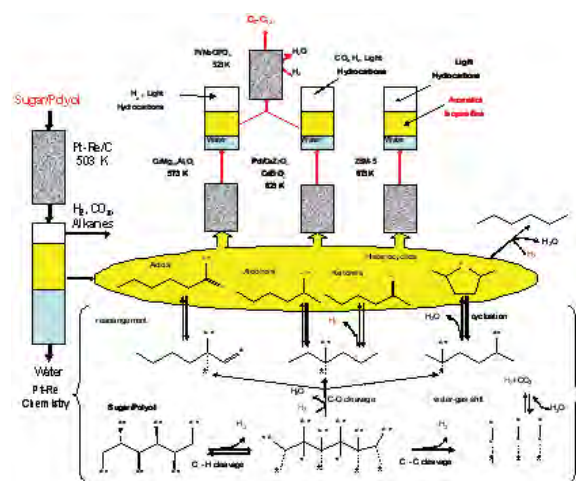


Figure 1. Schematic representation of carbohydrate processing for the production of monofunctional organic compounds, shown in yellow, and subsequent upgrading reactions in series. The proposed Pt–Re surface chemistries involved are shown in brackets where the asterisk signifies a catalyst site [2].

An essential requirement for overall effectiveness of our approach is to extend the process to utilize feeds derived directly from lignocellulose. Our initial work focused on the

C<sub>6</sub>-sugar glucose, the monomer that makes up cellulose, as well as the C<sub>6</sub>-polyol sorbitol. However, C<sub>6</sub> sugars only make up 25-40% of the total content of lignocellulose. Following our initial work, we have now successfully converted C<sub>5</sub> sugars, the main constituent of hemicellulose, to monofunctional organic intermediates with results comparable to C<sub>6</sub> sugar processing. With the combined processing of both C<sub>5</sub> and C<sub>6</sub> sugars, we are now able to convert approximately 70% of the content of lignocellulose into platform monofunctional organic compounds.

The next challenge in our work is to process feeds that include components that are not sugars, i.e., components derived from lignin. In particular, we are studying the effects of solubilized lignin components in mixed carbohydrate feeds to assess the role of these species in Pt-Re surface chemistry, and to determine whether these components can be upgraded to fuel components through our cascade mode catalytic approach.

#### References

1. B. Dale, *J. Agric. Chem.*, 56, 11 (2008).
2. E.L. Kunkes *et al.*, *Science* 322, 417 (2008).

tests indicate an approximately ten-fold higher production of C<sub>8</sub> to C<sub>18</sub> fatty acids in *E. coli* K-12 MG1655  $\Delta$ *fadD* when overexpressing ACC and BTE on plasmids. The predominant fatty acid chain length also shifts dramatically from C<sub>16</sub> to C<sub>12</sub> when BTE is expressed. Ongoing work includes optimization of ACC expression and identification of new metabolic and regulatory bottlenecks.

GTL

### Networks Contributing to Photosynthetic Biohydrogen Production in *Rhodobacter sphaeroides*

Wayne S. Kontur<sup>1\*</sup> (wskontur@wisc.edu), Eva C. Ziegelhoffer,<sup>1</sup> L. Safak Yilmaz,<sup>2</sup> Daniel R. Noguera<sup>2</sup> (noguera@engr.wisc.edu), and Timothy J. Donohue<sup>1</sup> (tdonohue@bact.wisc.edu)

<sup>1</sup>Dept. of Bacteriology and <sup>2</sup>Dept. of Civil and Environmental Engineering, University of Wisconsin, Madison, Wis.

<http://www.greatlakesbioenergy.org>

**Project Goals:** We seek to understand and optimize the light- and feedstock-powered production of hydrogen gas by the photosynthetic bacterium *Rhodobacter sphaeroides* for use as a clean and renewable source of energy. We are focusing our investigation on the various cellular pathways involved in hydrogen production. We want to determine the efficacy of hydrogen production from wild-type cells under different growth conditions, such as the carbon source used. We are analyzing these cultures via microarrays, proteomics, and other assays to determine the pathways that are utilized under the different conditions and, from this information, to gain insight into which pathways contribute to or detract from hydrogen production. We are also determining the effects of various mutations on hydrogen production. We will use these investigations to inform us on the design and development of strains, engineered through further mutation, that are optimized for hydrogen production under the conditions most relevant for large-scale systems.

*Rhodobacter sphaeroides* is a photosynthetic purple non-sulfur bacterium that can produce hydrogen gas from its nitrogenase enzyme, either concurrent with or independent of nitrogen fixation, during photoheterotrophic growth (and possibly from other, less-characterized sources). To assess the solar powered production of hydrogen under different conditions, *R. sphaeroides* wild-type strain 2.4.1 was grown photosynthetically in glutamate-containing medium (a poor nitrogen source that promotes the expression of nitrogenase) consisting of different carbon sources. Gas production analysis shows that the maximum rate of hydrogen production and the total amount of hydrogen produced from a culture depend on the identity of the carbon source used. We find general correlations between the ability of the bacteria to produce hydrogen and both the reducing potential

### Engineering *E. coli* for Production of Hydrocarbons

Brian Pflieger (pflieger@engr.wisc.edu) and Rebecca Lennen\* (lennen@wisc.edu)

University of Wisconsin, Madison, Wis.

<http://www.greatlakesbioenergy.org>

**Project Goals:** We have developed a strain of *E. coli* that overproduces fatty acids through metabolic engineering. Our goals for this year include: discovering limiting biochemical steps through transcriptomic and proteomic analysis and metabolic modelling; further optimizing expression of acetyl-CoA carboxylase and acyl-ACP thioesterase; expressing and characterizing activity of a type I fatty acid synthase in *E. coli*; and implementing a high-throughput screen for discovery of improved fatty acid biosynthetic genes.

Long-chain hydrocarbons have higher energy densities than alcohols and are immiscible with water. They are therefore an attractive target for development of an engineered micro-organism for sustainable fuel production. One potential platform for the synthesis of long-chain hydrocarbons in *Escherichia coli* is through reduction of fatty acids. The initial development of a fatty acid overproducing strain is reported. This strain features: (1) deletion of *fadD*, which encodes an acyl-CoA synthetase necessary for beta-oxidation; (2) over-expression of the four subunits of acetyl-CoA carboxylase (ACC), which converts acetyl-CoA to malonyl-CoA, a known bottleneck in fatty acid biosynthesis; and (3) heterologous expression of a codon-optimized plant medium chain acyl-acyl carrier protein thioesterase (BTE). Initial

GTL



and the chemical nature of the carbon source. Among the organic acids tested, the order of maximum production rate (in  $\mu\text{L H}_2/\text{mL culture/hr}$ ) is: lactate (~90) > succinate (~61) > malate (~50) > tartrate (~32) > gluconate (non-detectable). The three sugars tested (glucose, xylose, and fructose) have similar maximum rates of hydrogen production, ~30-40  $\mu\text{L H}_2/\text{mL culture/hr}$ . Cultures at the point of maximum hydrogen production rate were analyzed by microarrays, proteomics, and other assays to identify the metabolic pathways employed by *R. sphaeroides* that can contribute to or detract from hydrogen production. We find that a large percentage of the genes that are transcriptionally either up- or down-regulated in hydrogen producing cultures versus non-hydrogen producing cultures is related to metabolism. Among the genes up-regulated are those encoding subunits of the nitrogenase and hydrogenase proteins, as well as those related to electron transport, such as ferredoxins and flavins. Among the genes down-regulated are those involved in carbon fixation and the production of bioplastic polymers such as polyhydroxybutyrate, which are processes that compete with nitrogenase for reducing power. In addition to our studies of wild-type *R. sphaeroides*, we will report on studies to test the ability of various mutant strains (affected in electron flow pathways, light harvesting capacity, etc.) to produce hydrogen gas under comparable conditions.

GTL

## Bacterial Communities in the Rhizosphere of Biofuel Crops as Evaluated by 16S rRNA Pyrosequencing

Ederson da Conceição Jesus\* (ederson@msu.edu), Endang Susilawati (endangsusilawati@gmail.com), Qiong Wang (wangqiong@msu.edu), Jorge Rodrigues (jorge@uta.edu) and **James M. Tiedje** (tiedje@msu.edu)

Michigan State University, East Lansing, Mich.

<http://www.greatlakesbioenergy.org>

**Project Goals:** The goals of our project are to improve (a) the characteristics of biomass plants (Thrust 1); (b) the procedures for processing plant biomass (Thrust 2); (c) the biological or chemical processes to convert biomass into energy products (Thrust 3); and (d) the economic and environmental sustainability of the biomass-to-biofuel pipeline (Thrust 4). The specific goals of our group is to assess the structure and functional diversity of the rhizosphere community of biofuel energy crops, measure the extent of biological nitrogen fixation involved in biofuel crop production and describe and control interactions with symbiotic microbes and identify plant genes required for or regulating the establishment of AM symbioses.

The region around plant roots—the rhizosphere—harbors different microbial species which can fix nitrogen, protect plants against bacterial pathogens, produce plant growth

factors and aid soil structure, thus providing for the overall promotion of plant growth. Managing these microbial communities can help to improve biomass production and decrease production costs as well. One of the steps toward reaching this goal is to know which species are present in the rhizosphere community and how environmental and plant factors affect microbial community structure. We studied bacterial communities in soils cultivated with several crops with potential to be used as biofuel crops, i.e. switchgrass, big blue stem, orchard grass, tall fescue, corn, soybean, canola and sunflower. Bulk and rhizosphere soil communities have been analyzed. 16S rRNA gene sequences have been amplified from community DNA with tagged-primers and sequenced by 454 technology, generating thousands of sequences that have been used for analysis of community structure and composition. Preliminary results from soils cultivated with switchgrass, big blue stem, orchard grass and tall fescue show that differences in bacterial community structure and composition were correlated to plant genotype and soil attributes, such as pH and nutrient concentrations. This correlation accounted for about 29% of the variation in the studied bacterial communities. Clearer differences were observed when rhizosphere soil was examined compared to bulk soil, which can be explained by the fact that the influence of plant genotype is stronger in the rhizosphere than in the bulk soil. Several phyla were found with *Proteobacteria*, *Acidobacteria* and *Actinobacteria* among the most abundant.

GTL

## BioEnergy Cropping Systems on Marginal Land

Stephanie Smith\* (smith579@msu.edu) and **Kurt Thelen** (thelenk3@msu.edu)

Michigan State University, East Lansing, Mich.

<http://www.greatlakesbioenergy.org>

**Project Goals:** Evaluate bio-energy crop yield and fuel quality to determine differences in crops grown on marginal land compared to agriculturally productive soils. Extrapolate results based on soil type and climate zone throughout Michigan to determine the optimal regional bio-energy cropping system.

Agriculture faces unique challenges as increasing world population places unprecedented demands on food and energy resources. Global fossil fuels are finite in supply and are becoming more expensive to extract as supplies are diminished. Bioenergy crops are increasingly being seen as essential components of future energy plans. There is particular interest in the production of these bio-energy crops on marginally productive lands. As agricultural resources are stretched to meet food and energy demands, lower productivity land bases will likely be brought into production.

## Bioenergy Sciences Center (BESC)

GTL

### The BioEnergy Science Center—An Overview

M. Keller,<sup>1</sup> L. Lynd,<sup>2</sup> A. Davill,<sup>3</sup> and B. Davison<sup>1\*</sup>  
(davisonbh@ornl.gov)

<sup>1</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; <sup>2</sup>Thayer School of Engineering, Dartmouth College, Hanover, N.H.; and <sup>3</sup>Complex Carbohydrate Research Center, University of Georgia, Athens, Ga.

http://bioenergycenter.org

**Project Goals:** The challenge of converting cellulosic biomass to sugars is the dominant obstacle to cost-effective production of biofuels in sustained quantities capable of impacting U.S. consumption of fossil transportation fuels. The BioEnergy Science Center (BESC) research program addresses this challenge with an unprecedented interdisciplinary effort focused on overcoming the recalcitrance of biomass. By combining engineered plant cell walls to reduce recalcitrance with new biocatalysts to improve deconstruction, BESC within five years will revolutionize the processing of biomass. These breakthroughs will be achieved with a systems biology approach and new high-throughput analytical and computational technologies to achieve (1) targeted modification of plant cell walls to reduce their recalcitrance (using *Populus* and switchgrass as high-impact bioenergy feedstocks), thereby decreasing or eliminating the need for costly chemical pretreatment; and (2) consolidated bioprocessing, which involves the use of a single microorganism or microbial consortium to overcome biomass recalcitrance through single-step conversion of biomass to biofuels. Within five years the Center will remove biomass recalcitrance as a barrier to cost-effective biofuels production by achieving a minimum two-fold reduction in the projected cost of processing for conversion of biomass to ethanol. Through this effort we will greatly enhance our understanding of cell wall structure during synthesis and conversion. The benefits of the basic research will extend beyond the five-year program by laying the foundation for developing other biomass sources and fuel products, improving productivity of switchgrass and poplar, and ensuring sustainability of lignocellulosic biofuel production.

In addition to Oak Ridge National Laboratory (ORNL), the BESC core team consists of the Dartmouth College, the University of Georgia, the Georgia Institute of Technology, the University of Tennessee, the National Renewable Energy Laboratory, the Samuel Roberts Noble Foundation, and industrial partners ArborGen, Diversa, and Mascoma. Other individual PIs complete the team. The home base of BESC is the Joint Institute for Biological Sciences building, funded by the state of Tennessee and occupied in December 2007. Located on the ORNL campus, the facility is

designed specifically for interdisciplinary bioenergy research using systems biology tools. Other BESC anchor facilities include the University of Georgia's Complex Carbohydrate Research Center with extensive carbohydrate analytical and plant science expertise, the National Renewable Energy Laboratory's unique capabilities in comprehensive biomass analysis and bioprocessing, and ORNL's National Leadership Computing Facility

This poster will provide an overview of the BESC organization and year one activities. It will also provide context for the twenty other BESC posters which provide illustrative examples of specific BESC research efforts.

The BioEnergy Science Center is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

GTL

### The Use of TAIL PCR to Identify Genes Controlling Extreme Phenotypes in a *Populus* Activation Tagged Population

G.A. Tuskan<sup>1\*</sup> (gtk@ornl.gov), T. Tschaplinski,<sup>1</sup> U. Kalluri,<sup>1</sup> T. Yin,<sup>1</sup> L. Gunter,<sup>1</sup> M. Martin,<sup>1</sup> N. Engle,<sup>1</sup> S. Jawdy,<sup>1</sup> X. Yang,<sup>1</sup> Martin Keller,<sup>2</sup> V. Busov,<sup>3</sup> S. Strauss,<sup>4</sup> X. Ma,<sup>4</sup> M. Davis,<sup>5</sup> and R. Sykes<sup>5</sup>

<sup>1</sup>Environmental Sciences Division and <sup>2</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; <sup>3</sup>Michigan Technological University, Houghton, Mich.; <sup>4</sup>Oregon State University, Corvallis, Ore.; and <sup>5</sup>National Renewable Energy Laboratory, Golden, Colo.  
http://bioenergycenter.org

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

Bioenergy feedstock species provide the raw materials for biochemical conversion into sugars and ultimately liquid transportation fuels. The plant cell wall contains three major polymers including cellulose [i.e., a complex chain

of glucose molecules and the most abundant polymer on earth], hemicellulose [i.e., a complex mixture of five and six carbon sugars], and lignin [i.e., a polyphenolic matrix that protects the cell wall sugars]. Improving the yields of fermentable sugars from pretreated biomass is a key goal of the DOE Bioenergy Research Centers. Complete removal of lignin would be ideal for sugar yields but would be fatal to the plant. Thus, finding an optimum phenotype with a ratio of lignin to cellulose which results in high amounts of sugar and viability in the plant is one of our goals.

The detection of useful phenotypes is complicated by the long term nature of perennial woody plants such as *Populus*. Techniques to associate phenotypes with genotypes include conventional QTL studies, reverse genetics studies, naturally occurring mutants, as well as more sophisticated approaches such as activation tagging. During activation tagging a set of four tandemly repeated constitutive 35S promoters are randomly inserted into the genome of the target species. Plants are regenerated via tissue culture methods and the resulting intact plants are grown in the field for phenotypic evaluation.

Approximately 800 activation tagged lines of *Populus* were created in 2005 and established in field trials in eastern Oregon. Wood core samples and DNA were collected from these materials in February, 2008. Initial wood chemistry estimates, via molecular beam mass spectrometry (MBMS), were obtained from dried cores ground to 20 mesh size in a Wiley mill. Whole genome DNA template was extracted from leaf tissue and is being characterized via thermal asymmetric interlaced polymerase chain reaction (TAIL PCR). Sequences obtained from resulting TAIL PCR fragments are aligned to the *Populus* genome using BLASTN on the Joint Genome Institute (JGI) browser. Genes found within a  $\pm 5$  kb up and downstream region on the tag insert are candidate genes putatively affecting extreme wood chemistry phenotypes identified using MBMS.

Preliminary MBMS data suggests that the tested activation tagged lines have lignin values ranging from 17.6% (dry weight) to 23.0%, S/G ratios from 2.0 to 3.0, hemicellulose values from 20.0% to 25.6%, and cellulose values from 26.3% to 36.6%. Extreme phenotypes are defined as those values  $\pm 2.5$  standard deviations away from the wild type mean. Based on these criteria, 24 lines had at least one extreme phenotype, with three lines showing multiple extreme phenotypes. In these multiple phenotype lines it is likely that a single gene affected by the constitutive promoter is causing a change in the upstream portion of the carbohydrate biosynthesis pathway.

In order to identify the gene(s) associated with the extreme wood chemistry phenotype(s), TAIL PCR conditions were optimized using different combinations of reagent concentration, extension temperature and degenerative primers. Once conditions were optimized a defined tertiary reaction band was detectable on an agarose gel. These bands were excised, column purified and directly sequenced on an Applied Biosystem 3730 XL instrument. Individual reactions typically resulted in multiple bands, which may or may not be the same tagged locus [i.e. multiple insertions

commonly occur during activation tagging]. BLASTN results on an initial set of tagged lines used for optimization indicate that unique positions within the *Populus* genome, and consequently the affected genes, can be identified using this technique.

The submitted abstract has been authored by a contractor of the U.S. Government under contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes. 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.

GTL

## The pANIC Vector Set for Overexpression of Transgenes and RNAi-Mediated Knockdown of Native Genes in Switchgrass (*Panicum virgatum* L)

David G.J. Mann<sup>1\*</sup> (dmann1@utk.edu), Peter R. LaFayette,<sup>2</sup> Laura L. Abercrombie,<sup>1</sup> Wayne A. Parrott,<sup>2</sup> C. Neal Stewart Jr.,<sup>1</sup> and Martin Keller<sup>3</sup>

<sup>1</sup>Dept. of Plant Sciences, University of Tennessee, Knoxville, Tenn.; <sup>2</sup>Center for Applied Genetic Technologies, University of Georgia, Athens, Ga.; and <sup>3</sup>Biosciences Division, 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 multitolerant 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 detailed studies into poorly understood cell wall biosynthesis pathways. The Switchgrass activity is pursuing improved transformation methods, association studies and targeted pathways such as lignin synthesis.**

Switchgrass (*Panicum virgatum* L) is a promising feedstock for the production of cellulosic ethanol, and switchgrass biotechnology plays a pivotal role in the BioEnergy Science Center initiative. A crucial component of creating transgenic switchgrass is having the capability of transforming the explants with DNA sequences of interest using vector constructs. However, there are very limited monocot plant vectors currently available. With this in mind, we have designed

and constructed a versatile set of 24 Gateway-compatible destination vectors (termed “pANIC”) to be used in plants for transgenic crop improvement. Gateway compatibility allows for convenient insertion of any open reading frame (ORF) or other target sequence of interest. These vectors can be used for 1) transgene overexpression or 2) targeted gene silencing using double stranded RNA interference. Two main transformation methods exist for monocotyledonous plants: biolistic-mediated and *Agrobacterium*-mediated transformation. The pANIC vector set includes vectors which can be utilized for both applications, with all vectors containing three basic elements: 1) a Gateway cassette for overexpression or silencing of the target sequence, 2) a plant selection cassette and 3) a visual reporter cassette. The pANIC vector set allows for high throughput screening of sequences of interest in switchgrass, as well as other monocot plant species.

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

GTL

### Genetic Modification of Lignin Biosynthesis in Switchgrass

Hiroshi Hisano,<sup>1</sup> Hui Shen,<sup>2</sup> Rangaraj Nandakumar,<sup>1</sup> Chunxiang Fu,<sup>1</sup> Xirong Xiao,<sup>1</sup> Yaxin Ge,<sup>1</sup> Tim Hernandez,<sup>2</sup> Fang Chen<sup>2\*</sup> (fchen@noble.org), Richard A. Dixon,<sup>2</sup> Zeng-Yu Wang,<sup>1</sup> and **Martin Keller**<sup>3</sup>

<sup>1</sup>Forage Improvement Division and <sup>2</sup>Plant Biology Division, The Samuel Roberts Noble Foundation, Ardmore, Okla.; and <sup>3</sup>Biosciences Division, 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 multitasking 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 detailed studies into poorly understood cell wall biosynthesis pathways. The Switchgrass activity is pursuing improved transformation methods, association studies and targeted pathways such as lignin synthesis.**

Switchgrass (*Panicum virgatum*), a native C4 perennial grass throughout North America, is an excellent candidate for the production of cellulosic ethanol. The species has high

biomass yield, low input requirements, good stress tolerance and the ability to grow in marginal areas. Switchgrass contains abundant sugars in the form of cellulose and hemicellulose, which cannot be easily converted to ethanol. On the other hand, Lignin, a major component of the cell wall of switchgrass, has been recognized for its negative impact on cellulosic ethanol production. Therefore, genetic modification of lignin in switchgrass could lead to increased ethanol production.

Our project focuses on genetic manipulation of key lignin genes to reduce recalcitrance to saccharification and to improve ethanol production in switchgrass. The project involves isolation of genes involved in lignin biosynthesis, construction of transformation vectors, production of transgenic switchgrass plants with modified lignin content/composition, and chemical analyses of the transgenic materials. In this presentation, we will show our target genes and strategy of transformation 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.

GTL

### New Insights on the Mechanism of Xylan Biosynthesis

Maria J. Peña<sup>1\*</sup> (mpena@ccrc.uga.edu), Breeanna Urbanowicz,<sup>1</sup> Alan G. Davill,<sup>1,2</sup> Malcolm O’Neill,<sup>1</sup> William S. York,<sup>1,2</sup> and **Martin Keller**<sup>3</sup>

<sup>1</sup>Complex Carbohydrate Research Center and <sup>2</sup>Dept. of Biochemistry and Molecular Biology, University of Georgia, Athens, Ga.; and <sup>3</sup>Biosciences Division, 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 multitasking 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 detailed studies into poorly understood cell wall biosynthesis pathways. Biosynthesis research provides detailed studies in these and other plant models on key biosynthetic pathways for lignin, pectin, xylan and cellulose. The selection of target genes for transformation and validation has been coordinated and initiated.**

Xylan accounts for up to 30% of the secondary wall where it is believed to interact with cellulose and lignin to generate biomass that is recalcitrant to deconstruction by microbial enzymes. Understanding how xylan is synthesized and incorporated into the wall may allow us to engineer plants with biomass that is more readily converted to biofuel.

Our studies of *Arabidopsis* mutants that form defective secondary cell walls (Peña et al., *Plant Cell* 19:549–563 2007) have shown that *Arabidopsis* xylan contains a unique glycosyl sequence at its reducing end that is required for normal xylan synthesis and that xylan structure and biosynthesis are more complex than previously believed. Hardwood and softwood xylans also have this sequence at their reducing ends. Thus, it may be possible to modify xylan properties in biomass by targeting genes that control the biosynthesis of this sequence.

We have developed an *in vitro* xylan biosynthesis assay using microsomal membranes, fluorescence-labeled oligosaccharide acceptors, and UDP-sugar donors and have structurally characterized the products formed. These studies show that *Arabidopsis* microsomes contain xylosyltransferase and glucuronosyltransferase activities that likely work together in the synthesis of glucuronoxylan. Switchgrass microsomes contain enzymes that in the presence of UDP-Xyl also extend the Xyl acceptors. Our results suggest that the xylan backbone is extended by a comparable mechanism in monocots and dicots. We present the results of these studies and discuss their implications for the mechanism of xylan biosynthesis.

Supported by U.S. Department of Energy, Bioenergy Research Centers, grant DE-PS02-06ER64304.

GTL

## Functional Identification and Characterization of Sugar-1-P Kinases in *Arabidopsis*

Ting Yang<sup>1,2,3\*</sup> (goldfishing@gmail.com), Liron Bar-Peled,<sup>2</sup> Lindsay Gebhart,<sup>2</sup> John Glushka,<sup>2</sup> Maor Bar-Peled,<sup>1,2,4</sup> and **Martin Keller**<sup>5</sup>

<sup>1</sup>BioEnergy Science Center, <sup>2</sup>Complex Carbohydrate Research Center, <sup>3</sup>Dept. of Biochemistry and Molecular Biology, <sup>4</sup>Dept. of Plant Science, University of Georgia, Athens, Ga.; and <sup>5</sup>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 multitolerant 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 detailed studies into poorly understood cell wall biosynthesis pathways. Biosynthesis research provides detailed studies in these and other plant models on key biosynthetic pathways for lignin, pectin, xylan and cellulose. The selection of target genes for transformation and validation has been coordinated and initiated.**

The glycan salvage pathway is a metabolic process required during various stages of plant development: seed germination, pollen growth and embryo development, where storage carbohydrates are converted to simple sugars. The source of sugars for the salvage pathway likely utilizes monosaccharides derived from the hydrolysis of storage polysaccharides, free sugars that are released during wall remodeling, and during recycling of glycoprotein and glycolipids. Two major groups of enzymes exist in this pathway: sugar-1-P kinases and NDP-sugar pyrophosphorylases. Sugar-1-P kinases phosphorylate specific monosaccharide to their corresponding sugar-1-P, and subsequently the sugar-1-P may undergo pyrophosphorylation to form NDP-sugar.

D-galacturonic acid (D-GalA) is a major sugar residue found in pectin, plant primary cell wall polysaccharides, and in certain types of glycoproteins. Enzymes involved in synthesis of UDP-GalA in salvage pathway were identified over 40 years ago, yet the biological role of free GalA in plant metabolism remains elusive. Question remains as to the specificity of the kinase(s), and to the relative amount of GalA made in tissue; and the turn-over of pectin in various tissue to explain the need of GalA-kinase, remained unanswered. Here we report the discovery of a novel sugar-1-P kinase which phosphorylates GalA and to a lesser degree GlcA in *Arabidopsis*. The GalA-kinase was characterized and compared with galactose-kinase (GalK) using HPLC and NMR. To identify sugar and nucleotide specificities site-directed mutagenesis was applied.

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.

GTL

## GAUT12 (GALactURonosylTRANSFERase 12): A Putative Glycosyltransferase Involved in *Arabidopsis* Secondary Cell Wall Biosynthesis

Zhangying Hao<sup>1,3\*</sup> (haozhangying@gmail.com), Staffan Persson,<sup>4</sup> Kyle David Hutcheson,<sup>2,3</sup> Debra Mohnen,<sup>2,3</sup> and **Martin Keller**<sup>5</sup>

<sup>1</sup>Dept. of Plant Biology and <sup>2</sup>Dept. of Biochemistry and Molecular Biology, Complex Carbohydrate Research Center, BioEnergy Science Center (BESC), <sup>3</sup>University of Georgia, Athens, Ga.; <sup>4</sup>Max Planck Institute of Molecular

Plant Physiology, Potsdam, Germany; and <sup>5</sup>Biosciences Division, 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 multitolerant 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 detailed studies into poorly understood cell wall biosynthesis pathways. Biosynthesis research provides detailed studies in these and other plant models on key biosynthetic pathways for lignin, pectin, xylan and cellulose. The selection of target genes for transformation and validation has been coordinated and initiated.

GAUT12/IRX8 is a putative glycosyltransferase involved in *Arabidopsis* secondary cell wall biosynthesis. GAUT-type genes are widespread in plants (see poster by Y. Yin *et al.*). *Arabidopsis* is being used as a model for functional studies to be confirmed when transformants in *Populus* and switchgrass are ready. The *irregular xylem 8 (irx8)* mutant has reduced amounts of glucuronoxylan (GX) and, to a lesser degree, homogalacturonan (HG) in its walls. It was reported that the pentasaccharide sequence 4- $\beta$ -D-Xylp-(1-4)- $\beta$ -D-Xylp-(1-3)- $\alpha$ -L-Rhap-(1-2)- $\alpha$ -D-GalpA-(1-4)-D-Xyl previously identified in birch (*Betula verrucosa*) and spruce (*Picea abies*) was also found at the reducing end of GX in *Arabidopsis*. Due to the sequence similarity of GAUT12/IRX8 to GAUT1, a known HG: GalAT, we propose two hypotheses for GAUT12 function. Hypothesis 1: GAUT12/IRX8 may be involved in the synthesis of the *Arabidopsis* GX: reducing end pentasaccharide by catalyzing the addition to the  $\alpha$ -D-GalA onto the reducing end xylose. The moiety onto which the reducing end xylose is attached is not known. Hypothesis 2: The reduced amounts of GalA in *irx8* mutant walls suggests that GAUT12/IRX8 may be involved in the synthesis of a subfraction of HG (or possibly RG-I) to which xylan is covalently attached. Given the importance of understanding GAUT12/IRX8 function in wood formation and potentially in reducing biomass recalcitrance, we are working towards elucidating the enzyme function of GAUT12 by establishing robust biochemical assays for enzyme function. It is also our goal to identify potential protein partners that may interact with GAUT12 as it carries out its role in plant cell wall biosynthesis.

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.

## Isolation of Novel Biofuel-Relevant Thermophiles and the Identification of Extracellular Cellulolytic Enzymes Using Multi-Dimensional LC-MS/MS

James G. Elkins\* (elkinsjg@ornl.gov), Scott D. Hamilton-Brehm, Adriane Lochner, Richard J. Giannone, Tatiana Vishnivetskaya, Jennifer Mosher, Mircea Podar, Jennifer Morrell-Falvey, Steve Allman, Dawn Klingeman, Sue Carroll, Miriam Land, Steven D. Brown, Robert Hettich, Tommy Phelps, and **Martin Keller**

BioEnergy Science Center, Biosciences Division, 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 multitolerant 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 (e.g. *Clostridium thermocellum*) to understand novel strategies and enzyme complexes for biomass deconstruction. We are also searching natural diversity for enzymes and microbes—in particular, thermophilic anaerobes. Microbial research targets how the CBP microbe interacts with the features of the pretreated or native plant cell wall and testing strategies for using pure or mixed cultures for degradation.

Advanced conversion and fermentation technologies such as consolidated bioprocessing (CBP) offer the potential for a cellulose-based supply of liquid transportation fuels<sup>1</sup>. Ethanol production from CBP must include the use of robust microorganisms that are capable of efficient hydrolysis of lignocellulose with simultaneous fermentation of biomass-derived sugars to yield alcohol. Given the recalcitrance of biomass to enzymatic degradation, CBP must rely on a powerful suite of extracellular hydrolytic enzymes that are stable under what would likely be elevated temperatures. By integrating expertise in microbial ecology, microbiology, and omics-based approaches, we have established a microbial pipeline for the discovery and characterization of novel extremophiles that are able to break-down and metabolize plant-derived polymers.

Geothermally heated spring waters and sediments collected from Yellowstone National Park, Wyoming, USA, were selected as initial environments for screening and cultivation attempts. A total of 134 environmental samples were collected during two separate expeditions. Using these samples as a source of inoculum, enrichment cultures were established using dilute-acid pretreated switchgrass and *Populus*

as the primary carbon and energy source. Temperatures for growth corresponded to the original sample temperature and ranged from 60–85°C. To isolate individual organisms from stable enrichment cultures, a high-throughput (HT) isolation system using flow cytometry was developed that allowed rapid separation and growth of cellulolytic, extreme thermophiles in liquid culture. Based on 16S ribosomal RNA gene sequences, isolated organisms clustered within the bacterial divisions Firmicutes and Dictyoglomi and several previously uncultivated species of *Caldicellulosiruptor*, *Thermoanaerobacter*, and *Dictyoglomus* were recovered. Secondary screening for rates of growth on pretreated biomass and crystalline cellulose showed that isolates from the *Caldicellulosiruptor* group possessed the fastest growth rates on relevant plant materials and model substrates. An organism isolated from Obsidian Pool, YNP, designated *Caldicellulosiruptor* sp. OB47 grows optimally at 80°C and reaches cell densities >10<sup>8</sup> cells/ml on carbon sources such as cellobiose, Avicel (crystalline cellulose), xylan, pectin, filter paper, processed cardboard, and pretreated lignocellulosic biomass (switchgrass and *Populus*). In batch growth experiments, OB47 produced end-products of >2.5 mM acetate, >50 mM CO<sub>2</sub>, and >20 mM hydrogen. Comparative growth studies on 0.1% w/v Avicel between OB47 and known cellulolytic organisms, *Anaerocellum thermophilum*, *Caldicellulosiruptor saccharolyticus*, and *Clostridium thermocellum* revealed comparable growth rates which are among the fastest known for cellulolytic organisms<sup>2</sup>.

In order to gain more information regarding the functional enzyme system utilized by *Caldicellulosiruptor* sp. OB47 for growth on cellulose, genomic data was obtained from 454 pyrosequencing followed by assembly and annotation to construct a reference database of translated open reading frames. Initially, OB47 was grown on several cellulose-based substrates and cell pellets were harvested and prepared for proteomics analysis using multidimensional liquid chromatography mass spectroscopy (LC-MS/MS) to establish a baseline proteomic landscape. Using the translated draft OB47 genomic sequence as the reference protein database, roughly 1000 non-redundant protein identifications were obtained, corresponding to roughly 40% of the predicted expressed proteins. LC-MS/MS technology was then used to identify single components of the major extracellular hydrolytic enzymes expressed by OB47 during growth on filter paper. The extracellular protein fraction was also evaluated using filter paper cellulase assays to obtain general data about hydrolysis rate, optimum temperature, and enzymatic stability of the extracellular enzyme mixture. Over several biological and technical replicates, roughly 75 non-redundant proteins were identified by LC-MS/MS. Expectedly, several cellulases / glycoside hydrolases, cellobiosidases, solute-binding proteins, as well as S-layer domain-containing proteins were identified as the top hits when sorted by protein abundance. Using this analysis as a proof-of-concept, it was expanded to include growth on several different biofuel-relevant substrates such as switchgrass and *Populus*. In addition, this analysis can be readily expanded to include other candidate organisms which would provide a basis for comparison and allow for insights into different strategies (i.e. cellulosome vs. free enzymes) and particularly

which families of glycosyl hydrolases individual microorganisms are expressing during growth on cellulosic materials.

## References

1. Lynd, L.R., M.S. Laser, D. Bransby, B.E. Dale, B. Davison, R. Hamilton, M. Himmel, M. Keller, J.D. McMillan, J. Sheehan, and C.E. Wyman. How biotech can transform biofuels. *Nature Biotechnology* **26**, 169–172 (2008).
2. Lynd L.R., P.J. Weimer, W.H. van Zyl, and I.S. Pretorius. Microbial Cellulose Utilization: Fundamentals and Biotechnology. *Microbiol. Mol. Biol. Rev.* **66**, 506–577 (2002).

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.

GTL

## Degradation of Plant Biomass without Pretreatment by the Thermophilic Anaerobe, *Anaerocellum thermophilum*

Irina Kataeva<sup>1\*</sup> (kataeva@uga.edu), Sung-Jae Yang,<sup>1</sup> Nick Sennett,<sup>1</sup> Aleksandar Cvetkovic,<sup>1</sup> Farris L. Poole II,<sup>1</sup> Scott D. Hamilton-Brehm,<sup>1,3</sup> Richard J. Giannone,<sup>3</sup> Robert L. Hettich,<sup>3</sup> Nancy E. Engle,<sup>3</sup> Timothy J. Tschaplinski,<sup>3</sup> Janet Westpheling,<sup>1,2</sup> Michael W.W. Adams,<sup>1,2</sup> and **Martin Keller**<sup>3</sup>

Depts. of <sup>1</sup>Biochemistry & Molecular Biology and <sup>2</sup>Genetics, University of Georgia, Athens, Ga.; and <sup>3</sup>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 multitolerant 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 (e.g. *Clostridium thermocellum*) to understand novel strategies and enzyme complexes for biomass deconstruction. We are also searching natural diversity for enzymes and microbes—in particular, thermophilic anaerobes. Microbial research targets how the CBP microbe interacts with the features of the pretreated or native plant cell wall and testing strategies for using pure or mixed cultures for degradation.**

We are investigating the conversion of switchgrass and the hardwood poplar by the anaerobic thermophile *Anaerocellum thermophilum*. Our goals are to understand at molecular level the mechanisms by which lignocellulosic biomass is degraded, and how these mechanisms differ with the type of plant biomass that is used and with defined substrates such as crystalline cellulose and xylan.

We have demonstrated that *Anaerocellum thermophilum*, an anaerobic and thermophilic bacterium that grows optimally at 75 °C, is able to degrade various types of unprocessed plant biomass, as well as defined carbohydrates such as crystalline cellulose and xylan. The plant materials that are utilized include hardwoods such as poplar, low lignin grasses such as napier and bermuda, and those with high lignin content such as switchgrass. The predominant reduced end products from all of these growth substrates are hydrogen and lactate. Glucose, cellobiose and, to a lesser extent, celotriose (on crystalline cellulose) and xylose, xylobiose and some xylotriose (on xylan) also accumulated in the growth media during growth on the defined substrates, but these sugars did not accumulate during growth on the plant biomass. Substrate accessibility did not limit the growth of *A. thermophilum* on the plant material as the organism grew well on first- and second spent biomass derived from poplar and switchgrass. Spent biomass is the insoluble growth substrate recovered after the organism has reached late stationary phase. Extracellular extracts of *A. thermophilum* grown on various types of biomass have been subjected to proteomic, metabolomic, chromatographic, metal and enzymatic analyses. The results are being used to provide insights into the mechanisms of plant biomass conversions. We have also shown that a close relative of *A. thermophilum*, *Caldicellulosiruptor saccharolyticus* (T<sub>opt</sub> 70°C), grows well on switchgrass but not on poplar. Additional insights into mechanisms of biomass conversion are being obtained from a comparison of the genome sequences of these two thermophilic anaerobes.

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

GTL

## Integration of Genomics and Bioinformatics to Identify Genetic Differences in an Ethanol Tolerant *Clostridium thermocellum* ATCC27405 Strain

Steven D. Brown<sup>1\*</sup> (brownsd@ornl.gov), Tatiana Karpinets,<sup>1</sup> Jonathan R. Mielenz,<sup>1</sup> Shihui Yang,<sup>1</sup> Dawn M. Klingeman,<sup>1</sup> Miriam L. Land,<sup>1</sup> Loren J. Hauser,<sup>1</sup> Babu Raman,<sup>1</sup> Miguel Rodriguez Jr.,<sup>1</sup> Tingfen Yan,<sup>1</sup> Tatiana A. Vishnivetskaya,<sup>1</sup> Herbert J. Strobel,<sup>2</sup> Maria V. Sizova,<sup>3</sup> Nicolai Panikov,<sup>3</sup> Ying Xu,<sup>4</sup> Phuongan Dam,<sup>4</sup> Lee Lynd,<sup>3</sup> and **Martin Keller**<sup>1</sup>

<sup>1</sup>Biosciences Division and BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tenn.; <sup>2</sup>Dept. of Animal and Food Sciences, University of Kentucky, Lexington, Ky.; <sup>3</sup>Thayer School of Engineering, Dartmouth College, Hanover, N.H.; and <sup>4</sup>Institute of Bioinformatics, University of Georgia, Athens, Ga.

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 (e.g. *Clostridium thermocellum*) to understand novel strategies and enzyme complexes for biomass deconstruction. We are also searching natural diversity for enzymes and microbes—in particular, thermophilic anaerobes. Microbial research targets how the CBP microbe interacts with the features of the pretreated or native plant cell wall and testing strategies for using pure or mixed cultures for degradation.**

*Clostridium thermocellum* is a gram-positive, anaerobic, thermophilic bacterium that can ferment cellulose at one of the highest growth rates directly to ethanol via a large extracellular enzyme complex termed the cellulosome. *C. thermocellum* is a candidate industrial biocatalyst for future lignocellulosic fuel production. The metabolic byproducts of fermentation can inhibit fermentation performance and lignocellulosic biomass pretreatment processes also produce a variety of inhibitory chemicals that can adversely affect the fermentation. Limited information is available on the mechanisms and responses of *C. thermocellum* to different inhibitors. The genetic differences between wild-type *C. thermocellum* and an ethanol tolerant mutant have been identified through microarray based comparative genome sequencing and 454-pyrosequencing. We detected more than 400 differences in the ethanol tolerant mutant compared to the *C. thermocellum* wild-type strain. The resequencing data were in agreement with published membrane proteomic data and identified new mutations in key genes such as alcohol dehydrogenase. Bioinformatics analyses identified 16 mutational hot-spots in the ethanol tolerant strain, with 7 out of 16 related to cellulose degradation and likely accounted for the strain's decreased growth on cellulose. Further work to identify and verify important loci and physiological changes conferring tolerance to inhibitors will assist in the development of industrial strains for consolidated bioprocessing (CBP) of lignocellulosic biomass and therefore reduce biofuel production costs.

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.



GTL

## Advances in Microbial Cellulose Utilization: Methods Development and Kinetics

Nicolai Panikov<sup>1\*</sup> (Nicolai.Panikov@Dartmouth.edu), Maria Sizova,<sup>1</sup> Evert Holwerda,<sup>1</sup> Julie Paye,<sup>1</sup> Lois St. Brice,<sup>1</sup> Lee Lynd,<sup>1</sup> and **Martin Keller**<sup>2</sup>

<sup>1</sup>Thayer School of Engineering, Dartmouth College, Hanover, N.H.; and <sup>2</sup>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 multitolerant 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 (e.g. *Clostridium thermocellum*) to understand novel strategies and enzyme complexes for biomass deconstruction. We are also searching natural diversity for enzymes and microbes—in particular, thermophilic anaerobes. Microbial research targets how the CBP microbe interacts with the features of the pretreated or native plant cell wall and testing strategies for using pure or mixed cultures for degradation.**

Microbial cellulose utilization has great biotechnological promise, particularly for developing microbes capable of mediating consolidated bioprocessing. However, current understanding is profoundly limited with respect to most aspects of microbial cellulose fermentation. Methods for accomplishing tasks such as substrate and cell quantification are cumbersome, and the literature contains few if any controlled quantitative comparisons of the rates of microbial cellulose utilization. Owing partly to the absence of such comparative studies, we have no idea under what conditions microbial cellulose utilization occurs fastest. The first mechanistic kinetic model for microbial cellulose utilization has yet to be proposed.

BESC-supported research underway at Dartmouth aimed at meeting these research needs will be summarized. Particular topics addressed include:

- Development of advanced techniques for monitoring fermentation of insoluble cellulose, including inferring the cellulose concentration based on on-line measurements;
- Development of a quantitative assay for measuring and comparing rates of microbial cellulose utilization, functionally similar to the filter paper assay used for enzymatic studies;
- Development of a quantitative proteomic assay for the total cellulase concentration, including free cellulase

enzymes as well as cell-associated and cellulose-associated enzymes; and

- Documentation and investigation of an apparent “fast growing” phenotypic variant of *Clostridium thermocellum* able to utilize cellulose substantially more quickly than strains described in the literature.

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.

GTL

## Metagenomics for Mining New Deconstructive Enzymes, Exploring Enzyme Diversity and Screening Cellulolytic Activities

Luen-Luen Li,<sup>1</sup> Sean M. McCorkle,<sup>1</sup> Denise C. Monteleone,<sup>1</sup> Susannah G. Tringe,<sup>2</sup> Tanja Woyke,<sup>2</sup> Shi-You Ding,<sup>3</sup> Michael Himmel,<sup>3</sup> Safiyh Taghavi,<sup>1</sup> Carl Abulencia,<sup>4</sup> Deborah Balch,<sup>4</sup> Ying Hefner,<sup>4</sup> Melisa Low,<sup>4</sup> Steven Truong,<sup>4</sup> Peter Luginbühl,<sup>4</sup> Steve Wells,<sup>4</sup> Joel Kreps,<sup>4</sup> Kevin Gray,<sup>4</sup> Daniel van der Lelie<sup>1\*</sup> (vdlelie@bnl.gov), and **Martin Keller**<sup>5</sup>

<sup>1</sup>Brookhaven National Laboratory, Upton, N.Y.; <sup>2</sup>DOE Joint Genome Institute, Walnut Creek, Calif.; <sup>3</sup>National Renewable Energy Laboratory, Golden, Colo.; <sup>4</sup>Verenium Corporation, San Diego, Calif.; and <sup>5</sup>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 multitolerant 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 (e.g. *Clostridium thermocellum*) to understand novel strategies and enzyme complexes for biomass deconstruction. We are also searching natural diversity for enzymes and microbes—in particular, thermophilic anaerobes. BESC is studying and modeling the structure and activities of multi-enzyme complexes to design or screen for variants with better cell wall deconstruction capabilities.**

Plant biomass is the most abundant biopolymer on earth and has long been recognized as a potential sustainable source of mixed sugars for bioenergy production. Our goals are to understand the diversity, structure, functional interdependence, and metabolic capabilities of the natural cellulolytic microbial assemblages, and to exploit their dynamics for the conversion of plant biomass to a useful feedstock for biofuels production. Using metagenomics as an approach

allows the discovery of new enzyme diversity from microbial communities, especially from organisms that are unknown or have never been cultivated. From a microbial community actively decaying poplar biomass under anaerobic conditions, metagenomic DNA was isolated for further investigation. The distribution of microbial species in the community was investigated via 16S and 18S rRNA genes sequencing. *Saccharomyces* composed the major group among the Eukaryotes, and *Clostridiales* composed the major group among the Bacteria. No major population of Archaea was found as part of the microbial community. Using the 454 GS FLX Titanium pyrosequencing, approximately 580 Mbp metagenomic DNA was sequenced. Preliminary homology searches of metagenome sequences revealed a high diversity of glycosyl hydrolase homologs (approximately 4,000 glycosyl hydrolases were identified). Five candidate glycosyl hydrolases were initially selected for further investigation, based on homology to enzyme families of interest (GHase families 5, 9, 48, and 51 representing cellulase, hemicellulase and xylanase activities) and the quality of the sequences (length, homology, potential gene rearrangements, disruptions, deletions). Full-length open reading frames of these genes were obtained by using inverse PCR and DNA walking, and gene cloning is presently in the process. Another approach to discover new glycosyl hydrolases is by constructing lambda-based expression libraries and screening clones for glycosyl hydrolase activity. Libraries were constructed from a variety of likely cellulolytic environments such as the digestive tract of herbivorous mammals and insects, microbial 'biotraps' and the gills of marine shipworms. These libraries will be screened using Verenum's ultra high-throughput GigaMatrix® system that can screen up to 1 billion samples per day. In a preliminary screen, 10 million clones were screened and 353 primary hits identified from two environmental libraries. The number of active clones reduced to 61 after the tertiary screen and based on DNA sequence data, 29 unique, active enzymes were identified, 14 of which have known GH domains. Additional activity screening will be performed on these libraries, as well as large-scale sequencing at JGI of the original environmental DNAs. Our combined metagenomic studies and enzyme activity screens will provide insight into the microbial community compositions as well as provide a resource for discovering diverse, novel, community-encoded glycosyl hydrolases.

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

## The Improved Cellulosome: Computational Modeling to Minisomes

Yannick J. Bomble<sup>1\*</sup> (yannick\_bomble@nrel.gov), Michael F. Crowley,<sup>1</sup> Shi-you Ding,<sup>1</sup> Qi Xu,<sup>1</sup> Mark R. Nimlos,<sup>2</sup> Jiancong Xu,<sup>3</sup> Moumita Saharay,<sup>3</sup> Hong Guo,<sup>4</sup> John W. Brady,<sup>5</sup> David B. Wilson,<sup>6</sup> Jeremy C. Smith,<sup>3</sup> Michael E. Himmel,<sup>1</sup> and **Martin Keller**<sup>7</sup>

<sup>1</sup>Chemical and Biosciences Center and <sup>2</sup>National Bioenergy Center, National Renewable Energy Laboratory, Golden, Colo.; <sup>3</sup>Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, Tenn.; <sup>4</sup>Dept. of Biochemistry and Cell and Molecular Biology, University of Tennessee, Knoxville, Tenn.; <sup>5</sup>Dept. of Food Science and <sup>6</sup>Section of the Biochemistry, Molecular, and Cell Biology, Cornell University, Ithaca, N.Y.; and <sup>7</sup>Biosciences Division, 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 multitolerant 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 (e.g. *Clostridium thermocellum*) to understand novel strategies and enzyme complexes for biomass deconstruction. We are also searching natural diversity for enzymes and microbes—in particular, thermophilic anaerobes. BESC is studying and modeling the structure and activities of multi-enzyme complexes to design or screen for variants with better cell wall deconstruction capabilities.**

This study aims at understanding the mechanisms involved in the sequential binding of the cellulosomal enzymes on the scaffold of *C. thermocellum*. The first coarse grained model to study the formation and function of the cellulosome was developed within CHARMM. Some of the binding constants between cohesins and dockerins were derived from all-atom simulations. Individual subdomains were also studied with CHARMM and Amber on cellulose surfaces or with individual cellulose chains. These domains include catalytic domains, carbohydrate binding domains, and fibronectins. All five cellulosomal fibronectins (Fn3) of *C. thermocellum* have been identified, overexpressed and purified, the crystal structure of one of them has been solved, and this provided experimental structure for the computational modeling of cellulosomal Fn3 function.

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

## A Rapid Analytical Pyrolysis Method for Investigating Genetic Modification of the Lignin Pathway in Alfalfa (*Medicago sativa*)

A. Ziebell<sup>1\*</sup> (angela\_ziebell@nrel.gov), R. Sykes,<sup>1</sup> D. Astling,<sup>1</sup> K. Reichel,<sup>1</sup> C. Doeppke,<sup>1</sup> M. Davis,<sup>1</sup> and **Martin Keller<sup>2</sup>**

<sup>1</sup>National Renewable Energy Laboratory, Golden, Colo.; and <sup>2</sup>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 multitolerant microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC researchers in characterization, modeling, education and data management seek to provide integrating data and understanding across the center. We use advanced technologies to analyze chemical and structural changes within biomass. We have developed a coordinated characterization pipeline to provide composition, pretreatment, and enzyme digestibility data from native and modified plant samples and to store this data within a data management system. We are developing knowledge environments to allow access to improved data analysis and modeling tools. Knowledge gained by thoroughly characterizing biomass chemistry and structure will drive coordinated development of improved plant biomass and degradation microbes. Native, genetically modified, and partially deconstructed lignocellulosic samples will be analyzed.**

Pyrolysis-Molecular Beam Mass Spectrometry (py-MBMS) and pyrolysis-Gas Chromatography Mass Spectroscopy (py-GCMS) were used to study changes in lignin which occur on down regulation of two genes which form part of the biosynthetic route to S and G lignin. Antisense down regulated p-coumarate 3-hydroxylase (C3H) and hydroxycinnamoyl transferase (HCT) mutants of Alfalfa (*Medicago sativa*) were used to investigate the impact of changes in total lignin content, as well as H, G and S lignin ratios. Both genes are involved in the S and G but not H biosynthetic pathways, down regulation, therefore, results in reduced production of S and G and an increased proportion of H lignin. The change in lignin content is assessed by measuring the mass spectrum of a pyrolyzed sample. Distinctive peaks are known to come from certain compounds in biomass allowing changes in those peaks to be attributed to changes in the compounds.

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

## Integrated High Throughput Pretreatment and Enzymatic Hydrolysis in 96 Well Plates

Michael H. Studer<sup>1\*</sup> (michael.studer@ucr.edu), Jaclyn DeMartini,<sup>1</sup> Heather McKenzie,<sup>1</sup> Charles E. Wyman,<sup>1</sup> and **Martin Keller<sup>2</sup>**

<sup>1</sup>Dept. of Chemical and Environmental Engineering, Center for Environmental Research and Technology, University of California, Riverside, Calif.; and <sup>2</sup>Biosciences Division, 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 multitolerant microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC researchers in characterization, modeling, education and data management seek to provide integrating data and understanding across the center. We use advanced technologies to analyze chemical and structural changes within biomass. We have developed a coordinated characterization pipeline to provide composition, pretreatment, and enzyme digestibility data from native and modified plant samples and to store this data within a data management system. We are developing knowledge environments to allow access to improved data analysis and modeling tools. Knowledge gained by thoroughly characterizing biomass chemistry and structure will drive coordinated development of improved plant biomass and degradation microbes. Native, genetically modified, and partially deconstructed lignocellulosic samples will be analyzed.**

One of the aims of the BioEnergy Science Center (BESC) is the engineering of advanced plants with reduced recalcitrance for sugar release. As this requires screening of many different natural and genetically modified biomass types to identify those with a lower recalcitrance for sugar release, a new high-throughput (HTP) tool integrating pretreatment and enzymatic hydrolysis in the same multi-well configuration was developed. Water only or dilute acid pretreatment of 1% biomass slurries is performed in a custom-made 300 $\mu$ L 96 well plate made of metal to withstand heating to temperatures up to 180°C in a steam chamber as well as prevent corrosion. Furthermore, our so-called co-hydrolysis approach adds citric acid buffer, sodium azide, and enzymes (e.g., cellulase and xylanase) directly to each well without separating the solid and liquid after pretreatment, with enzyme loadings based on the original glucan and xylan content of the raw biomass. Next, the plate is incubated at 50°C for 72 h, and the release of sugars is quantified by HPLC. To prove the feasibility of this concept, performance for the co-hydrolysis process was compared to that for conventional washed solids hydrolysis, and about 100 mg

of enzyme protein/g glucan and xylan in the raw biomass was shown to overcome inhibition by products formed or released during pretreatment and result in similar yields. Operational testing demonstrated that the custom-made well-plates did not leak during pretreatment, and heat-up and cool-down required less than 45 s at an operating temperature of 180°C. The standard deviation in total sugar yields was only 4.1% across the 96 wells for combined pretreatment and co-hydrolysis of poplar, and yields for co-hydrolysis using the multi-well system were virtually identical to those for co-hydrolysis with standard tube reactors as well as washed solids hydrolysis using these standard vessels.

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.

GTL

## Advanced Imaging Projects in the BioEnergy Science Center (BESC)

Shi-You Ding<sup>1\*</sup> (shi\_you\_ding@nrel.gov), Yining Zeng,<sup>1</sup> Thomas Haas,<sup>1</sup> Yu-San Liu,<sup>1</sup> Mike Himmel,<sup>1</sup> X. Sunney Xie<sup>2</sup> (xie@chemistry.harvard.edu), and **Martin Keller**<sup>3</sup>

<sup>1</sup>Chemical and Biosciences Center, National Renewable Energy Laboratory, Golden, Colo.; <sup>2</sup>Dept. of Chemistry and Chemical Biology, Harvard University, Cambridge, Mass.; and <sup>3</sup>Biosciences Division, 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 multitolerant microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC researchers in characterization, modeling, education and data management seek to provide integrating data and understanding across the center. We use advanced technologies to analyze chemical and structural changes within biomass. We have developed a coordinated characterization pipeline to provide composition, pretreatment, and enzyme digestibility data from native and modified plant samples and to store this data within a data management system. We are developing knowledge environments to allow access to improved data analysis and modeling tools. Knowledge gained by thoroughly characterizing biomass chemistry and structure will drive coordinated development of improved plant biomass and degradation microbes. Native, genetically modified, and partially deconstructed lignocellulosic samples will be analyzed.**

Deeper understanding of biomass structure and processes for its conversion to sugars and other useful products is

required to enable economic lignocellulosic biofuels production. The BioEnergy Science Center (BESC) has targeted these objectives for critical investigation. Imaging techniques are used to study the cell walls from wild type, genetically modified, and chemically and/or biologically treated switchgrass and poplar to identify the molecular characteristics that govern the emergent property of recalcitrance. Our projects focus on four research areas:

*Sub-Nanometer Scale Imaging of the Plant Cell Walls:* The structural networks in higher plant cell walls are composed of polysaccharides, pectins, lignins and glycoproteins, which are evolved to be recalcitrant composites that are difficult to deconstruct to fermentable sugars. It remains a challenge; however, to characterize plant cell walls at the molecular scale. New imaging tools based on scanning probe microscopy (SPM) are capable of characterizing the plant cell wall at the nanometer scales. Specifically, atomic force microscopy (AFM) combined with near-field optics is now used to image cell walls with considerable success.

*High Resolution Chemical Imaging of Biomass:* It is important to identify the chemistry and kinetics of the biomass conversion processes *in vivo*. Non-linear optical microscopy has been used to map plant cell wall chemistry in the BESC. Specifically, coherent anti-Stokes Raman scattering (CARS), stimulated Raman scattering (SRS) (see also Poster by Saar *et al.*), and second (SHG) and third (THG) harmonic generation are currently used to enable new imaging tools for the characterization of biomass samples.

*Microbe/Enzyme/Biomass Interaction:* In nature, biomass degradation is a process of molecular interaction and reaction between plant cell wall polymers (i.e., cellulose and matrix polymers) and cellulolytic microbes and their secreted enzymes. An integrated system has been set up to combine microscopic and spectroscopic modules that allow us to characterize biomass conversion processes at high spatial and chemical resolution. For example, AFM is used to map the surface topography of the plant cell wall and the binding of microbial cells and enzymes to the walls; total internal reflection fluorescence (TIRF) microscopy with fluorescent probe labeling techniques is used to track the distributions and movements of labeled microbial cells and enzymes; and spectroscopy is used to monitor the resultant chemical changes in cell wall polymeric component during biochemical, as well as chemical conversions of biomass.

*Single Enzyme Molecule Spectroscopy:* It is believed that efficient deconstruction of complex plant cell walls requires the synergistic reaction of many polysaccharide-degrading enzymes. In some cases multiple enzyme complexes (i.e., cellulosomes) perform the efficient hydrolysis of the plant cell walls. It is not well known; however, the ways that these enzyme complexes are assembled and made functional. We have shown that single molecule approaches are capable of tracking the behaviors of individual enzyme and binding domains known to be critical for the cellulase function. At NREL, various carbohydrate-binding modules and enzymes have been tagged with fluorescent probes so that TIRF and non-linear optics can be used to image these molecules at the single molecular level.

We note that integrated imaging modalities are required to characterize biomass, both in the wild type form and during the conversion process. The knowledge derived from the use of these advanced characterization tools could be facilitated by combining our ability to image cell walls at the meso- and micro- scales, i.e., by the simultaneous imaging of single molecule features upon a complex chemical background. Example images of plant tissues, cellulose microfibrils, microbial cells, enzyme complexes, and single enzymes will be presented and discussed.

## References

1. Ding S-Y and Himmel ME, (2006) The Maize Primary Cell Wall Microfibril: A New Model Derived from Direct Visualization. *J. Agric. Food Chem.* 54, 597-606.
2. Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, and Foust TD (2007) Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. *Science* 315, 804-807.
3. Ding S-Y, Himmel ME and Xie SX, (2008) Identify Molecular Structural Features of Biomass Recalcitrance Using Nondestructive Microscopy and Spectroscopy. *Microscopy and Microanalysis*, Vol. 14, Suppl. S2, pp 1494-1495 doi: 10.1017/S1431927608088582.
4. Ding S-Y, Xu Q, Crowley M, Zeng Y, Nimlos M, Lamed R, Bayer EA and Himmel ME, (2008). A biophysical perspective on the cellulosome: new opportunities for biomass conversion *Curr. Opin. Biotechnol.* 19:218-227.
5. Freudiger CW, Min W, Saar BG, Lu S, Holtom GR, He H, Tsai JC, Kang JX, Xie SX, (2008). Label-free Biomedical Imaging with high sensitivity by stimulated Raman scattering microscopy, *Science*, in press.
6. Xu Q, Tucker MP, Arenkiel P, Ai X, Rumbles G, Sugiyama J, Himmel ME, and Ding S-Y, (2008) Labeling the planar face of crystalline cellulose using quantum dots directed by type-I carbohydrate-binding modules. *Cellulose*. DOI 10.1007/s10570-008-9234-4
7. Ding S-Y, Xu Q, Ali MK, Baker JO, Bayer EA, Barak Y, Lamed R, Sugiyama J, Rumbles G, and Himmel ME, (2006). Versatile derivatives of carbohydrate-binding modules for imaging of complex carbohydrates approaching the molecular level of resolution. *Biotechniques* 41, 435-443.

This work is funded by the US Department of Energy, the Office of Science, Biological and Environmental Research (BER) program under two projects, DOE Genomics:GTL BioEnergy Science Center (BESC) and "Study of Lignocellulosic Material Degradation with CARS Microscopy" Grant DE-FG02-07ER64500.

## Analytical BESC Advances in Characterization of Biomass and Recalcitrance

Y. Pu<sup>1\*</sup> (Yunqiao.Pu@ipst.gatech.edu), N. Jiang,<sup>1</sup> H. Li,<sup>1</sup> M. Foston,<sup>1</sup> R. Samuel,<sup>1</sup> J. Seokwon,<sup>1</sup> A.J. Ragauskas,<sup>1</sup> and **Martin Keller**<sup>2</sup>

<sup>1</sup>School of Chemistry and Biochemistry, Institute of Paper Science and Technology, Georgia Institute of Technology, Atlanta, Ga.; and <sup>2</sup>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, education and data management seek to provide integrating data and understanding across the center. We use advanced technologies to analyze chemical and structural changes within biomass. We have developed a coordinated characterization pipeline to provide composition, pretreatment, and enzyme digestibility data from native and modified plant samples and to store this data within a data management system. We are developing knowledge environments to allow access to improved data analysis and modeling tools. Knowledge gained by thoroughly characterizing biomass chemistry and structure will drive coordinated development of improved plant biomass and degradation microbes. Native, genetically modified, and partially deconstructed lignocellulosic samples will be analyzed.**

The design of low recalcitrance biomass for deconstruction to bioethanol and related biofuels is predicated on understanding the fundamental relationship between plant polysaccharides, lignin and how these biopolymers are integrated in the plant cell wall. This poster examines how advanced 1D and 2D NMR techniques can be utilized to identify structural elements of importance to the recalcitrance of genetically engineered alfalfa. In addition, on going studies utilizing deuteronic ionic liquid solvents with NMR and MALDI techniques are being developed such that rapid whole plant cell wall structures can be determined without the need for laborious isolation and purification techniques. We will also report crystallinity and other measurements as part of understanding the structure and chemistry of these samples. These advances will provide researchers new analytical tools to address the challenge of high-throughput/high resolution analysis of biomass polymers generated in state-of-the-art plant genomics programs.

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.

## Computer Simulation of Lignocellulosic Biomass

Loukas Petridis<sup>1\*</sup> (petridisl@ornl.gov), Benjamin Lindner,<sup>1</sup> Jeremy C. Smith,<sup>1</sup> and **Martin Keller**<sup>2</sup>

<sup>1</sup>Center for Molecular Biophysics and <sup>2</sup>Biosciences Division, 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 multitolerant microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC researchers in characterization, modeling, education and data management seek to provide integrating data and understanding across the center. We use advanced technologies to analyze chemical and structural changes within biomass. We have developed a coordinated characterization pipeline to provide composition, pretreatment, and enzyme digestibility data from native and modified plant samples and to store this data within a data management system. We are developing knowledge environments to allow access to improved data analysis and modeling tools. Combining experimental data with biological models will help define the genomic and physical basis of plant cell-wall recalcitrance and deconstruction.**

Plant cell wall lignocellulosic biomass is a complex material composed of crystalline cellulose microfibrils laminated with hemicellulose, pectin, and lignin polymers. The aim of this project is to use computer simulation to complement experiments in understanding the physical properties of this biomass. Atomistic models of softwood lignin and cellulose have been created and combined to form an initial model of lignocellulose. The models were built based on input from experiments on lignin and cellulose composition from the characterization group of BioEnergy Science Center. Molecular dynamics simulation of the models is performed to examine for structural properties of the system.

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

## The Plant Cell Wall Biosynthesis Related Galacturonosyltransferase (GAUT) and GAUT-Like (GATL) Genes Have a Different Origin than the Other Glycosyltransferase Family 8 Genes

Yanbin Yin<sup>1,3\*</sup> (yinyb@csbl.bmb.uga.edu), Hailing Chen,<sup>1</sup> Michael Hahn,<sup>2,3</sup> Debra Mohnen,<sup>2,3</sup> Ying Xu,<sup>1,3</sup> and **Martin Keller**<sup>3</sup>

<sup>1</sup>Computational System Biology Laboratory, Dept. of Biochemistry and Molecular Biology, and Institute of Bioinformatics, and <sup>2</sup>Complex Carbohydrate Research Center, University of Georgia, Athens, Ga.; and <sup>3</sup>BioEnergy Science Center, Biosciences Division, 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 multitolerant microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC researchers in characterization, modeling, education and data management seek to provide integrating data and understanding across the center. We use advanced technologies to analyze chemical and structural changes within biomass. We have developed a coordinated characterization pipeline to provide composition, pretreatment, and enzyme digestibility data from native and modified plant samples and to store this data within a data management system. We are developing knowledge environments to allow access to improved data analysis and modeling tools. Combining experimental data with biological models will help define the genomic and physical basis of plant cell-wall recalcitrance and deconstruction.**

The galacturonosyltransferases (GAUT) and GAUT like (GATL) genes, belonging to the glycosyltransferase family 8 (GT8), have been indicated to be involved in the plant cell wall biosynthesis (pectin and xylan synthesis) (see poster by Z. Hao et al.). In this study, we have identified, classified and evolutionarily studied the GT8 family in 13 fully sequenced plant and green algal genomes as well as in the NCBI non redundant protein database (NCBI-nr DB). We found that there are three major GT8 protein classes in nature. The GAUTs and GATLs are in class I, while galactinol synthases (GolSs) and plant glycogenin-like starch initiation proteins (PGSIPs) are in class II. The class III includes almost all bacterial sequences. A moss-specific subfamily, a metazoan subfamily, some algal sequences and a sequence from the cyanobacteria *Synechococcus elongatus* PCC 7942 are found within class I too. Some sequence motifs such as the DXD and HXXGXXKPW are conserved across all GT8

subfamilies, while others are specific to certain subfamilies. We conclude that the plant cell wall synthesis related GAUTs and GATLs have a different origin than the other plant GT8 genes, possibly anciently acquired from some certain cyanobacteria.

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.

GTL

## Computational Prediction of Golgi Resident Proteins in *Arabidopsis thaliana*

Wen-Chi Chou<sup>1,3\*</sup> (wcc@uga.edu), Maor Bar-Peled,<sup>2,3</sup> Ying Xu,<sup>1,3</sup> and **Martin Keller**<sup>3</sup>

<sup>1</sup>Computational Systems Biology Laboratory, Dept. of Biochemistry and Molecular Biology, and Institute of Bioinformatics and <sup>2</sup>Center for Complex Carbohydrate Research and Dept. of Plant Biology, University of Georgia, Athens, Ga.; and <sup>3</sup>DOE 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 multitasking microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC researchers in characterization, modeling, education and data management seek to provide integrating data and understanding across the center. We use advanced technologies to analyze chemical and structural changes within biomass. We have developed a coordinated characterization pipeline to provide composition, pretreatment, and enzyme digestibility data from native and modified plant samples and to store this data within a data management system. We are developing knowledge environments to allow access to improved data analysis and modeling tools. Combining experimental data with biological models will help define the genomic and physical basis of plant cell-wall recalcitrance and deconstruction.**

Proteins residing in plant Golgi play key functional roles in cell wall synthesis and protein targeting. It will provide a much improved understanding about the detailed mechanism of cell wall syntheses if we know which functional proteins reside in Golgi. We present here a computational method for identification of the Golgi resident proteins in *Arabidopsis thaliana*. We have compiled a list of Golgi residing proteins from the published literature, and studied a number of sequence and structural features of these proteins, potentially useful in distinguishing between Golgi resident and non-Golgi proteins. Among these features, we found

that transmembrane domains and a few functional signatures associated with known Golgi resident proteins are particularly useful. Based on these features, we have developed a classification method based on a support vector machine, for predicting Golgi resident proteins. Our prediction program achieved 91% and 93% prediction sensitivity and specificity, respectively, on a test set consisting of 1,434 proteins of *Arabidopsis thaliana*, substantially better than any of the existing prediction programs. Using this program, we have predicted around 2700 Golgi resident proteins in *Arabidopsis thaliana*, providing a useful data source to other researchers.

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.

GTL

## The BioEnergy Science Center Laboratory Information Management System (LIMS)

Susan K. Holladay\* (holladaysk@ornl.gov), Denise D. Schmoyer, Sheryl A. Martin, Edward C. Uberbacher, and **Martin Keller**

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 multitasking microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC researchers in characterization, modeling, education and data management seek to provide integrating data and understanding across the center. We use advanced technologies to analyze chemical and structural changes within biomass. We have developed a coordinated characterization pipeline to provide composition, pretreatment, and enzyme digestibility data from native and modified plant samples and to store this data within a data management system. We are developing knowledge environments to allow access to improved data analysis and modeling tools. Combining experimental data with biological models will help define the genomic and physical basis of plant cell-wall recalcitrance and deconstruction. Laboratory Information Management Systems and knowledgebase environments have separate key roles. The LIMS is the central data repository for all information related to processing and analysis of the BioEnergy Science Center (BESC) samples. It maintains a detailed history for each sample by capturing processing parameters, protocols, statistical data, stocks, tests and analytical results for the complete life cycle of the sample. It is also used to track protocols so that data can be linked to its source and history. The LIMS is also linked to the BESC knowledgebase which**

**contains the results of experiments interpreted in the context of biological systems.**

A laboratory information management system has been implemented using the Nautilus commercial package and modified and configured for BESC operations. The major data generation steps used by BESC experimental campaigns in Year 1 have been coded into this LIMS system. These include data and metadata from various instruments and processes covering sample acquisition, biomass grinding, biomass pretreatment, compositional analysis before and after cellulolytic processing, the analysis of complex environmental community microbial samples, and sequence analysis using 454. The LIMS has been used to capture results and process for the eight campaigns undertaken by BESC in Year 1. For each of these campaigns a detailed workflow, including sample generation, sample splitting and shipping, sample processing, protocols, controls, replicates, and metadata and results have been captured. This has involved extensive collaboration and communication between the LIMS team and experimental labs. The tracking of materials, samples and processes is central to the proper functioning of BESC both from the standpoint of ensuring meaningful and reproducible results, but also to track the generation of intellectual property. All materials and samples shipped between institutions are tracked at the level of the institution using a material transfer information system attached to the LIMS. The LIMS has been used to capture large amounts of primary data including over 800 results from compositional analysis studies. The LIMS team has provided initial documentation on LIMS use for sample tracking on the Wiki site. Additionally several seminars about LIMS use have been provided with additional training workshops to be held in Year 2. Other outreach mechanisms including mailings and helpline have been provided. An ORACLE database has been designed and implemented to support LIMS operations. It stores data, metadata and links to data files from BESC experiments and procedures. The database server has been set up in a special area of the ORNL network outside the firewall (Open Research Enclave), which also houses the LIMS system and interfaces. Processes have been established for user account and password generation that ensure proper data security. Using the database and available LIMS interface, users can access data about samples or analysis results from any campaign.

A standardized approach to bar coding has been implemented in BESC, with bar codes generated by the LIMS system. Bar codes are used to initiate (generate) samples and also when samples are sub-sampled, or used in multi-well plate operations as specified in each workflow. In addition to unique IDs contained in the LIMS, human readable fields are available on the barcodes. Major labs in BESC have all purchased the same barcode reader purchased as a result of substantial testing by NREL. A series of reports is now available for BESC members including reports for material transfer, shipping, sample history and provenance, campaign results.

This submission has been authored by a contractor of the U.S. Government under contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a nonexclusive, royalty-free

license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes. 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.

GTL

**The BESC Knowledgebase: An Infrastructure for Biological Discovery**

Morey Parang<sup>1,2\*</sup> (parangm@ornl.gov), Tatiana Karpinets,<sup>1,2</sup> Doug Hyatt,<sup>1,2</sup> Mike Leuze,<sup>1</sup> Honney Park,<sup>1</sup> Bob Cottingham,<sup>1</sup> Michael Galloway,<sup>1</sup> Loren Hauser,<sup>1</sup> Mustafa Syed,<sup>1</sup> Ed Uberbacher,<sup>1</sup> and **Martin Keller**<sup>1</sup>

<sup>1</sup>Oak Ridge National Laboratory, Oak Ridge, Tenn.; and <sup>2</sup>University of Tennessee, Knoxville, Tenn.

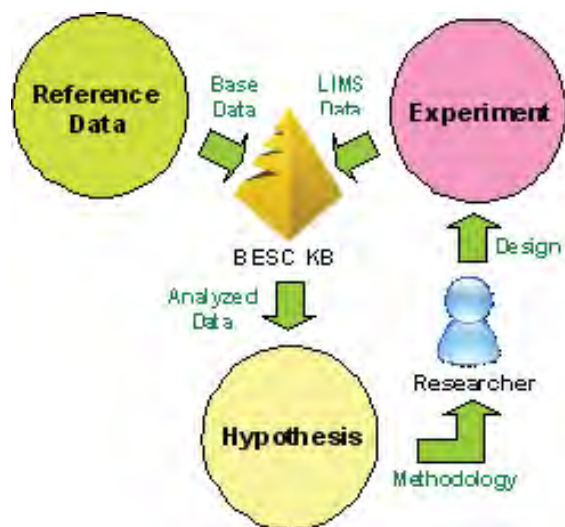
<http://bioenergycenter.org>

**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 multi-talented microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). BESC researchers in characterization, modeling, education and data management seek to provide integrating data and understanding across the center. We use advanced technologies to analyze chemical and structural changes within biomass. We have developed a coordinated characterization pipeline to provide composition, pretreatment, and enzyme digestibility data from native and modified plant samples and to store this data within a data management system. We are developing knowledge environments to allow access to improved data analysis and modeling tools. Combining experimental data with biological models will help define the genomic and physical basis of plant cell-wall recalcitrance and deconstruction. Laboratory Information Management Systems and knowledgebase environments have separate key roles.**

The BioEnergy Science Center (BESC) is a multi-institutional center which is undertaking large experimental campaigns to understand and reduce biomass recalcitrance. As such, the center is expected to generate large volumes of diverse data including genome sequences, omics data, protein structures, images, mass spectrometry as well as NMR spectra, and various assay results. The BESC knowledgebase (KB), currently under development, provides a novel means for creative exploration and sharing of the data, materials, and experimental processes across the distributed enterprise. The BESC KB is an infrastructure designed specifically for providing "knowledge" of BESC-related research. It serves as a participatory biological discovery platform for the BESC community. The BESC KB links and correlates the results of analyses from BESC analysis processes to



genomes, pathways, regulatory knowledge and molecules from plants and microbes. The KB platform provides an array of applications and tools to extract and combine knowledge from isolated data, helping the BESC researchers in interpretation and understand of the experimental results and to form rapid hypotheses (Figure A).



**Figure A.** BESC Knowledgebase Concept and Rational



**Figure B.** BESC Knowledgebase BeoCyc

An ultimate goal of the BESC knowledgebase is to develop an infrastructure that (i) provides comprehensive information on metabolic processes in the organisms related to production of biofuel, (ii) facilitates analysis and querying of the information in a cell systems context and (iii) overlays the information with the collected experimental data and other information in the BESC knowledgebase.

The knowledgebase addresses the main BESC focus areas through development of five “knowledge” domains: (i) Plant Genomes and Cell Wall Biosyntheses; (ii) Cellulolytic Microbes; (iii) Biomass Characterization Data; (iv) Enzymes and Cellulosomes; (v) Consolidated Bioprocessing. Currently only the first two domain areas have been

implemented. The Plant Genomes and Cell Wall Biosyntheses component primarily targets the Biomass Formation and Modification focus area knowledge requirements. This domain contains reference data for twenty one plant genomes along with a diverse set of bioinformatic tools and applications. These tools include annotation pipelines, manual curation of gene models, visualization tools, and search engines.

The Cellulolytic Microbes domain targets the Characterization and Modeling focus area, which includes the BeoCyc implementation. BeoCyc is an initial element of the infrastructure (Figure B) representing a collection of Pathway/Genome Databases (PGDBs) in a set BioEnergy related microOrganisms (Beo). The PGDBs were generated automatically using the Pathway Tools software and include annotation of genes in the microorganisms with metabolic pathways from the MetaCyc database. BeoCyc provides a web interface to the PGDBs and to diverse computational, visualization and analytical tools based on the generated databases, which include searching of the databases, browsing of the microbial genomes and generated annotations, downloading the annotations, overview of the metabolic pathways and their overlaying with the experimental data. At present BeoCyc includes PGDB reconstructions for 14 microbial organisms involved in degradation of cellulose and fuel productions and for some bacteria that are specific targets of BESC investigators.

The submitted abstract has been authored by a contractor of the U.S. Government under contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes. 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.

GTL

## BioEnergy Science Center Education and Outreach

P. Doney<sup>1,3\*</sup> (patdoney@uga.edu), S. Fowler,<sup>2,3</sup> B. Davison,<sup>2,3</sup> **M. Keller**,<sup>2,3</sup> J. Westpheling,<sup>1,3</sup> S. Kral,<sup>4</sup> J. Griffin,<sup>4</sup> and H. Schulson<sup>4</sup>

<sup>1</sup>University of Georgia, Athens, Ga.; <sup>2</sup>Oak Ridge National Laboratory and <sup>3</sup>BESC, Oak Ridge, Tenn.; and <sup>4</sup>The Creative Discovery Museum, Chattanooga, 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 multitasking microbes for converting plant biomass into biofuels in a single step (consolidated bioprocessing). Education and outreach to**

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

In addition to our efforts to prepare a new generation of scientists for the emerging fields of bioenergy through the interdisciplinary training of graduate students and postdocs, our center has taken a novel approach in that our education efforts begin with fifth graders. We have developed lesson plans aimed at 4<sup>th</sup>, 5<sup>th</sup> and 6<sup>th</sup> 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 biobased fuel economy. The hands-on activities and guided questions are also designed to meet educational objectives for these grades. These lessons have been piloted in a hundred classrooms in North Georgia and Tennessee and will be made available to schools nationwide in the fall of 2009. We have also begun to pilot interactive “science night” programs offered to students and the general public through local schools, museums and community centers. We will present details of some of the lessons and science night activities.

The BioEnergy Science Center is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

## USDA–DOE Plant Feedstock Genomics for Bioenergy

USDA-DOE

### Towards a Map of the *Populus* Biomass Protein-Protein Interaction Network

Eric Beers<sup>1\*</sup> (ebeers@vt.edu) Amy Brunner,<sup>2</sup> Allan Dickerman,<sup>3</sup> Xiaoyan Sheng,<sup>2</sup> Chengsong Zhao,<sup>1</sup> and Mingzhe Zhao<sup>1</sup>

<sup>1</sup>Dept. of Horticulture, <sup>2</sup>Dept. of Forestry, and <sup>3</sup>Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Va.  
<http://xylome.vbi.vt.edu/>

**Project Goals:** 1) Clone approximately 400 Gateway-compatible ORFs corresponding to the poplar xylem gene set (PxORF); 2) Identify poplar proteins that co-purify with selected TAP-tagged PxORF proteins expressed in poplar; 3) Identify high-confidence Y2H interactions for a subset of approximately 60 PxORFs comprising putative regulators of lignocellulose synthesis screened against a poplar xylem cDNA library; 4) Identify Y2H interactions resulting from a matrix of pair-wise assays between all

PxORF proteins; 5) Produce a protein-protein interaction map that incorporates interactions identified from the three screens; and 6) Maintain a web site to make results available and facilitate distribution of clones.

As one of the fastest-growing and most productive trees in North America, poplar is a model biomass crop for producing “green” energy such as power, heat, and biofuels. The lignocellulosic (woody) portion of the plant is a sink for CO<sub>2</sub>, and can be converted to ethanol to provide a renewable energy source. Therefore, a more detailed understanding of the molecular biology and genomics of wood formation in poplar trees is needed for the development of novel strategies, which will ultimately lead to the creation of low-cost and more efficient bioenergy feedstocks. The poplar genome is the first completely sequenced tree genome containing over 45,000 protein-coding genes. Our previous microarray data have shown that approximately 400 poplar genes are highly up-regulated in xylem. These genes have been designated as the poplar biomass gene set. In this project, our overall goal is to identify protein-protein interaction networks associated with biomass production in the woody tissues of poplar by performing yeast two-hybrid matrix assays using all possible pair-wise combinations from the biomass gene set tested in the same yeast cell for the strength of their interaction. Furthermore, selected genes will be tested in the yeast two-hybrid system for their abilities to interact with any protein found in a library of proteins derived from all the genes expressed in poplar wood-forming tissues. Finally, selected genes will be expressed in plants or *E. coli* as affinity-tagged proteins and tested for their abilities to form complexes with other proteins extracted from woody tissue. Members of protein complexes isolated from plant extracts will be identified using mass spectrometry. All of the interactions revealed as a result of these three types of protein-protein interaction studies will be combined and processed by bioinformatical programs to produce a biomass protein-protein interaction map. Significant progress has been made in constructing our poplar biomass ORF collection. To date, we have successfully cloned into Gateway entry vectors more than 120 open reading frames (ORFs) that encode members of the poplar biomass gene set. Sequencing of these biomass ORFs has revealed numerous splice variants and allelic sequences that diverge from the JGI model ORFs. Approximately 90 ORFs have been recombinationally cloned into yeast vectors as DNA-binding domain (DB) and activation domain (AD) fusions for two-hybrid assays. All DB fusions are tested for auto-activation of reporter genes without the presence of an AD fusion vector. Auto-activating DB fusions are eliminated from future yeast two-hybrid assays or re-cloned minus their putative activation domains. A pilot 30 x 30 yeast two-hybrid matrix experiment is being performed focusing on genes involved in secondary cell wall synthesis and signal transduction. Eleven genes have been affinity-tagged and transformed into poplar. Independent transgenic poplar lines have been tested by real-time RT-PCR and/or Western blotting to confirm the expression of affinity-tagged proteins.

USDA-DOE

## Development of Genomic and Genetic Tools for Foxtail Millet, and Use of These Tools in the Improvement of Biomass Production for Bioenergy Crops

**Jeffrey L. Bennetzen**<sup>1\*</sup> (maize@uga.edu), Katrien M. Devos<sup>2</sup> (kdevos@uga.edu), Andrew Doust<sup>3</sup> (andrew.doust@okstate.edu), Janice Zale<sup>4</sup> (janiczale@comcast.net), and Ryan Percifield<sup>1</sup>

<sup>1</sup>Dept. of Genetics, <sup>2</sup>Dept. of Crop and Soil Science, and Dept. of Plant Biology, University of Georgia, Athens, Ga.; <sup>3</sup>Botany Dept., Oklahoma State University, Stillwater, Okla.; and <sup>4</sup>Dept. of Plant Sciences, University of Tennessee, Knoxville, Tenn.

**Project Goals:** This project will generate a variety of genomic and genetic tools for foxtail millet, including SNPs, BAC libraries, optimized foxtail millet transformation technology, and a high density QTL and genetic map of foxtail millet for significant biomass traits. These resources will complement the DOE Joint Genome Institute whole genome sequencing of foxtail millet, enhancing its value as a functional genomic model for second generation bioenergy crops such as switchgrass.

Foxtail millet, *Setaria italica*, has the full set of attributes that will make it a model plant for basic and applied studies, particularly for its close relatives like switchgrass, an important bioenergy crop. Foxtail millet has small and simple chromosomes that will be sequenced in 2009 by the DOE's Joint Genome Institute (JGI). This sequence will enormously enhance further study and understanding of the genetic basis of biomass production, but it will only be fully useful if additional technologies are developed for use in foxtail millet. This proposal described experiments to improve the genetic and genomic toolkit for foxtail millet, and thus assist with the assembly of the foxtail millet genomic sequence, anchor the sequence to the genetic map and lay the foundation for future molecular, developmental, physiological and genetic studies.

The objectives and approaches in this proposal were 1) to construct a >12X redundant clone (BAC) library from foxtail millet inbred Yugu1 DNA, the DNA that will be sequenced by JGI, 2) to sequence the ends of the BACs so these data can help in the assembly and annotation of the complete foxtail millet genome sequence, 3) to confirm the assembly and detailed annotation, 4) to generate a new and large mapping population from the cross of Yugu1 foxtail millet with its wild progenitor, *Setaria viridis*, 5) to sequence the *S. viridis* genome with next-generation technology to find single nucleotide polymorphism (SNP) markers, 6) to construct a >1000 marker genetic map from the new crossing population using these SNPs, 7) mapping of tillering, shade response and biomass yield QTL in this new population, and 8) optimization of foxtail millet transformation

technology and its initial use to generate transgenics with altered expression of biomass candidate genes.

As of December 1 of this year, three months after grant activation, BAC library construction (obj. 1) has been completed and the BACs have been sent to JGI for end sequencing (obj. 2). The assembly and annotation of the complete foxtail millet genome sequence (obj. 3) will be in collaboration with the JGI, emphasizing the high throughput expertise of the JGI group and the manual confirmation skills of the UGA group. The generation of a new and large mapping population of a cross between foxtail millet and green millet (obj. 4) has begun, with an F<sub>2</sub> population now in hand. Further crosses are planned to increase the F<sub>2</sub> population and to try other green millet accessions as parents. To help integrate the first genetic map into the planned new map, 110 previous RFLP markers are being converted into PCR-based markers.

USDA-DOE

## Identifying Genes Controlling Feruloylation in Grass Cell Walls

**M.M. de O. Buanafina**<sup>\*</sup> (mmb26@psu.edu) and D. Braun (dda10@psu.edu)

Biology Dept., Pennsylvania State University, University Park, Pa.

**Project Goals:** The focus of this proposal is to identify and characterize new genes controlling feruloylation in grasses, as well as new genes that are responsible for the assembly of lignin into the cell wall and for biomass conversion. This will provide fundamental knowledge concerning the most crucial factors that influence grass cell wall degradability.

The major wall phenolics in grasses include lignin and hydroxycinnamic acids (HCA). HCAs constitute 1 to 2% of grass cell walls by dry weight and are important structural components as they are ester and ether linked to arabinoxylans. Ferulic acid (4-hydroxy-3-methoxy-cinnamic acid) is the major HCA identified in both primary and secondary cell walls of grasses. Ferulic acid residues attached to arabinoxylans have the ability to form ferulate dimers functioning in cell wall cross-linking. They are also proposed to act as nucleation sites for the formation of lignin and for the linkage of lignin to the xylan/cellulose network. Such coupling reactions, which occur predominantly in grasses, significantly decrease cell wall degradability and thus work as a barrier against efficient utilization of cell walls as a source of biomass for bioenergy production.

We have shown previously that the expression of ferulic acid esterase FAE in different grass species resulted in a substantial reduction in cell-wall-esterified ferulates and diferulates. FAE was shown to impact cell wall hydrolysis, resulting in increased yield of reducing sugars by cellulase treatment as

well as increased digestibility, reinforcing the importance of feruloylation and cross linking for cell wall degradability.

Controlling the level of total feruloylation should have a direct impact on the level of cross-linking and thereby on cell wall degradation. Currently, the genes underlying arabinoxylans feruloylation have not been identified and the isolation of such genes could be of great importance in manipulating ferulates accretion to the wall in grass species. Accordingly, we have targeted the feruloylation pathway as a means of increasing the efficiency of energy production from plant biomass. To identify these genes we are taking a forward genetic approach combined with a spectroscopic screen followed by detailed genetic and phenotype analyses. We have chosen *Brachypodium distachyon* as our model grass system because of its unique characteristics such as small genome size. *Brachypodium* seeds have been mutagenized with ethyl-methane sulfonate. We propose to identify *Brachypodium* mutants of interest by high-throughput spectroscopy-based leaf assay screening for change in the level of cell-wall esterified ferulic acid. Mutation of the feruloyl transferase gene(s) should lead to accumulation of soluble ferulates – as they will not be secreted to the cell wall – and less ferulate cross-linked to xylan.

If successful, the identification of these genes in *Brachypodium* can work as a handle for gene discovery in other important grass crop species because of conservation of genome organization and gene order in grasses.

USDA-DOE

## Developing Association Mapping in Polyploidy Perennial Biofuel Grasses: Establishing the Baseline on Genome-Size Variation

Denise E. Costich,<sup>1\*</sup> Bernd Friebe,<sup>2</sup> Michelle E. Denton,<sup>3</sup> Elhan Ersoz,<sup>3</sup> Michael D. Casler,<sup>4</sup> Jerome H. Cherney,<sup>5</sup> and **Edward S. Buckler**<sup>1</sup> (dc58@cornell.edu)

<sup>1</sup>USDA-ARS and Institute for Genomic Diversity, Cornell University, Ithaca, N.Y.; <sup>2</sup>Wheat Genetics and Genomic Resources Center, Kansas State University, Manhattan, Kans.; <sup>3</sup>Cornell University, Ithaca, N.Y.; <sup>4</sup>USDA-ARS, U.S. Dairy Forage Research Center, Madison, Wis.; and <sup>5</sup>Dept. of Crop & Soil Sciences, Cornell University, Ithaca, N.Y.

**Project Goals:** 1) Assemble association panels of diverse populations and linkage populations for switchgrass and reed canarygrass (~1000 clones each); 2) Trait evaluation for key biofeedstock characteristics in these panels; 3) Develop high density SNP markers in switchgrass; 4) Genotype association panels and linkage populations in switchgrass; 5) Evaluate population structure and germplasm diversity in switchgrass; and 6) Establish association mapping and estimate marker based breeding values in switchgrass.

Perennial grasses are considered to be primary candidates for biofuel production. However, many aspects of the basic biology of these species, including the number of sets of chromosomes in their genomes (ploidy), are still poorly understood. In fact, nineteen of the twenty-two grass species that were screened by the Oak Ridge National Laboratory's Herbaceous Energy Crops Program in the 1980s are polyploid, a phenomenon where genetically related individuals possess chromosome numbers that are multiples of each other. Furthermore, ten of these species have variable ploidy, where 2-6 different levels (4X-12X) have been observed. Two of these species, Switchgrass (*Panicum virgatum*) and Reed canarygrass (*Phalaris arundinacea*), are the study organisms for our joint linkage and association mapping project. In addition to ploidy-level variation, these species also show evidence of loss and gain of individual chromosomes (aneuploidy).

This variation in genome size will have major implications for genetic marker development and mapping. For example, chromosome-number variation, resulting in gene-copy number variation, could underlie a significant proportion of the phenotypic variation in the biofuel-related traits we will be measuring. We are determining genome size in two ways, via flow cytometry to measure nuclear DNA content and phase contrast microscopy to count chromosomes, in a representative sample of switchgrass plants that will become part of our permanent live germplasm collection. Establishing the correlation between these two measurements will allow us to use the faster method (flow cytometry) to phenotype thousands of plants in our field populations, reserving the more labor-intensive method (microscopy) for outliers. We present here our progress to date in this study and provide recommendations to the scientific community developing genetic resources for biofuel grasses.

USDA-DOE

## Computational Resources for Biofuel Feedstock Species

Kevin L. Childs\* (kchilds@plantbiology.msu.edu) and **C. Robin Buell** (buell@msu.edu)

Michigan State University, East Lansing, Mich.

<http://bfgr.plantbiology.msu.edu/>

**Project Goals:** To provide computational tools and resources for data-mining genomic based information available for biofuel feedstock species. Through the Biofuel Feedstock Genomics Resource (<http://bfgr.plantbiology.msu.edu/>), we will create a comprehensive, uniform, well annotated resource for data-mining genomic data for biofuel feedstock species.

While current production of ethanol as a biofuel relies on starch and sugar inputs, it is anticipated that sustainable production of ethanol for biofuel use will utilize lignocellulosic feedstocks. Candidate plant species to be used for lignocel-

lulosic ethanol production include a large number of species within the Poaceae, Pinaceae, and Salicaceae families. For these biofuel feedstock species, there are variable amounts of genome sequence resources available, ranging from complete genome sequences (e.g. sorghum, poplar) to transcriptome data sets in the form of Expressed Sequence Tags (e.g. switchgrass, pine). While obtaining genome or transcript sequence is the initial step in a genomics-based approach to biological questions, the more challenging step in genomics the process of understanding gene function and how genes and their products confer the underlying processes/traits in plant biology. This is attributable to a large extent on two issues: a large percentage of genes within genomes have no known function and experimental approaches to determining gene function on a per gene basis are fiscally prohibitive. One method to improve our understanding of gene function is through comparative approaches in which sequence similarity is used to cross-annotate orthologs and paralogs thereby leveraging all available functional annotation data to improve the annotation of genes in species with limited annotation data. We will provide computational tools and resources for data-mining genomic based information available for biofuel feedstock species. Through the Biofuel Feedstock Genomics Resource (<http://bfgr.plantbiology.msu.edu/>), we will create a comprehensive, uniform, well annotated resource for data-mining genomic data for biofuel feedstock species. Functional and comparative genomics will be used to improve the quality and quantity of annotation available for biofuel feedstock species. Model genomes will be included in the comparative analyses to leverage the wealth of information and resources currently available. We will generate new hypotheses regarding gene function using a systems biology approach in which gene associations in rice are linked to orthologous sequences in other Poaceae biofuel feedstock species. Collectively, this project complements existing genomic and bioinformatic resources in that it is focused on integrating sequences and annotation data already available and leverages existing data to allow for translation of information from well understood genes and genomes to biofuel feedstock species, especially those in which only partial genome sequence, annotation and functional genomics resources are available.

USDA-DOE

## Translational Genomics for the Improvement of Switchgrass

Bryan W. Penning, Anna T. Olek, Maureen C. McCann, and **Nicholas C. Carpita\*** ([carpita@purdue.edu](mailto:carpita@purdue.edu))

Purdue University, West Lafayette, Ind.

Switchgrass is targeted to become a future biomass crop, but the discovery of genes underlying biomass-relevant traits is compromised in switchgrass by the paucity of genetic resources. Maize provides a genetic resource for improvement of distinct cell walls of switchgrass and other energy grasses. Comparative genomics of maize, rice, and *Arabidopsis* sequences reveal marked differences in gene family struc-

ture between grass species and dicotyledonous species. A description of function for an estimated 1500 genes related to cell wall biology will be provided for switchgrass, based on homology to maize and rice sequences, and augmented with gene families that are currently of unknown function. Maize cell wall gene family members that are highly expressed during primary and secondary wall formation will be determined, and regulatory small RNA sequences will be identified. As proof of concept, we will test mutants and transgenic lines of maize for release of glucose and xylose in functional analyses of genes that impact structure and degradability of non-cellulosic polysaccharides. Analyses of cell wall structures and saccharification potential will be applied to developing stems and stover of maize and switchgrass.

USDA-DOE

## Genomic Knowledgebase for Facilitating the Use of Woody Biomass for Fuel Ethanol Production

**Vincent Chiang** ([Vincent\\_Chiang@ncsu.edu](mailto:Vincent_Chiang@ncsu.edu)), Ron Sederoff, Ho-min Chang, Ying-Hsuan Sun,\* Ting-Feng Yeh, Shanfa Lu, Quanzi Li, and Rui Shi

Dept. of Forestry and Environmental Resources and Dept. of Wood and Paper Science, North Carolina State University, Raleigh, N.C.

**Project Goals: We propose to develop new, genome-wide understanding of the regulatory control of cellulose, xylan and lignin accumulation in wood cell walls of *Populus*. To do this, transcriptional profiling will be applied to two *Populus* systems: (1) our existing transgenic *Populus* lines having increased cellulose as a result of lignin reduction and (2) mechanical stress induced tension wood known for its augmented cellulose and reduced xylan and lignin formation. The combination of these two systems is unique in that it may provide information about transcriptional control for preferential distribution of carbon between cellulose, xylan or lignin. Transcriptional profiling of a biological system where the process of interested is preferentially regulated is most effective for identifying the responsible genetic control. We wish to establish a comprehensive database to include all the possible genes or gene networks that may enable effective manipulation of the biosynthesis of cellulose, xylan and lignin. Our goal is to provide genomics based approaches for developing novel, high energy woody biomass for the production of ethanol. *Populus* is the only DOE target energy crop whose genome has been sequenced (<http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>), offering a unique opportunity for achieving this goal.**

Wood, consisting of cellulose, hemicelluloses and lignin, is a major renewable lignocellulosic resource that can be used as feedstock for fuel ethanol production. However, lignin and xylan hemicellulose are two major barriers to an

efficient lignocellulose conversion to ethanol. Tension wood is a specialized wood tissue formed only in angiosperm trees species, such as *Populus*, to counteract mechanical and gravitational stresses. Tension wood contains substantially increased amount of cellulose and reduced quantity of lignin and xylan. It also contains increased number of fiber cells, of which walls are enriched with more extractable lignin (high lignin S/G component ratio). These tension wood properties are highly desirable for ethanol production. The knowledge of tension wood development provides insights into regulatory networks associated with the control of the biosynthesis of cellulose, hemicelluloses and lignin, forming the basis for strategies to create more desirable wood or lignocellulosics for fuel ethanol production.

The sequenced genome of *Populus trichocarpa* allows us to take a genome-wide approach to studying regulatory mechanisms underlying the formation of tension wood. We used one-year-old greenhouse grown *P. trichocarpa* trees to study the development of tension wood by bending the tree stem to create the mechanical stress-induced tension wood. The bending was performed for 1, 3, 7, 21, 35 and 49 days. The developing xylem tissues of the bent stem as well as stem of normal trees (control) were collected for profiling gene expression using Affymetrix *P. trichocarpa* full genome microarrays. Cell wall sugar contents (for glucose, xylose, mannose rhamnose and arabinose), xylan synthase and galactan synthase activities were quantified and correlated with microarray-derived gene expression patterns.

In addition to the expected changes, such as a sharp reduction in xylan content, we also found a drastic increase of galactan quantity during tension wood formation. Galactan, a C-6 hemicellulose, is normally present as a minor hemicellulose in wood of angiosperms. Consistently, we found increased galactan and decreased xylan synthase activities in developing xylem of tension wood. Microarray experiments identified 1220 differentially expressed genes during tension wood formation based on a statistical robust mixed-model analysis and stringent Bonferroni correction for multiple tests. These differentially expressed genes include those involved in major sugar biosynthesis pathways. Based on the Pearson's correlation coefficient between the expression profiles and galactan synthase activities, we identified 6 putative galactosyl transferases (Pearson's  $r$  range from 0.83 to 0.93) that may be responsible for the increased galactan content in tension wood. Correlation analysis, however, did not identify genes that are highly correlated (Pearson's  $r > 0.8$ ) with the decreased xylan synthase activity. The decreased xylan content found in tension wood may suggest the involvement of a post-transcriptional control of xylan biosynthesis. A proteomic approach using 2-D PAGE LC-MS/MS was also applied to complement the transcriptome results.

Parallel experiments using transgenic *P. trichocarpa* trees with reduced lignin and increased lignin S/G ratio were also carried out to identify genes that may be associated with changes in these lignin traits. Data are being analyzed and will be discussed.

## Systematic Modification of Monolignol Pathway Gene Expression for Improved Lignocellulose Utilization

**Richard A. Dixon\*** (radixon@noble.org), Fang Chen, Rui Zhou, Lisa Jackson, Gail Shadle, and Qiao Zhou

Plant Biology Division, Samuel Roberts Noble Foundation, Ardmore, Okla.

<http://www.noble.org/PlantBio/Dixon/Index.htm>

**Project Goals: 1) To gain a detailed understanding of the biosynthetic pathways to the lignin building blocks in alfalfa, with focus on potentially independent pathways to guaiacyl and syringyl lignin; 2) To evaluate the saccharification efficiency of cell walls from alfalfa lines with various targeted modifications to lignin content and composition; 3) To develop an unbiased forward genetic screen for *Medicago truncatula* mutants with altered levels and cellular patterns of lignification, and to begin to characterize the affected genes; and 4) To provide lignin analytical support for the project "Genetic dissection of the lignocellulosic pathway of grasses" from Kansas State University (Bikram Gill, PI).**

Alfalfa (*Medicago sativa*) has recently been promoted as a potential bioenergy crop. Currently, it is the world's major forage legume, with an average annual value of more than \$8 billion in the USA alone, supported by extensive agronomic improvement programs in the private sector. Although its yields do not approach those of the currently favored bioenergy crops switchgrass, *Miscanthus* or poplar, alfalfa is a high yielding perennial with the added benefits of nitrogen fixation, well defined agronomic practices, seed industry support, and extensive translational genomics resources through studies on the closely related model species *M. truncatula*. It also has potential bioenergy uses in rotation with corn.

The compositions and structures of lignified plant cell walls have been identified as key factors limiting effective biomass to biofuel conversion in processes where the sugar components of polysaccharides are released (saccharification) prior to fermentation to ethanol. Lignin is one of the major components contributing to the recalcitrance of plant cell walls to saccharification.

Prior to the current DOE funding, we had identified the majority of the genes involved in the biosynthesis of monolignols in *Medicago*, and had generated stably transformed alfalfa lines independently down-regulated in seven different enzymatic steps of the monolignol biosynthetic pathway (L-phenylalanine ammonia-lyase, cinnamate 4-hydroxylase [C4H], hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase [HCT], coumaroyl shikimate 3-hydroxylase [C3H], caffeoyl CoA 3-O-methyltransferase [CCoAOMT], ferulate (coniferaldehyde) 5-hydroxylase [F5H], or caffeic acid 3-O-methyltransferase [COMT]). This resulted in lines with wide differences in both lignin

content and composition in the same genetic background. During the present granting period, we have generated additional alfalfa lines down-regulated in cinnamoyl CoA reductase and cinnamyl alcohol dehydrogenase. Analysis of the above lines revealed that reduction in lignin content, more so than altering lignin composition, can result in large improvements in cell wall saccharification efficiency, and potentially obviate the need for acid-pretreatment, thus facilitating consolidated bioprocessing [1,2]. These lines have also provided baseline “improved” material for the complete repertoire of analytical procedures available within the DOE Bioenergy Sciences Center.

Although recent models of the monolignol pathway show an essentially linear pathway as far as coniferaldehyde, earlier models proposed a more complex metabolic grid. We have shown both biochemically and genetically that COMT may function at the level of 3-*O*-methylation of caffeoyl aldehyde or caffeoyl alcohol, preferred substrates for the enzyme from *Medicago*. This raises the question of the origin of caffeoyl aldehyde. We have now demonstrated that *Medicago* possesses two distinct cinnamoyl CoA reductases, one with preference for feruoyl CoA, the other with preference for coumaroyl CoA and caffeoyl CoA (the precursor of caffeoyl aldehyde). Their kinetic properties suggest that they might function to control independent pathways to the different monolignols. This is being tested through the analysis of *M. truncatula* mutant lines harboring transposon insertions in these genes. We have also identified two *Medicago* cinnamyl alcohol dehydrogenase genes encoding proteins with substrate specificities consistent with this model.

We have developed a method for the large scale screening of *M. truncatula* transposon insertion lines [3] with altered levels/patterns of lignification, by visualization of lignin/cell wall phenolic autofluorescence in stem cross sections by UV microscopy. We will describe the visible and chemical phenotypes of selected lines that have been confirmed through analysis of homozygous progeny. We currently have around 20 confirmed mutants with interesting properties, varying from an overall reduction in lignin level to ectopic lignin deposition (lacking in some cell types but not others, or present in cell types where it is not usually found). The affected genes causing these phenotypes are being pursued through analysis of genomic sequences flanking the Tnt1 insertions coupled with segregation analysis.

Finally, we have continued to provide lignin analytical support for the project “Genetic dissection of the lignocellulosic pathway of grasses”. Initially, we concentrated on plant material generated through virus-induced gene silencing of wheat monolignol pathway genes, and have more recently been analyzing stable transgenic wheat lines.

## References

1. Chen, F., and Dixon, R.A. (2007) Lignin modification improves fermentable sugar yields for biofuel production. *Nature Biotechnology* 25, 759-761.
2. Jackson, L.A., Shadle, G.L., Zhou, R., Nakashima, J., Chen, F., and Dixon, R.A. (2008). Improving saccharification efficiency of alfalfa stems through modification of

the terminal stages of monolignol biosynthesis. *Bioenergy Research* 1, 180-192.

3. Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., Cayrel, A., Endre, G., Zhao, P.X., Chabaud, M., Ratet, P., and Mysore, K.S. (2008) Large scale insertional mutagenesis using Tnt1 retrotransposon in the model legume *Medicago truncatula*. *Plant Journal* 54, 335-347.

USDA-DOE

## Genetic Dissection of the Lignocellulosic Pathway of Wheat to Improve Biomass Quality of Grasses as a Feedstock for Biofuels

W. Li<sup>1\*</sup> (wli@ksu.edu), C. Bi,<sup>1</sup> S.K. Sehgal,<sup>1</sup> N. Rothe,<sup>1</sup> H. Trick,<sup>1</sup> F. Chen,<sup>2</sup> R.A. Dixon,<sup>2</sup> and B.S. Gill<sup>1</sup> (bsgill@ksu.edu)

<sup>1</sup>Dept. of Plant Pathology, Kansas State University, Manhattan, Kans.; and <sup>2</sup>Dept. of Plant Biology, Noble Foundation, Ardmore, Okla.

http://www.ksu.edu/wgrc

Project goal is to identify the major determinants of lignocellulosic degradation in grasses. Our objective is to dissect the lignin pathway of grasses using wheat as a model by coupling functional genomics and metabolite profiling. Our specific objectives are: 1) to investigate the expression of candidate genes for lignin biosynthesis; 2) to silence these genes by RNAi; 3) to develop knockout mutants of these genes and; 4) to characterize the silenced tissues and knockout mutants by metabolite profiling. Combining database mining and expression pattern studies, we identified 21 candidates for the 10 families of monolignol biosynthetic genes. Expression of these 21 genes is under strong developmental regulation and they are expressed at much higher level in the stem as compared to the leaf sheath and blade. This is correlated with higher lignin content in stem as compared to leaf. We have developed RNAi constructs for these 21 genes and obtained RNAi transgenic plants for 12 genes. Expression of the target genes was silenced by up to 90%. At the same time, we have developed a mutagenesis population of 3000 single M2 plants. The metabolite profiling of the RNAi transgenic plants and screening of lignin mutants by TILLING is under way.

USDA-DOE

## Development of Genomic Tools for the Improvement of Prairie Cordgrass (*Spartina pectinata*), a Highly Productive Bioenergy Feedstock Crop

Jose L. Gonzalez\* (jose.gonzalez@sdstate.edu), Arvid Boe, and Vance Owens

Dept. of Plant Sciences, South Dakota State University, Brookings, S.D.

**Project Goals: Develop SSR markers and construct a linkage map for prairie cordgrass (*Spartina pectinata*) to help in the development of this species as a biomass crop.**

Considerable attention has been turned towards the production of biofuels from cellulosic biomass crops. Prairie cordgrass (*Spartina pectinata*) is a North American native perennial C<sub>4</sub> grass which yields up to 22 metric tons/ha in South Dakota plantings, compared to 11 metric tons/ha for other C<sub>4</sub> grasses. Due to the potential of Prairie cordgrass as a cellulosic biomass crop species, a research project has been instigated to assist in this grasses domestication and utilization.

An initial step is the production of the first molecular map of prairie cordgrass. The molecular map is being produced using microsatellite markers developed from sequenced genomic libraries enriched for the microsatellite sequences; (CA)<sub>n</sub>, (GA)<sub>n</sub>, (AAG)<sub>n</sub> and (CAG)<sub>n</sub>. Initial investigations of 100 clone sequences from the four libraries identified a 60% enrichment rate. Primers were designed for the enriched sequences and tested in a sub-sampling of 16 genotypes, identifying numerous polymorphisms. Due to the positive result from the analysis of these 100 clones, a total of 3,072 clones from the (CA)<sub>n</sub> library were sequenced producing 2475 viable sequences, of which 1678 contained simple SSR regions and were used to develop primers. In total from the initial 100 sequences from the four libraries and the 1678 sequences from the (CA)<sub>n</sub> library, 551 putative loci were identified. Primers were designed for all 551 putative loci and tested for viability and polymorphism. PCR conditions were optimized to allow high through put screening of a mapping population, thus two annealing temperatures (53°C and 55°C) were utilized. Results from the primer screening identified 97 primer pairs which were polymorphic at 53°C, 134 primer pairs which were polymorphic at 55°C, 83 primer pairs amplified but were not polymorphic and 197 which didn't amplify under the high through put conditions.

The 231 polymorphic primer pairs have been used to develop a molecular map using a pseudo test cross approach using an F<sub>1</sub> generation of a reciprocal cross between two open-pollinated parents, collected originally from the Red River valley. The final map will be utilized for marker assisted selection in a breeding program with SDSU forage and biomass crops breeder Dr. Arvid Boe. Utilizing

association mapping techniques, previously validated by our research group, quantitative traits such as yield and disease resistance will be expeditiously incorporated into the cultivated crop.

Additional research efforts include a germplasm collection throughout the Midwestern states of the USA. The genetic diversity of this core genotype collection has been evaluated with AFLPs. At the same time a disease survey is also being conducted. Additional research being undertaken to assist the domestication of Prairie cordgrass includes a characterization of the transcriptome using 454 technology, and the development of a BAC library. These efforts complement efforts in the breeding program and in crop management techniques.

USDA-DOE

## Profiling the Small RNAs of *Brachypodium distachyon*

Matthias Ganssmann,\* Marcelo German, Blake C. Meyers, and Pamela J. Green (green@dbi.udel.edu)

Delaware Biotechnology Institute, University of Delaware, Newark, Del.

**Project Goals: Our project aims to examine small RNA regulation in *Brachypodium distachyon* in key tissues and in seedlings under different abiotic stresses. By profiling small RNAs and mRNAs, and investigating the targets of miRNAs, this project will enhance the value of the genome and EST information now being generated. Ultimately, this knowledge should improve our ability to devise strategies to enhance growth and yield of grasses that are less amenable to functional genomic analysis.**

MicroRNAs (miRNAs) are a class of small (~21-24 nucleotide) non-coding RNAs with a major role in regulating gene expression in most multicellular eukaryotic organisms. Base pairing of a miRNA to its target mRNA prevents the translation of the mRNA either by cleavage and subsequent degradation of the transcript, or translational repression by a yet unknown mechanism. Important biological processes such as structural organization of leaves and inflorescences are regulated by miRNAs. Recent findings, mainly from the model plant *Arabidopsis thaliana*, revealed the association of miRNAs with abiotic stress responses. The goal of this study is to gain insight on small RNA mediated gene regulation in *Brachypodium distachyon* (*Brachypodium*). The close relation to major cereal grain species and to Poaceae species important for energy production, together with *Arabidopsis*-like features accentuates the suitability of *Brachypodium* to serve as a model organism for temperate grasses in functional genomic research.

Our project has four specific aims: 1. Profile small RNAs and mRNAs from key tissues, stresses, and small RNA biosynthesis mutants. 2. Characterize the regulation and target RNA cleavage products of selected miRNAs that are



associated with specific stresses. 3. Compare the *Brachypodium* data to more limited information from switchgrass to identify conserved miRNAs. 4. Provide public access to the data through a web interface.

Completion of our project objectives should result in the identification of miRNAs and target RNAs that may have important roles in stress responses in *Brachypodium* and related grasses that can be studied in mechanistic detail in the future. In addition to providing the initial results and tools for the latter, by profiling miRNAs, other small RNAs, and mRNAs very deeply, this project will greatly enhance the value of the genome and cDNA sequences now being generated. These resources can then be examined for small RNAs that have the potential to regulate most any plant process.

We have prepared small RNA libraries from major tissues of *Brachypodium* and subjected these to deep sequencing with Illumina's sequencing by synthesis method. This has yielded at least 4 million reads from each tissue and we are using this data to identify the small RNAs in the population and their abundances. The results from this initial profiling analysis will be presented and the miRNAs exhibiting the strongest differential regulation in key tissues will be identified. We have also generated mRNA libraries from several tissues and small RNA libraries from plants subjected to several environmental stresses. These are currently in various stages in our sequencing and analysis pipeline. In addition, three small RNA libraries have been sequenced from switchgrass under an NSF project that will allow conserved miRNAs to be identified.

Supported by grant DE-FG02-07-ER64450

USDA-DOE

### Transcription Factor Protein Interactions with Cell Wall Gene *cis*-Regulatory Regions and Their Overall Role in Bioenergy Feedstock Properties

Naomi Young,<sup>1</sup> Ghislain Breton,<sup>2</sup> Todd C Mockler,<sup>3</sup> Steve A Kay,<sup>2</sup> and **Samuel P Hazen**<sup>1\*</sup> (hazen@bio.umass.edu)

<sup>1</sup>Dept. of Biology, University of Massachusetts, Amherst, Mass.; <sup>2</sup>Section of Cell and Developmental Biology, University of California San Diego, La Jolla, Calif.; and <sup>3</sup>Center for Genome Research and Biocomputing and Dept. of Botany and Plant Pathology, Oregon State University, Corvallis, Ore.

**Project Goals: The long-term goal of this project is to understand the regulatory mechanism underlying plant cell wall biosynthesis at the transcriptional level. Results from these pursuits will ultimately lead to a better understanding of growth and development and cell wall biosynthesis as well as a systems level insight into regulatory**

### networks affecting monocot and dicot bioenergy-related properties.

There is extensive heterogeneity in cell wall composition among and within species and certainly within a single plant. Walls are modified to fulfill the particular needs of a cell type at a specific developmental stage. There should indeed be several orders of complexity to control different pathways leading to the appropriate synthesis and assembly of wall components leading to a particular wall function. We have begun to screen for transcription factor (TF) protein interactions with promoters of key cell wall genes involved in cellulose, hemicellulose, and lignin synthesis. In a pilot screen using a subset (n= 250) of the complete library of *Arabidopsis* TFs, we identified novel interactions between the Trihelix and GRAS families of TFs and the phenylpropanoid metabolic grid. Our current strategy targets the direct regulators of processive as well as the non-processive glycosyltransferases. As a follow-up of the yeast one-hybrid assay, we will use ChIP-seq to capture and identify the DNA bound to the identified transcription factor proteins using overexpressed tag versions and define the extent of their genomic targets. Concurrently, we will develop gain- and loss-of-function mutants in the dicot model *Arabidopsis* and the monocot model *Brachypodium* to determine the effects of these perturbations on cell wall properties and feedstock quality. Results from these pursuits will ultimately lead to a better understanding of growth and development and cell wall biosynthesis as well as a systems level insight into regulatory networks affecting monocot and dicot bioenergy-related properties.

USDA-DOE

### Biochemical Genomics of Wood Formation: O-acylesterification for Alteration of Lignocellulosic Property

**Chang-Jun Liu**\* (cliu@bnl.gov), Xiao-Hong Yu, and Jin-Ying Gou

Biology Dept., Brookhaven National Laboratory, Upton, N.Y.

**Project Goals: The goals of the project are to systemically characterize plant acyl-CoA dependent acyltransferases and acylesterases; by which, to identify the specific enzymes involved in lignocellulosic acylesterification, and further to explore the underlying biochemical mechanisms and biological functions of cell wall acylesterification.**

Enzymatic O-acylesterification is a common modification of plant lignocelluloses. Acylesterification affects cell wall's structural property, degradability, and eventually, the growth, development and reproduction of plant feedstocks. The enzymatic O-acylation and deacylation reactions also occur in the formation of a variety of secondary metabolites that are required for disease resistance and forming heart wood in trees. Despite the wide occurrence of O-acylesterification

in plant metabolism and cell wall modification, the enzymes involved in such structural modification remain largely elusive, and the biological roles of different types of cell wall acylesterification are unclear. We are employing biochemical genomics and reverse genetics approaches to systemically characterize plant acyl-CoA dependent acyltransferases; by which, we are going to identify the specific enzymes involved in lignocellulosic acylesterification, and further to explore the underlying biochemical mechanisms and biological functions of cell wall acylesterification.

In previous fiscal years, we have identified 94 and 61 putative acyl-CoA dependent acyltransferase and 10 acylesterase genes from poplar and *Arabidopsis*, respectively. To further probe gene expression pattern, we conducted “in silico” transcript analysis and RT-PCR for all of the identified putative poplar genes by using RNAs prepared from poplar leaf, root, developing stem, apical bud, cortex of bark, phloem, developing wood and lignified wood. The bioinformatics study allowed us to reveal unique features of acyltransferase family genes in their gene structure, organization and distribution on the genomes of both *Arabidopsis* and poplar; and the gene expression profiling resulted in the discovery of a number of acyl-CoA dependent *O*-acyltransferase and acylesterase genes preferentially expressed in secondary cell wall forming tissues. These studies laid a strong basis for understanding the evolutionary relationship among acyltransferase family genes, and the gene functions implicated in plant growth, development and metabolism.

In order to understand the molecular mechanism of cell wall acylesterification, we analyzed the wall-bound acylesters of poplar by using Liquid Chromatography-Mass Spectrometry (LC-MS), Fourier Transform-InfraRed (FT-IR) microspectroscopy, and synchrotron InfraRed (IR) imaging facility. The results revealed that the cell wall of dicotyledonous poplar, as the walls of many monocot grasses, contains a considerable amount of acylesters, primarily acetyl and *p*-hydroxycinnamoyl molecules. The “wall-bound” acetate and phenolics display a distinct tissue specific-, mechanical stress responsible- and developmental accumulation pattern, indicating distinct roles of different “wall-bound” acylesters in poplar cell wall structural construction and/or metabolism of cell wall matrix components.

To functionally identify the putative acyltransferases, we performed comprehensive *in vitro* functional screening for a number of the produced recombinant proteins by incubating with a collection of thioester donors and acceptors. The studies let us characterize several novel acyltransferases in relation to cell wall biogenesis and modification, which include (i) a novel monolignol acetyltransferase specifically responsible for the modification of lignin biosynthetic precursor, sinapyl alcohol. The biochemical property and subcellular localization of this enzyme were further investigated. The correlation of gene expression to lignin acetylation was explored and the transgenic plants were created. (ii) two hydroxycinnamoyltransferases that display activities for synthesis of both monolignol biosynthetic intermediates and chlorogenic acid, a metabolite acting as phytoalexin in plant defense responses. Overexpressing these

genes in tobacco resulted in the interesting changes in lignin content, compositions, and xylem morphology. Additionally, we also characterized several other novel hydroxycinnamoyl-transferases responsible for the modification of a variety of phenyl or aliphatic alcohols. These identified acyltransferase genes will be useful molecular tools for re-directing photosynthetic carbons from lignin polymer biosynthesis into the value-added metabolites in plant feedstocks.

USDA-DOE

## Characterization of Pectin Acetylerase Reveals Critical Roles of Cell Wall Acylesterification in Plant Growth and Development

Jin-Ying Gou\* and Chang-Jun Liu (cliu@bnl.gov)

Biology Dept., Brookhaven National Laboratory, Upton, N.Y.

**Project Goals: The goals of the project are to systemically characterize the enzymes involved in lignocellulosic acylesterification, and to understand the biochemical mechanisms and biological functions underlying cell wall modification.**

Pectin is one of the structural components of plant cell wall, predominant in the primary walls of fast growing tissues. In plant, pectin is developmentally modified by methylesterification and acetylerification. The acetylerification of homogalacturonan usually occurs on the *O*-2 or/and *O*-3 positions in the galacturonic acid residues but its physiological function has not been well studied. In our group, a putative pectin acetylerase homolog from poplar, ptPAE1, was cloned and heterologously expressed in *E. coli*. The recombinant protein was purified as an active enzyme form that could release acetyl moieties from the acetylated polysaccharides, pectin, xylan, and arabinogalactan. The ptPAE1 exhibits optimal activity at pH 7 and 35°C on the acetylated pectin and kinetically prefers substrate pectin over xylan. The identified esterase gene driven by a double 35S promoter was over-expressed in tobacco. The promoted de-acetylation of polysaccharides due to the constitutive expression of ptPAE1 resulted in obvious morphological phenotype. The apical buds of transgenic plants displayed constant senescence at flowering stage. In the flowers of transgenic plants, the development of filament was impeded, which resulted in short stamen and, consequently, the seed pods of the transgenic plants were greatly reduced both in size and weight, compared to the control plants. The offspring seedlings harboring transgene generated shorter hypocotyls and roots than non-transgenic control plants. The growth retardation resulted from the disturbance of pectin acetylation in the fast elongation tissues indicates that acetylerification of pectin has significant roles in plant growth and development.

USDA-DOE

## *Brachypodium* Transcriptomics

Henry D. Priest<sup>1\*</sup> (priesth@onid.orst.edu), Samuel Fox,<sup>1</sup> Wenqin Wang,<sup>2</sup> Todd P. Michael,<sup>2</sup> and **Todd C. Mockler<sup>1</sup>** (tmockler@cgrb.oregonstate.edu)

<sup>1</sup>Center for Genome Research and Biocomputing and Dept. of Botany and Plant Pathology, Oregon State University, Corvallis, Ore.; and <sup>2</sup>Rutgers, The State University of New Jersey, Dept. of Plant Biology and Pathology, The Waksman Institute of Microbiology, Piscataway, N.J.

**Project Goals:** *Brachypodium distachyon* is a model grass plant with close evolutionary relationships to economically important species including the forage and turf grasses, temperate cereals, and several potential biofuel feedstock crops. We are designing an Affymetrix *Brachypodium* whole-genome microarray that will be used for whole-genome expression profiling over major developmental stages, diurnal and circadian time-courses, and stress conditions. The resulting *Brachypodium* gene expression atlas will allow us to better understand the gene regulation networks underlying traits of major importance for both the quality and quantity of biomass.

*Brachypodium distachyon* is a model grass plant with close evolutionary relationships to economically important species including the forage and turf grasses, temperate cereals, and several potential biofuel feedstock crops. To better understand the gene regulation networks underlying traits of major importance for both the quality and quantity of biomass, we are pursuing a hypothesis-generating approach involving whole-genome expression profiling over major developmental stages, diurnal and circadian time-courses, and stress conditions. We are designing an Affymetrix *Brachypodium* whole-genome microarray, which will become available for purchase by any researcher. In collaboration with MIPS and JGI we are currently annotating the *Brachypodium* genome and transcriptome. A portion of the microarray space will be used to tile each predicted exon and intron with multiple probes. The remainder of the genome will be tiled to represent predicted non-coding/intergenic regions. We anticipate that the array design will be completed and Affymetrix will manufacture microarrays in the first quarter of 2009. We will use these arrays to map major gene expression changes of relevance to important traits of grass crops. We have been collecting *Brachypodium* tissues representing a number of different developmental stages, environmental conditions, and stress treatments—including (a) different tissues from plants of different developmental stages (roots, shoots, leaves, developing flower spikes, developing seeds, germinating seeds, etiolated seedlings); (b) diurnal and circadian sampling in light/dark/temperature cycles; and (c) plants exposed to abiotic stresses (heat, cold, salt, dehydration, nutrient limitation, and light stress). We are currently preparing RNAs from these samples for use with the *Brachypodium* genome arrays to establish baseline expression profiles for each sample. We will use these data

to create a comprehensive public gene expression atlas. We will analyze and compare the expression patterns of *Brachypodium* genes with respect to their inferred biological functions. We will cluster co-expressed genes and perform promoter analyses to predict transcription factor binding sites in their upstream regions. Our results will be integrated into the gene annotations available through BrachyBase.org and other annotation databases, and we will develop bioinformatic tools for public access to these resources.

USDA-DOE

## Identification of Cell Wall Synthesis Regulatory Genes Controlling Biomass Characteristics and Yield in Rice (*Oryza sativa*)

Zhaohua Peng<sup>1\*</sup> (zp7@BCH.msstate.edu), Pamela Ronald,<sup>2</sup> and Guo-liang Wang<sup>3</sup>

<sup>1</sup>Dept. of Biochemistry and Molecular Biology, Mississippi State University, Miss.; <sup>2</sup>Dept. of Plant Pathology, University of California, Davis; and <sup>3</sup>Dept. of Plant Pathology, Ohio State University, Columbus, Ohio

Understanding the regulatory mechanisms of cell wall synthesis is essential for the improvement of biomass characteristics critical to biofuel production. To study the cell wall synthesis regulation, we are examining the cell wall regeneration process in protoplasts of rice (*Oryza sativa*). In protoplasts, the cell wall synthesis is highly activated due to the removal of cell wall. We are analyzing the transcriptome and proteome dynamic changes in response to the removal of cell wall using DNA oligo array and shotgun proteomics, respectively. The differentially regulated genes and proteins are further analyzed using systems biology method to elucidate possible pathways involved in cell wall synthesis and regeneration. In addition, mutants and transgenic lines of putative regulatory genes and critical metabolic pathway genes have been generated. Some of the mutants/transgenic lines have displayed changes in cellulose, hemicelluloses and lignin contents.

USDA-DOE

## NIRS Prediction of Corn Stover Cell Wall Composition and Conversion Potential, and Relationships among these Traits

H.G. Jung<sup>1,2\*</sup> (Hans.Jung@ars.usda.gov), M.L. Lewis,<sup>2</sup> R.E. Lorenzana,<sup>2</sup> and **R. Bernardo<sup>2</sup>** (bernardo@umn.edu)

<sup>1</sup>USDA-Agricultural Research Service, Plant Science Research Unit, St. Paul, Minn.; and <sup>2</sup>Dept. of Agronomy and Plant Genetics, University of Minnesota, St. Paul, Minn.

[http://agronomy.cfans.umn.edu/corn\\_cellulosic\\_ethanol.html](http://agronomy.cfans.umn.edu/corn_cellulosic_ethanol.html)

**Project Goals: 1) Determine the prospects of and identify challenges in marker-assisted breeding for both corn grain yield and stover quality traits important for ethanol production. 2) Determine if genome-wide selection (which does not require finding markers with significant effects) is superior to the usual approach of selecting only for significant markers, with the goal of simultaneously improving corn grain yield and stover quality.**

Implementation of a breeding program (conventional or marker assisted) for corn stover quality traits related to cellulosic ethanol production is dependent on having a reliable, accurate, rapid, and inexpensive system for phenotyping plant material. Determination of cell wall composition and testing for fermentable sugar release in a cellulosic ethanol conversion process by conventional wet chemistry methods is laborious and expensive. However, the feed grain and forage industries have demonstrated that near-infrared reflectance spectroscopy (NIRS) is a routine and reliable system for predicting quality parameters rapidly and at low cost. We report here on the NIRS prediction equations we have developed for phenotyping corn stover in support of our project on marker-assisted selection to improve corn stover as a cellulosic ethanol feedstock. Also, correlations among the quality traits were examined for prediction of conversion potential of corn stover to determine if information on cell wall composition would be sufficient for selecting corn genotypes that have high conversion potential.

Testcrosses of 223 recombinant inbred lines (RIL) derived from B73 x Mo17 and the two parental inbreds to a tester inbred were grown at four locations in Minnesota in 2007. Corn stover, minus the ear (grain, cob, husk, and shank), was collected at grain maturity. Stover was chopped, dried at 60°C, and ground prior to scanning by NIRS (1100 to 2500 nm). A calibration sample set (N=154) was selected from the 900 total samples based on a combination of spectral diversity and physical design of the experiment. Calibration samples were analyzed for cell wall composition by the two-stage acid hydrolysis method of Theander et al. (1995) and conversion potential in a dilute acid/high temperature pretreatment and enzymatic saccharification process (Dien et al., 2006). Percent cell wall sugars released by the conversion process were determined as the difference between stover composition and the wall sugars remaining in the insoluble residue from the conversion process. NIRS calibration equations for stover quality were developed by modified partial least squares regression. Stover quality was predicted for all samples from their NIRS spectra. Mean data for corn lines were used for correlation analysis of relationships among the cell wall components and with sugar release.

The NIRS calibration equation statistics indicated that stover cell wall components can be predicted with varying precision. Glucose (primarily from cellulose) was well predicted but the other major components gave only moderate fits in the calibrations. Better calibrations have been observed for cell wall composition of corn plant part, alfalfa stem, and switchgrass herbage sample sets; however, those studies included more variability in composition due to multiple maturity stages which generally improves NIRS calibration

statistics. It was observed that calibration of glucose release by the conversion process had a high  $R^2$  whereas release of other cell wall sugars gave poor calibrations. In contrast, *in vitro* rumen digestibility of all cell wall sugars have high calibration statistics. This difference can be explained by the fact that glucose release in the conversion process and rumen digestibility are both enzymatic (biochemical) processes and cell wall matrix structure interacts with enzyme accessibility to polysaccharides. This is unlike hydrolysis of non-glucan polysaccharides in the conversion process which was primarily an acid (chemical) reaction that should be less influenced by wall matrix structure. NIRS measures organic matrix structure of biomass in the form of hydrogen bond stretching, not actual degradation of polysaccharides by either chemical or biochemical reactions.

Although NIRS calibration statistics were variable among the cell wall traits, we were able to determine that the corn RILs contained significant genetic variation for all traits and identify important quantitative trait loci. As expected for a biochemical process, release of glucose was negatively correlated with lignin concentration ( $r = -0.67$ ), similar to what has been observed for rumen digestibility. Xylose and other non-glucan sugars were correlated with glucose release to varying degrees ( $r = -0.13$  to  $0.56$ ), but were also correlated with lignin concentration. These cell wall sugar correlations with glucose release are most likely a reflection of extent of cell wall development, rather than an impact on glucose release *per se*. Glucose concentration was not related to glucose release. Release of xylose and arabinose were not correlated with lignin concentration, and uronic acid release was only poorly correlated with lignin ( $r = -0.25$ ). Again, because non-glucan sugar release was a chemical hydrolysis process we did not expect large effects of lignin on their release by this conversion process. It appears that release of glucose, but not other wall sugars, can be effectively predicted from cell wall composition.

**NIRS calibration statistics for corn stover quality traits and the observed range for predicted data.**

Trait	Samples Used	Mean	SEC	R <sup>2</sup>	Range
		g kg-1 DM			g kg-1 DM
Klason lignin	137	167	6	0.66	161 - 174
Glucose	135	344	5	0.92	322 - 361
Xylose	138	203	8	0.46	194 - 212
Arabinose	141	27	1	0.67	26 - 29
Uronic acids	138	27	1	0.50	26 - 28
		%			%
Glucose release	138	51.8	1.6	0.83	48.0 - 55.5
Xylose release	142	85.0	1.4	0.12	84.6 - 85.6
Arabinose release	136	88.1	1.2	0.30	87.4 - 89.0
Uronic acid release	137	75.4	2.3	0.12	74.9 - 76.4

## References

1. Dien, B.S., H.G. Jung, K.P. Vogel, M.D. Casler, J.F.S. Lamb, L. Iten, R.B. Mitchell, and G. Sarath. 2006. Chemical composition and response to dilute-acid pretreatment and enzymatic saccharification of alfalfa, reed canarygrass, and switchgrass. *Biomass Bioenergy* 30:880-891.
2. Theander, O., P. Aman, E. Westerlund, R. Andersson, and D. Pettersson. 1995. Total dietary fiber determined as neutral sugar residues, uronic acid residues, and Klason lignin (The Uppsala Method): collaborative study. *J. AOAC Int.* 78:1030-1044.

---

 USDA-DOE
Epigenomics of Development in *Populus*

Kyle R. Pomraning,<sup>2</sup> Palitha D. Dharmawardhana,<sup>3</sup> Henry D. Priest,<sup>4</sup> Olga Shevchenko,<sup>3</sup> Caiping Ma,<sup>3</sup> Elizabeth Etherington,<sup>3</sup> Larry Wilhem,<sup>4</sup> Todd Mockler,<sup>1,4</sup> Michael Freitag,<sup>1,2</sup> and **Steven Strauss**<sup>1,3\*</sup>

<sup>1</sup>Center for Genome Research and Biocomputing, <sup>2</sup>Dept. of Biochemistry and Biophysics, <sup>3</sup>Dept. of Forest Ecosystems and Society, and <sup>4</sup>Dept. of Botany and Plant Pathology, Oregon State University, Corvallis, Ore.

**Project Goals: Epigenetics is defined by long-lasting or heritable changes in gene expression that are not associated with changes in DNA sequence. It is mainly reflected in methylation of DNA and chemical changes in DNA-associated chromosomal proteins such as histones. Recognition of its importance as a means for control of plant development has increased significantly in recent years, however, little is known about epigenetic controls in the life of trees and other woody plants. Many traits important to biomass growth and adaptability in trees may be under epigenetic control, thus may be useful for their breeding and biotechnology. This includes timing of flowering and flower structure; dormancy induction and release; shoot and leaf architecture; amenability to organ regeneration; stress tolerance; and phase-associated changes in wood structure. We will use poplar (genus *Populus*, including aspens and cottonwoods), because it has been designated as a model woody biomass species for genomic studies, and is a major source of wood, energy, and environmental services in the USA and throughout the world. We will characterize epigenetic changes in DNA methylation and two kinds of histone modification via a combination of antibody-based chromatin immunoprecipitation and DNA sequencing ("ChIP-sequencing").**

Poplar (genus *Populus*) is a deciduous tree that provides a variety of environmental services and is used in many countries for agroforestry, bioenergy, pulp and wood products. Poplar is widely considered the model taxon for woody plant genomics and biotechnology. To better understand how chromatin structure varies during tree development, we are determining the distribution and variation of cytosine DNA methylation and specific histone modifications (1) in different poplar tissues, (2) during the annual dormancy cycle,

(3) during in vitro regeneration, and (4) after transgenic suppression of DNA methylation by RNAi-mediated silencing of the two predicted poplar ddm1 homologues. We collected root, leaf, bud, xylem, phloem, and callus samples, and subjected them to methylated-DNA immunoprecipitation (MeDIP) and chromatin immunoprecipitations (ChIP) followed by high throughput Illumina/Solexa sequencing. By using antibodies specific against 5-methylcytosine, trimethylated histone H3 lysine 4 (H3 K4<sup>Me3</sup>), and H3 K9<sup>Me3</sup> and H3 K27<sup>Me3</sup>, we will determine the genome-wide distribution of methylated DNA and the selected H3 marks. Except for H3 K4<sup>Me3</sup> all of these modifications are considered heterochromatic (silencing) marks. As expected, MeDIP and quantitative PCR studies of leaf chromatin have revealed high levels of cytosine DNA methylation at a variety of predicted transposable elements, and low levels of cytosine methylation at predicted active regions with high transcriptional activity. We are currently optimizing ChIP conditions and in vitro library construction for sequencing, and have begun to collect high-throughput DNA sequencing data. Epigenomic data are analyzed by a custom-developed, efficient pipeline for Solexa data analysis and by graphical representation on a poplar genome browser.

---

 USDA-DOE
An EST-Microsatellite Linkage Map of Switchgrass (*Panicum Virgatum* L.) and Comparison Within the Poaceae

**Christian Tobias**<sup>1</sup> (christian.tobias@ars.usda.gov), Miki Okada<sup>1\*</sup> (miki.okada@ars.usda.gov), Christina Lanzatella-Craig,<sup>1</sup> Brindha Narasimhamoorthy,<sup>2</sup> Malay Saha,<sup>2</sup> and Joe Bouton<sup>2</sup>

<sup>1</sup>USDA-ARS Western Regional Research Center, Calif.; and <sup>2</sup>Samuel Roberts Noble Foundation, Inc., Forage Improvement Division, Okla.

**Project Goals: Create a gene based linkage map of tetraploid switchgrass for marker-trait association.**

Switchgrass is widely viewed as a promising crop for bioenergy production. However, development of improved cultivars optimized for bioenergy through breeding involves improving yields and altering feedstock composition so that competition for limited arable land is minimized and process efficiencies are fully realized. Fundamental to any advanced breeding program are availability of molecular markers and genetic linkage maps that facilitate modern cultivar development through marker assisted selection (MAS). New crops such as switchgrass stand to benefit from the application of MAS techniques and through comparative approaches with other grasses that will provide a multitude of candidate gene-loci for traits considered important for bioenergy.

As a precondition to SNP and microsatellite development 500,000 EST were produced at the DOE Joint Genome

Institute from 10 distinct sources that included multiple genotypes and tissues. These were assembled into 74,869 consensus sequences; a large number which reflects the allelic diversity of this species. Approximately seventy percent of the assembled consensus sequences could be aligned with the sorghum genome at an *E*-value of  $<1 \times 10^{-20}$  indicating a high degree of similarity. Splice junction and coding sequence predictions were produced based on these alignments. The representation in the libraries of gene families known to be associated with C4 photosynthesis, cellulose and beta-glucan synthesis, phenylpropanoid biosynthesis, and peroxidase activity indicated likely roles for individual family members. Comparisons of synonymous codon substitutions were used to assess genome sequence diversity and indicated an overall similarity between the two genome copies present in the tetraploid.

Identification of EST-SSR markers and amplification on two individual parents of a tetraploid F1 mapping population yielded an average of 2.18 amplicons per individual and 35% of the markers produced useful fragment length polymorphisms. We have produced a microsatellite map from this F1 mapping population that is both an initial step toward QTL mapping and part of a larger project to produce an integrated genetic map that incorporates genic and intergenic markers, determines the extent of preferential pairing/disomic inheritance in switchgrass and delimits the two distinct genomes. Mapping of an advanced F2 generation which will facilitate linkage phase and haplotype determination has also been initiated. This F2 population is being established in a replicated field trial for analysis of biomass characteristics and mapping of quantitative trait loci. Thus far a total of 522 markers at 308 loci have been mapped in both parents with map lengths of 1947 cM and 2345 cM for each parent. Currently we detect 38 linkage groups organized into 9 homeologous groups. However we expect this number to stabilize at 36 when genomic SSR and EST-SSR datasets are integrated. Extensive collinearity with sorghum is apparent with the differences in base chromosome number appearing to result from the fusion of sorghum chromosomes 8 and 9.

USDA-DOE

## Fast Detection of Improved Hydrolysis in Plants with Genetically Modified Lignin

C. Chapple<sup>1</sup> (chapple@purdue.edu), R. Meilan,<sup>2</sup> X. Li,<sup>1</sup> E.A. Ximenes,<sup>3,4</sup> Y. Kim,<sup>3,4</sup> and M.R. Ladisch<sup>3,4\*</sup>

<sup>1</sup>Dept. of Biochemistry, <sup>2</sup>Dept. of Forestry and Natural Resources, <sup>3</sup>Dept. of Agricultural and Biological Engineering, and <sup>4</sup>Laboratory of Renewable Resources Engineering, Potter Engineering Center, Purdue University, West Lafayette, Ind.

**Project Goals: The phenolic cell wall polymer lignin constitutes a significant barrier to biomass conversion but, at the same time, is essential to normal plant growth and development. This research addresses efforts to**

**rationally engineer lignification in species used as energy crops in a manner that will help to maximize biomass-to-energy conversion without impacting the vital functions of lignin. The work is utilizing molecular strategies and appropriate transgenes for manipulating the expression of lignification-associated genes (Chapple lab); generating and evaluating poplar engineered for altered lignin content and/or monomer composition, and field-testing this perennial woody plant for fitness (Meilan lab); analyzing the impact of these transgenic strategies on metabolism in general and lignin biosynthesis in particular (Chapple lab); and evaluating the ease with which cell wall deconstruction can be accomplished using both chemical and enzymatic means (Ladisch lab). The research has four major objectives: 1) Generation of transgenic poplar up- or down-regulated for four enzymes known to impact lignin quantity and quality; 2) Development of metabolic profiling methods for poplar and their application to greenhouse- and field-grown wild-type and transgenic plants; 3) Field-plot morphometric analysis of transgenic lines; and 4) Cell wall deconstruction analysis of wild-type and lignin-modified transgenic lines.**

The phenol cell wall polymer, lignin, constitutes a significant barrier to biomass conversion, but at the same time is essential to normal plant growth and development. Plant genetic engineering has the potential to reduce costs for biofuel production by enabling modifications in plant cell wall structure and lignin to achieve higher digestibility of the cellulose and reducing the need for pretreatment. Pretreatment disrupts the cell wall structure so that cellulase enzymes may readily access the cellulose and break it down to glucose. This research addresses efforts to rationally engineer lignification in hybrid poplar (species *Populus*) in order to minimize the severity of the pretreatment and thereby reduce the cost and enhance the transformation of structural polysaccharides in plant cells to fermentable sugars.

The ability to control cell-wall composition without compromising plant performance is a challenge and key objective of bioenergy crop improvement. Identification of successful plant modifications entails metabolic profiling methods applied to greenhouse- and field-grown wild-type and transgenic plants. While profiling gives insights on effects of the up- or down-regulation of the four enzymes known to impact lignin synthesis, comparison of changes in genetics to changes in bioprocessing properties that enhance the production of ethanol is needed. A method for the rapid detection of biodegradability in genetically modified plants (*Arabidopsis thaliana* and *Populus*) that vary in lignin content and/or composition is addressed as part of this research. Pretreatment is carried out by pressure cooking 50 mg cellulosic material in water at 200 C for 10 min. The samples are then cooled, enzyme is added, and hydrolysis carried out for 30 min at 50 C and pH 4.8. The liquid is assayed for the amount of glucose formed. The method enables the quick evaluation of individual plants, grown in the greenhouse, since each plant weighs 50 to 70 mg, and hence one plant is equivalent to one sample. Using the reported method, we have been able to rapidly and reproducibly identify genetically modified plants that have improved biodegradability.

Extended hydrolysis times of 48 to 72 hours for *Populus* have resulted in conversions approaching 80%. These methods are identifying both *Arabidopsis* and *Populus* plants for use in further developing lignocellulosic feedstocks for biofuels production.

USDA-DOE

## Genome-Enabled Discovery of Carbon Sequestration Genes

**Gerald A. Tuskan**<sup>1</sup> (tuskanga@ornl.gov), Timothy J. Tschaplinski\* (tschaplinstj@ornl.gov), Udaya Kalluri,<sup>1</sup> Tongming Yin,<sup>1</sup> Xiaohan Yang,<sup>1</sup> Xinye Zhang,<sup>1</sup> Nancy Engle,<sup>1</sup> Priya Ranjan,<sup>1</sup> Manojit Basu,<sup>1</sup> Lee Gunter,<sup>1</sup> Sara Jawdy,<sup>1</sup> Madhavi Martin,<sup>1</sup> Alina Campbell,<sup>1</sup> Stephen DiFazio,<sup>2</sup> John Davis,<sup>3</sup> Maud Hinchee,<sup>8</sup> Christa Pinnacchio,<sup>6</sup> Rick Meila,<sup>4</sup> Victor Buso,<sup>7</sup> and Steve Strauss<sup>5</sup>

<sup>1</sup>Environmental Science Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; <sup>2</sup>University of West Virginia, Morgantown, W. Va.; <sup>3</sup>University of Florida, Gainesville, Fla.; <sup>4</sup>Purdue University, West Lafayette, Ind.; <sup>5</sup>Oregon State University, Corvallis, Ore.; <sup>6</sup>DOE Joint Genome Institute, Walnut Creek, Calif.; <sup>7</sup>Michigan Technological University, Houghton, Mich.; and <sup>8</sup>ArborGen LLC, Summerville, S.C.

**Background and goals** - The fate of carbon below ground is likely to be a major factor determining the success of carbon sequestration strategies involving plants. Despite their importance, molecular processes controlling belowground C allocation and partitioning are poorly understood. This project is leveraging the *Populus trichocarpa* genome sequence to discover genes important to C sequestration in plants and soils. The focus is on the identification of genes that provide key control points for the flow and chemical transformations of carbon in roots, concentrating on genes that control the synthesis of chemical forms of carbon that result in slower turnover rates of soil organic matter (i.e., increased recalcitrance). We propose to enhance carbon allocation and partitioning to roots by 1) modifying the auxin signaling pathway, and the invertase family, which controls sucrose metabolism, and by 2) increasing root proliferation through transgenesis with genes known to control fine root proliferation (e.g., ANT), 3) increasing the production of recalcitrant C metabolites by identifying genes controlling secondary C metabolism by a major mQTL-based gene discovery effort, and 4) increasing aboveground productivity by enhancing drought tolerance to achieve maximum C sequestration. This broad, integrated approach is aimed at ultimately enhancing root biomass as well as root detritus longevity, providing the best prospects for significant enhancement of belowground C sequestration.

**Transgenic poplar lines with altered root phenotypes** - Phenotypic analysis of root development in Aux7, Aux17, Aux16-1, 717-1B4, and transformation control were carried out as these lines had a visual aboveground phenotype.

By week 8, the number of primary and lateral roots was significantly lower in Aux7 than any other line. A number of activation-tagged poplar clones with altered rooting were also identified. Those with increased rooting included D24-4, K694-2-1, D23-12, D16-20, D18-29, and D18-16. Clones with decreased rooting included D18-84, D18-49, D23-2, E4-9, and E18-14.

### Metabolic profile of activation-tagged poplar mutants -

The metabolic profiling of activation-tagged poplar clones (roots and leaves) that displayed extremes (increase and decrease) in rooting response was completed. The profiles range from very few differences between the mutant and wild type control to a large number of extensive changes. Three of the clones with reduced rooting had a reduced concentration myo-inositol in roots. Myo-inositol was shown to be highly correlated with rooting in poplar stem cuttings (Tschaplinski and Blake 1989). The clones with increased rooting tended to have increases in aspartic acid and other amino acids in the roots, as seen in ANT mutants with increased rooting. This mutant also had large accumulations of complex phenolic glycosides at the expense of lower MW phenolics. Although some responses are common among the mutants, each mutant needs to be considered separately. The bases for the underlying changes will be elucidated by recovering the 35S promoter region using TAIL-PCR and sequencing in both directions to identify the genes that may be directly or indirectly contributing to the observed profiles.

**Metabolite QTL and underlying candidate genes** - Several secondary carbon metabolite quantitative trait loci (mQTL) were identified in a hotspot on poplar LG X. The first five candidate genes underlying the identified mQTL were selected. Constructs for up and down regulation were created and submitted to ArborGen for transformation into *P. deltoides*. We anticipate receiving these transformants early next year.

**Productivity QTL** - A primary focus of this project has been the examination of the genetic control of biomass allocation above and below ground. We examined controls of productivity in TDxD' Family 52124 containing over 800 hybrid clones. We developed a hydroponic system that provides robust growth and used this system to assess biomass allocation in 252 genotypes from this pedigree using two replicates in a completely randomized design. The clonal repeatability, or genotypic broad-sense heritability, ranged from 0.286 to 0.333 for three biomass-allocation traits, indicating that the phenotypes are repeatable between replicates. These traits are currently being mapped as QTL, utilizing a genetic map developed for this pedigree.

**Poplar drought tolerance** - Cisgenic *Populus deltoides* clones were down- and up-regulated with a constitutive promoter with drought tolerance genes identified in QTL intervals. The genes selected were all known transcription factors that were previously identified in other species. An up-regulated popMYB clone was 16% more productive than the wild-type control and 54% more productive than the empty vector control. A number of the transgenic lines had unusual growth forms, exhibiting a loss of erect growth habit as a result of the constitutive promoter. A second round of trans-

genic plants was created using a drought-inducible promoter identified from qRT-PCR analyses as highly responsive in a controlled drought experiment, increasing 27238-fold in a *P. deltoides* clone under severe drought stress. A global transcript profile is being conducted in *P. deltoides* that will identify novel drought response pathways. Enhanced drought tolerance will ensure sustained productivity and maximum C sequestration potential.

This research was supported by the Office of Biological and Environmental Research of the U.S. Department of Energy. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725.

USDA-DOE

## Genetic Dissection of Bioenergy Traits in Sorghum

**Wilfred Vermeris**\* (wev@ufl.edu), Ana Saballos,<sup>1</sup> Zhanguo Xin,<sup>2</sup> Jeffery Pedersen,<sup>3</sup> Scott Sattler,<sup>3</sup> William Rooney,<sup>4</sup> Seth Murray,<sup>4</sup> and Stephen Kresovich<sup>5</sup>

<sup>1</sup>University of Florida, Gainesville, Fla.; <sup>2</sup>USDA-ARS, Lubbock, Tex.; <sup>3</sup>USDA-ARS, Lincoln, Nebr.; <sup>4</sup>Texas A&M University, College Station, Tex.; and <sup>5</sup>Cornell University, Ithaca, N.Y.

**Project Goals:** The overall goal of the project is to identify the genetic basis of a number of traits that enhance the value of sorghum as a bioenergy crop. These traits relate to the production of fermentable sugars from the sugar-rich juice in the stems of sweet sorghums and from the cell wall polysaccharides in sorghum stover or bagasse. The specific objectives are to: 1) identify the gene(s) underlying a major QTL for stem sugar concentration, 2) identify QTL for stem juice volume and stalk sugar concentration, 3) classify approximately 60 novel sorghum brown midrib (*bmr*) mutants from the USDA TILLING population in allelic groups based on cell wall chemistry and allelism tests, 4) select representative *bmr* mutants from each allelic group for their potential value as feedstock for ethanol production, and 5) clone and characterize those *Bmr* genes that represent loci other than the two cloned *Bmr* genes, using a mapping and a candidate gene approach.

Sorghum (*Sorghum bicolor* (L.) Moench) is a multi-source bioenergy crop that can produce grain, sugar-rich juice and lignocellulosic biomass as feedstocks for fermentable sugars for the production of liquid transportation fuels and green chemical feedstocks. Alternatively, production schemes in which some of the products are used for other purposes, including animal feed or biomass for the generation of heat or electricity are possible. Sorghum requires lower inputs than maize and sugar cane and is tolerant of a wide range in temperature and soil composition.

We are investigating the genetic basis of a number of traits in order to enhance their utility and facilitate plant breeding efforts. The *brown midrib* (*bmr*) mutations affect the chemi-

cal composition of the cell wall. Some of the *bmr* mutations have been shown to result in at least a 50% increase in the yields of glucose and xylose after thermo-chemical pretreatment and enzymatic saccharification of sorghum stover. There are at least four *Bmr* genes, two of which have been cloned and shown to encode enzymes in the monoglucosyl biosynthetic pathway. An EMS-mutagenized population containing close to 60 *bmr* mutants generated by the USDA-ARS is being used to identify additional *bmr* loci and novel alleles of already established loci, and to establish the function of the *Bmr* genes that have not yet been cloned.

The stems of sweet sorghums contain sugar-rich juice that can be used directly for fermentation. Total sugar yield is determined by juice volume and sugar concentration. We are using a recombinant inbred line (RIL) population derived from a cross between a sweet sorghum and a dry-stalk, non-sweet sorghum to map quantitative trait loci (QTL) involved in determining juice volume. Prior research using a different RIL population led to the identification of a major QTL for sugar concentration in sweet sorghum. We are using genetically similar lines with contrasting genotypes for this QTL derived from this population as a basis for high-throughput expression profiling based on the 454 sequencing platform to identify the gene(s) in this QTL.

Our combined efforts are expected to yield novel insights in the genetic control of primary metabolism in sorghum, and result in genetic markers that will greatly expedite the development of the next generation of bioenergy sorghums.

USDA-DOE

## Insertional Mutagenesis of *Brachypodium distachyon*

**John Vogel**\* (john.vogel@ars.usda.gov), Jennifer Bragg, Jiajie Wu, Yong Gu, Gerard Lazo, and Olin Anderson

USDA-ARS, Western Regional Research Center, Albany, Calif.

**Project Goals:** 1) Generate >7,500 insertional mutants in the model grass *Brachypodium distachyon*; 2) Sequence DNA flanking >6,000 insertion sites and annotate the genes affected using the genomic sequence; and 3) Establish a website where researchers could search the flanking sequence database and order knockouts in genes of interest.

*Brachypodium distachyon* (*Brachypodium*) is an excellent model for studying the basic biological processes underlying the traits that determine the utility of grasses as energy crops. Several important genomic resources have been developed in *Brachypodium*, including the recently completed whole genome sequence, ESTs, a high-density genetic linkage map and germplasm resources. The objective of our work is to generate >7,500 insertional mutants and to sequence the regions flanking >6,000 insertion sites. The location of insertions in the genome will be determined



by comparing flanking sequences to the complete genome sequence. This information will be loaded into a searchable website to provide researchers with a means to order T-DNA lines with mutations in genes of interest. Thus, this project will provide a large, freely available collection of sequence-indexed mutants to researchers studying grasses and grains. This poster describes our experiments aimed at optimizing transformation, evaluating transposon tagging systems and sequencing flanking DNA. As a first step toward designing constructs for large scale mutagenesis, we compared several vectors and observed considerable variation in efficiency. The average transformation efficiency (from 6 independent transformations) of the best construct was 55% with a mean survival rate of 88% after the plants were transferred to soil. We improved our transformation method by modifying the callus initiation media and omitting a recovery step. These changes increased transformation efficiency, decreased labor by eliminating a transfer step and decreased the time needed to create transgenics. To date, we have generated >1,300 T<sub>0</sub> T-DNA mutant lines. Over 95% of T<sub>0</sub> plants tested showed expression of a GUS reporter gene, and all were positive by PCR for the presence of the selectable marker. T<sub>1</sub> seed has been harvested for 450 lines and 225 lines have been planted for phenotypic evaluation. To determine if transposon tagging is more efficient than T-DNA tagging we evaluated both Ac/Ds and En/Spm transposon systems in *Brachypodium*. Both transposons were lethal to the majority of transgenic plants indicating that neither transposon system is compatible with *Brachypodium*. We are currently comparing TAIL-PCR and inverse-PCR to sequence flanking DNA. Initial experiments suggest that TAIL-PCR is less efficient than inverse-PCR. To date, we have generated 79 flanking sequences from 121 T<sub>0</sub> plants using inverse-PCR. Of these sequences, 45 (57%) contain *Brachypodium* genomic sequence, and 34 (43%) contained only vector sequence. We are currently transitioning from optimization to production scale generation of T-DNA lines.

USDA-DOE

## Identification of Genes That Control Biomass Production Using Rice as a Model System

**Jan E. Leach**<sup>1\*</sup> (jan.leach@colostate.edu), Daniel R. Bush<sup>2,3</sup> (dbush@colostate.edu), Hei Leung<sup>3</sup> (h.leung@cjar.org), and John McKay<sup>1,3</sup> (jkmckay@lamar.colostate.edu)

<sup>1</sup>Bioagricultural Sciences and Pest Management and <sup>2</sup>Department of Biology, Colorado State University, Ft. Collins, Colo.; and <sup>3</sup>International Rice Research Institute, Manila, Philippines

**Project Goals: Our goal is to provide the applied biomass research community and industry with information to allow exploitation of the genes and pathways relevant to biomass accumulation in grasses.**

Phenotypic variation in cellulosic biomass accumulation has not been systematically investigated in any plant species. We propose a critical first step to identifying the genotypic variation that controls biomass productivity, using rice as a model grass. Using the genetic and genomic tools available in rice, we aim to establish causal links between the genes (or genetic regions in the case of multigenic traits), gene expression profiles, and phenotype (biomass accumulation). Identifying the loci that are essential for optimal cellulosic biomass in rice will provide a set of loci and networks that are molecular targets for improving target biomass crops (e.g., *Miscanthus*, switchgrass). We have identified three lines of rice that span the variation in biomass accumulation and for which we have comprehensive SNP data. One population has already been advanced to 700 F6 recombinant inbred lines (RILs); development of the second population is in progress. These genetic resources will be used to dissect the genetic basis of biomass accumulation.

Using these two populations, we will (1) identify QTL for biomass accumulation phenotypes, (2) refine these QTLs using an integrated analysis of variation at both the DNA and expression levels as well as analysis of mutant populations, and (3) validate gene candidates for biomass traits using mutant analysis, gene silencing strategies or overexpression analysis. Detailed physiological and morphological descriptions of the parents, high-biomass RILs and mutants, and transgenic plants used in validation experiments will parallel genetic analysis. Gene expression profiles will be generated from near-isogenic lines (NILs) at two growth stages in a single season to identify genomic regions where gene expression of clustered genes is correlated with phenotype. All data will be integrated into a convenient gene browser to facilitate identification of gene candidates for the biomass traits. Finally, we will validate the significance of a suite candidate genes or pathways in biomass accumulation by mutant analysis, gene silencing (RNAi), and/or overexpression. Screening and field assessments as well as data integration will involve exchange of PIs and postdoctoral fellows between CSU and IRRI. Deliverables from this project include: (1) Characterized genomic regions in rice that are associated with variation relevant to biomass production, (2) Permanent rice genetic stocks (e.g., RILs and NILs) that will enable further exploitation of biomass diversity and other desirable traits, (3) selected candidate genes associated with biomass production, (4) validation of function of genes predicted to be involved in biomass production.

In the first few months of funding, we have completed a comprehensive phenotyping for biomass traits (plant size and architecture, photosynthetic rates and assimilate partitioning) for the three rice lines that serve as parents for the RIL populations, and based on these results, have initiated phenotypic screening of the available RIL population. Development of the second RIL population is underway. A comprehensive SNP dataset and phenotypic and genetic mapping information are being used to associate genotype and phenotype and identify candidate biomass genes. Over 15 putative mutants altered in biomass accumulation (large plants) have been identified through phenotypic screens of

a deletion collection. After confirmation of genotype, DNA from these mutants will be hybridized to rice oligonucleotide arrays for discovery of deleted regions.

USDA-DOE

## Resource Development in Switchgrass, a Polyploid Perennial Biofuel Grass

**Katrien M. Devos**<sup>1,2\*</sup> (kdevos@uga.edu), Jeffrey L. Bennetzen,<sup>3</sup> Joe H. Bouton,<sup>4</sup> E. Charles Brummer,<sup>1</sup> and Malay C. Saha<sup>4</sup>

<sup>1</sup>Institute of Plant Breeding, Genetics, and Genomics, Dept. of Crop and Soil Sciences, <sup>2</sup>Dept. of Plant Biology, and <sup>3</sup>Dept. of Genetics, University of Georgia, Athens, Ga.; and <sup>4</sup>Forage Improvement Division, The Samuel Roberts Noble Foundation, Ardmore, Okla.

**Project Goals:** This project aims to develop tools such as DNA markers, genetic maps and trait maps to enhance the breeding of switchgrass for bioenergy production. We are also interested in gaining insight in the structural organization of the switchgrass genome, and understanding the relationship of the switchgrass genome with that of the model species foxtail millet as a foundation for the future sequencing of the complex switchgrass genome.

Switchgrass, *Panicum virgatum*, is a perennial grass that is native to the United States. It is a polyploid that has been bred as a forage crop for more than a decade. Its high biomass potential under low inputs also makes switchgrass a good candidate biofuel crop. The efforts to develop switchgrass as a dedicated biofuel crop have gained impetus in the past few years with the funding of several collaborative projects to develop genetic tools and populations for trait mapping. Our project is focused on the development of resources for switchgrass, including SSR markers, genetic maps, trait maps, a fosmid library and a limited amount of sequence information to get insight into the structural organization of the switchgrass genome. Mapping will be carried out in a biparental population developed from a cross between a lowland Alamo genotype and an upland Summer genotype. This population has been planted in two locations and a first round of phenotyping has been carried out in the fall of 2008. Genetic mapping is in progress. Markers for mapping include EST-derived and genomic simple sequence repeats (SSRs), candidate genes for biomass yield and, potentially, single nucleotide polymorphism (SNP) markers. The genetic maps will be aligned with the maps of other grass species, in particular foxtail millet. Foxtail millet is a small-genome relative of switchgrass whose whole genome sequence is expected to become available in Spring 2009 and will act as a model for switchgrass genomics. The switchgrass and comparative tools will greatly accelerate the breeding of switchgrass varieties targeted at enhanced bioenergy conversion potential.

## Identification of Genes that Regulate Phosphate Acquisition and Plant Performance During Arbuscular Mycorrhizal Symbiosis in *Medicago truncatula* and *Brachypodium distachyon*

S. Karen Gomez,<sup>1</sup> Nathan Pumplin,<sup>1</sup> Aynur Cakmak,<sup>1</sup> Jeon Hong,<sup>1</sup> Ying Li,<sup>2</sup> Matthew E. Hudson,<sup>2</sup> and **Maria J. Harrison**<sup>1\*</sup> (mjh78@cornell.edu)

<sup>1</sup>Boyce Thompson Institute for Plant Research, Ithaca, N.Y.; and <sup>2</sup>Dept. of Crop Sciences, University of Illinois, Urbana, Ill.

**Project Goals:** The overall goals are to identify and characterize plant genes involved in the regulation and functioning of the AM symbiosis in *Medicago truncatula* and *Brachypodium distachyon*. The specific objectives are: 1) Identify and characterize *M. truncatula* transcription factors expressed in mycorrhizal roots. 2) Identify transcription factors required for development of the arbuscule-cortical cell interface and symbiotic phosphate transport. 3) Investigate the molecular basis of functional diversity among AM symbioses in *B. distachyon* and identify gene expression patterns associated with AM symbioses that differ functionally with respect to plant performance. 4) Analyze the predicted *B. distachyon* ortholog of MtPT4 and its role in symbiotic phosphate transport.

Most vascular flowering plants are able to form symbiotic associations with arbuscular mycorrhizal (AM) fungi. The symbiosis develops in the roots and has a profound effect on plant productivity, largely through improvements in plant mineral nutrition (1). All proposed bioenergy crops including legumes, grasses and trees, are capable of forming AM symbioses, and therefore have the potential to benefit from phosphorus and nitrogen acquisition through the symbiosis. This is significant because phosphorus and nitrogen are the two mineral nutrients whose availability is most frequently limiting for plant growth (2).

To ensure that future crops contain the optimal set of alleles to benefit maximally from an AM symbiosis, it is necessary first to understand which genes control the AM symbiosis and the molecular basis of their function. To address this we aim to identify plant transcription factors that regulate development of the AM symbiosis and symbiotic phosphate transfer in *Medicago truncatula*. In mycorrhizal roots, symbiotic phosphate transfer occurs in the cortical cells harboring fungal arbuscules and these cells comprise only a small proportion of the total cortical cell population within the root. To identify transcription factors operating in this cell type, we have developed laser capture microdissection to enable the capture of cells containing arbuscules. Analysis of transcription factor gene expression in this cell type is in progress. In parallel, we are taking advantage of genomics resources developed in *M. truncatula* to enable bioinformatics approaches to identify potential cis-regulatory motifs

associated with promoters of AM symbiosis-induced genes (3).

*M. truncatula* is widely used as a model for studies of AM symbiosis but it is not an ideal model for analysis of AM symbiosis in grass species. *Brachypodium distachyon* is a wild grass species that serves as a model for the temperate grasses, including those proposed as bioenergy crops. There are several detailed ecological studies of AM symbioses of grasses, including studies of the genus *Brachypodium* (4,5), and these indicate the potential of the AM symbiosis for increasing plant growth in low phosphate soils. However, these studies also illustrate significant differences in plant performance depending on the AM fungal species involved. Variation in plant performance during symbiosis with different AM fungal symbionts is a well documented phenomenon; however, the molecular basis is not understood. We aim to use *B. distachyon* to investigate this and will use sequence-based transcript profiling and statistical analyses to compare transcript profiles in *B. distachyon* AM symbioses that differ in their effect on plant performance. The goal is to identify transcript profiles and potentially plant processes associated with maximal plant performance. Experiments to evaluate the response of *B. distachyon* to different AM fungi are in progress.

## References

1. Javot H, Penmetsa RV, Terzaghi N, Cook DR, Harrison MJ. 2007. A *Medicago truncatula* phosphate transporter indispensable for the arbuscular mycorrhizal symbiosis. *PNAS* 104: 1720-5
2. Vance C. 2001. Symbiotic Nitrogen Fixation and Phosphorus Acquisition. *Plant Nutrition in a World of Declining Renewable Resources. Plant Physiology* 127: 390-7
3. Hudson ME, Quail PH. 2003. Identification of promoter motifs involved in the network of phytochrome A-regulated gene expression by combined analysis of genomic sequence and microarray data. *Plant Physiology* 133: 1605-16
4. van der Heijden MGA, Wiemken A, Sanders IR. 2003. Different arbuscular mycorrhizal fungi alter coexistence and resource distribution between co-occurring plant. *New Phytologist* 157: 569-78
5. van der Heijden MGA, Streitwolf-Engel R, Riedl R, Siegrist S, Neudecker A, et al. 2006. The mycorrhizal contribution to plant productivity, plant nutrition and soil structure in experimental grassland. *New Phytologist* 172: 739-52

## Biofuels > Analytical and Imaging Technologies for Studying Lignocellulosic Material Degradation

GTL

### Three-Dimensional Spatial Profiling of Lignocellulosic Materials by Coupling Light Scattering and Mass Spectrometry

P.W. Bohn<sup>1\*</sup> (pbohn@nd.edu) and J.V. Sweedler<sup>2</sup> (sweedler@scs.uiuc.edu)

<sup>1</sup>Dept. of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, Ind.; and <sup>2</sup>Dept. of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Ill.

**Project Goals:** The project goals are to combine Raman imaging and MALDI and SIMS mass spectrometric imaging to accomplish four main objectives: (1) develop correlated optical and mass spectrometric imaging approaches for sub-surface imaging to address the highly scattering nature of LCMs in different states of processing; (2) adapt mass spectrometric imaging protocols to enable spatially-resolved LCM characterization; (3) create surface optical and mass spectrometric measures of lignin-hemicellulose-cellulose degradation at specific processing stages; and (4) correlate the in situ optical (Raman and SH-OCT) and mass spectrometric information to generate depth-resolved maps of chemical information as a function of spatial position and processing time.

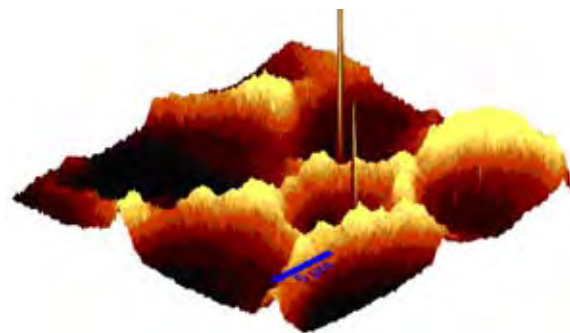


Figure. 3D rendering of lignin from untreated *Miscanthus*.

The physical and chemical characteristics of lignocellulosic materials (LCMs) pose daunting challenges for imaging and molecular characterization: they are opaque and highly scattering; their chemical composition is a spatially variegated mixture of heteropolymers; and the nature of the matrix evolves in time during processing. Any approach to imaging these materials must (1) produce real-time molecular

speciation information *in situ*; (2) extract sub-surface information during processing; and (3) follow the spatial and temporal characteristics of the molecular species in the matrix and correlate this complex profile with saccharification. To address these challenges we are implementing tightly integrated optical and mass spectrometric imaging approaches. Employing Raman microspectroscopy (RM) provides real-time *in situ* information regarding the temporal and spatial profiles of the processing species and the overall chemical degradation state of the lignin heteropolymer; while MALDI and SIMS provide spatially-resolved information on the specific molecular species produced by pre-enzymatic processing. The goals of the approach are to: (1) develop correlated optical and mass spectrometric imaging approaches for sub-surface imaging to address the highly scattering nature of LCMs in different states of processing; (2) adapt mass spectrometric imaging protocols to enable spatially-resolved LCM characterization; (3) create surface optical and mass spectrometric measures of lignin-hemicellulose-cellulose degradation at specific processing stages; and (4) correlate the *in situ* optical (Raman and SH-OCT) and mass spectrometric information to generate depth-resolved maps of chemical information as a function of spatial position and processing time. Raman and MALDI/SIMS imaging results from native (unprocessed) and acid-treated *Miscanthus x giganteus* will be described. Approaches for generating chemically specific information in plan view and three-dimensional profiling (*viz.* figure) will be also be presented.

GTL

## Visualization of Acid-Pretreatment Effects on Lignocellulose by Integration of Neutron Scattering and Computer Simulation

**Barbara R. Evans**<sup>1\*</sup> (evansb@ornl.gov), Jeremy C. Smith,<sup>2,3</sup> Loukas Petridis,<sup>2</sup> Volker Urban,<sup>1,4</sup> Sai Venkatesh Pingali,<sup>1,4</sup> Dean A. Myles,<sup>1,4,5</sup> William T. Heller,<sup>1,4</sup> Hugh M. O'Neill,<sup>1,4</sup> Joseph McGaughey,<sup>1</sup> Art Ragauskas,<sup>6</sup> Marcus Foston,<sup>6</sup> and Ida Lee<sup>1,7</sup>

<sup>1</sup>Chemical Sciences Division and <sup>2</sup>Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, Tenn.; <sup>3</sup>Governor's Chair, University of Tennessee, Knoxville, Tenn.; <sup>4</sup>Center for Structural Molecular Biology and <sup>5</sup>Neutron Scattering Science Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; <sup>6</sup>Institute of Paper Science and Technology, School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Ga.; and <sup>7</sup>Dept. of Electrical and Computer Engineering, University of Tennessee, Knoxville, Tenn.

**Project Goals: The overarching goal is the integration of neutron scattering techniques, surface force imaging, and computer simulation technology to provide *in situ* real-time multi-scale visualization of lignocellulose structure. The comprehensive molecular and nanoscale information thus obtained will enable the targeted improvement**

**of biomass deconstruction and saccharification that is needed for efficient cellulosic ethanol production. This new imaging technology will integrate the continuously evolving power and capabilities of neutron scattering and high-performance leadership computer simulation at Oak Ridge National Laboratory to synergistically derive information on lignocellulosic degradation at an unprecedented level of detail.**

A detailed physicochemical understanding of lignocellulose structure and its deconstruction during processing at multiple length scales is needed to enable cost-effective production of fuels from biomass. To attain this goal, we are developing a real-time *in situ* multi-length scale neutron scattering-computer simulation technology with complimentary surface force recognition imaging. Key to this effort is the design and employment of multipurpose neutron imaging chambers that will enable *in situ* dynamic observation of the structural evolution of biomass under pretreatment conditions. This visualization capability resulting from the combination of neutron scattering and high-performance computer simulation afforded through facilities at ORNL will provide fundamental information about biomass morphological degradation during heat, pressure, chemical or enzymatic treatment at an unprecedented level of detail. The synergistic capabilities of the Spallation Neutron Source (SNS), the High Flux Isotope Reactor (HFIR) and the National Leadership Computing Facility (NCLF) at ORNL will be combined with supporting information from biochemistry, X-ray diffraction, small-angle x-ray scattering, and physical and chemical characterization of lignocellulosic materials through surface force recognition imaging. This novel combination of experimental and computational techniques will produce the fundamental molecular scale understanding of biomass degradation required to develop cost-effective approaches to cellulosic ethanol production.

We characterized samples of standard biomass feedstocks of loblolly pine, switch grass, and hybrid poplar provided by the Institute of Paper Science and Technology (IPST) at Georgia Tech. The IPST provided technical background on composition of the lignocellulose in the feed stock plants and samples of the component biopolymers cellulose, hemicellulose, and lignin that provided the separate scattering signatures of the components to assist in the analysis of the scattering and diffraction patterns of whole lignocellulose. Small-angle neutron scattering (SANS) experiments were carried out on the Bio-SANS beamline at the ORNL High Flux Isotope Reactor (HFIR). In the preliminary experiments, untreated biomass samples loblolly pine, switch grass, and hybrid poplar, and the individual components extracted from these samples were examined. Matched sets of the isolated components cellulose, hemicellulose, and lignin were prepared at IPST by standard methods. The scattering patterns of these components are being used to assist in analysis of the patterns obtained from the corresponding whole plant material. Additionally, the structural consequences of dilute sulfuric acid pretreatment of switch grass and poplar were studied by measuring a pretreatment reaction-time series. These results are being used to determine the parameters required for design and use of the real-time visualization

chambers. The results suggest that very little changes in the sample over the length scales investigated during this standard pretreatment process. A force field for molecular dynamics of cellulose has been developed (Petridis and Smith, *J. Comp. Chem.*, 2008, in print) and atomistic models of lignin and cellulose were constructed and combined to form an initial model for lignocellulose. Molecular dynamics simulation is performed to examine structural properties of the system. The results of the simulation are compared with SANS data and the computational models are further refined to achieve better agreement with experiments. Development of a simplified modeling approach for the large scale structural organization of lignocellulose has begun. Switch grass, wheat, and other cellulose sources are being cultivated in deuterated media to provide perdeuterated lignocellulose that will enable higher contrast with neutron scattering and examination of surface accessibility.

This work is funded by the U. S. Department of Energy, Office of Biological and Environmental Research Genomics:GTL Program under FWP ERKP704, "Dynamic Visualization of Lignocellulose Degradation by Integration of Neutron Scattering Imaging and Computer Simulation".

GTL

## Synchrotron Infrared Spectromicroscopy of Cellulose Degradation Strategies of Living Cellulolytic Bacteria

Hoi-Ying N. Holman<sup>1\*</sup> (hyholman@lbl.gov), Zhao Hao,<sup>1</sup> Ping Hu,<sup>1</sup> Robin Miles,<sup>2</sup> Albert Pisano,<sup>3</sup> Inkyu Park,<sup>3</sup> and Sun Choi<sup>3</sup>

<sup>1</sup>Chemical Ecology and Spectromicroscopy Group, Lawrence Berkeley National Laboratory, University of California, Berkeley, Calif.; <sup>2</sup>Microfluidics and Chemical/Bio Devices Group, Lawrence Livermore National Laboratory, Livermore, Calif.; and <sup>3</sup>Micro-Electromechanical Systems Group, University of California, Berkeley, Calif.

**Project Goals:** Our goal is to provide a new research tool for investigating, directly and in a single experiment, the conversion of biomass from its deconstruction to its conversion to ethanol with unparalleled resolution. To achieve this goal, we have designed technology research and development efforts to meet the following specific aims. 1) to develop an integrated set of non-invasive micro-analytical and imaging technologies to directly characterize lignocellulose degradation by microorganisms in situ; 2) to use these technologies to advance the fundamental understanding of lignocellulose degradation by currently known microorganisms and microbial communities; and 3) to design and confirm the strategy for integrating the SIR technology of live bacterial cells and communities with 3D tracking fluorescence microscopy of single molecules.

We are developing a synchrotron radiation-based Fourier transform infrared (SR-FTIR) spectromicroscopy technology with uninterrupted and non-invasive *in vivo* tracking capabilities. Progress in sequencing and in structural and functional genomics of cellulolytic bacteria must be followed by studying both their strategies of actions on solid substrates such as lignocellulose, inhibitory effects by end products of cellulose degradation, and their adaptation response to toxicity from the fermentation products, such as ethanol. Non-invasive technologies to study such lignocellulose-bacteria systems are not well established.

Our goal is to provide a new research tool for investigating, directly and in a single experiment, the conversion of biomass from its deconstruction to its conversion to ethanol with unparalleled resolution. This technology will have a temporal resolution of seconds-to-minutes and a spatial resolution of several micrometers. Essential to this endeavor is the development of experimental technologies and methods that are robust and allows for the comprehensive measurement and analysis of lignocellulose degradation and the definition of the microbial role. A microfluidic incubation platform and a multi-phasic array of nanowire sensors are being developed to provide a controlled experimental environment that can probe cellular chemistry and manipulate experimental conditions, rather than simply observing processes. Microbial cells and their functions are sensitive to immediate environmental conditions. Genetic strategies are being developed for using model microbes that allow direct confirmation of their viability, metabolic activity, and normal functions inside the microfluidic system during their exposure to SR-FTIR spectromicroscopy measurements.

Today, we are employing lignocellulose-conversion by the mesophilic *Clostridium cellulolyticum* as our model system with which to develop and optimize the selection and tracking of spectral markers, and the performance of synchrotron infrared spectromicroscopy technologies. *C. cellulolyticum* is a representative of anaerobic, mesophilic, cellulolytic bacterium which producing complex cellulose systems (cellulosomes). The cellulosome is thought to allow concerted enzyme activity in close proximity, enabling optimum synergism between the cellulases. An easier DNA transformation system and more biochemical and physiological data make the study of *C. cellulolyticum* valuable. Since the genetic systems are not yet well-established in *Clostridia*, reporter strains are not readily available, and gene/pathway modifications are not easily done. To thoroughly test the integrity and functions of these living cells in the SIR technologies, we are using *E. coli* reporter strains to provide complementary information.

GTL

for biomass utilization, that the resulting oxidized polymer does not need to be removed for sugars to be released.

## A New Solution-State NMR Approach to Elucidate Fungal and Enzyme/Mediator Delignification Pathways

**Kenneth E. Hammel**<sup>1\*</sup> (kehammel@wisc.edu), Daniel J. Yelle,<sup>1</sup> John Ralph,<sup>2</sup> Dongsheng Wei,<sup>1</sup> and Alexander N. Kapich<sup>1</sup>

<sup>1</sup>USDA Forest Products Laboratory, Madison, Wis.; and <sup>2</sup>Dept. of Biochemistry, University of Wisconsin, Madison, Wis.

**Project Goals:** The mechanisms that lignocellulose-degrading fungi employ to bypass lignin and access the polysaccharides in plant biomass are not well understood, in part because the chemical changes they cause in lignin have not been well characterized. We are applying new methods for the complete dissolution of lignocellulose to identify diagnostic degradative reactions in lignin by solution-state NMR spectroscopy.

Brown rot basidiomycetes remove cellulose from wood efficiently, even though this sugar polymer is initially shielded by a biochemically recalcitrant barrier of lignin. During this process, the lignin appears to remain in situ, which raises the question of how the polysaccharide-degrading systems of brown rot fungi circumvent the lignin to access their substrates. New results based on solution-state NMR analysis of ball-milled, dissolved, brown-rotted wood are now available to clarify this picture. We obtained short-range <sup>1</sup>H-<sup>13</sup>C (HSQC) spectra of aspen degraded by the brown-rotter *Postia placenta*, and found that the residual material was about 25% deficient in the major arylglycerol-β-aryl ether structure of lignin, relative to the methoxyl content of the sample. To identify some of the processes responsible for lignin depletion, we next performed additional NMR analyses on decayed, ball-milled aspen that had been treated with a cellulase mixture to remove polysaccharides. A <sup>13</sup>C NMR spectrum of this enzyme-treated sample showed that it contained benzoic acid and benzaldehyde residues that were absent in the undecayed aspen. Moreover, a short-range <sup>1</sup>H-<sup>13</sup>C (HSQC) spectrum of the same sample displayed signals characteristic of phenylglycerol residues that were absent in the undecayed wood. The identity of the benzoic acids and benzaldehydes was confirmed in a long-range <sup>1</sup>H-<sup>13</sup>C (HMBC) spectrum, and the identity of the phenylglycerols was confirmed in a total correlation spectroscopy (TOCSY) experiment. The presence of benzoic acid and benzaldehyde residues in the brown-rotted wood shows that its lignin sidechains had been cleaved between C<sub>α</sub> and C<sub>β</sub>, whereas the presence of phenylglycerols shows that intermonomer ether linkages between lignin sidechains had been cleaved. Our results demonstrate that *Postia placenta* is actually ligninolytic, contrary to the prevailing view of fungal brown rot. Presumably, ligninolysis is required for fungal polysaccharide hydrolases to operate effectively, but it is interesting, in light of current efforts to improve methods

GTL

## In Vivo Mapping of ROS Produced by Wood Decay Fungi during Early Colonization

**Christopher G. Hunt**<sup>1\*</sup> (cghunt@fs.fed.us), Kenneth E. Hammel<sup>1,2</sup> (kehammel@wisc.edu), and Don Carl Jones<sup>1</sup>

<sup>1</sup>U.S. Forest Service Forest Products Laboratory; and <sup>2</sup>University of Wisconsin, Madison, Wis.

**Project Goals:** Develop a modular system to image and quantify ROS (or other metabolite) from fungi in their natural substrate with minimal disturbance. Use the knowledge of ROS concentrations to understand the mechanisms at work during incipient decay.

### Background

Wood decay fungi are successful in removing carbohydrate from wood, despite the presence of lignin. We hope to improve prospects for woody biomass saccharification by understanding the mechanisms of fungal wood decay.

To make lignocellulosics permeable to enzymes, filamentous fungi use a variety of small, diffusible reactive oxygen species (ROS) such as hydroxyl radicals, peroxy radicals, and possibly phenoxy radicals. These radicals diffuse into the cell walls and initiate biodegradative radical reactions. When lignin is the target, radical attack results in various extents of oxidation and depolymerization.

These small diffusible oxidative species are important tools used by filamentous fungi to make the cell wall accessible to enzymes. Despite this, we have a poor knowledge of how these oxidants are spatially distributed in biodegrading lignocellulose relative to the fungal hyphae that produce them. The goal of this project is to remedy this deficit through fluorescence microscopy of newly designed sensors that will serve as in situ reporters of biodegradative radical production. We will use these sensors to produce oxidative maps that will help us to understand how fungi generate ROS and how they use these ROS to make cell walls more accessible to enzymes.

### Method

We are attaching fluorescent dyes to silica beads. Our first bead has BODIPY 581/591<sup>®</sup> on a 3μm porous HPLC bead. This dye's emission changes irreversibly from red to green upon oxidation by ROS. The ratio of red to green emission provides a quantitative measure of the cumulative oxidation at that point in space. Dyes with reactivity to specific ROS, pH, or other metabolites of interest are envisioned.

There are many advantages gained by fixing the dye to bead. We design the bead to emit two fluorescent signals, so that the ratio of the two signal intensities provides quantitative information. Immobilized dyes are prevented from moving

after reaction, so partitioning is impossible, they cannot be ingested, and the fluorescence from the dye is clearly distinguishable from background.

Beads are placed on wood samples and imaged with a confocal microscope during fungal colonization. Images can be analyzed to provide the analyte concentration maps as well as an overlay of the location of fungal hyphae.

## Results

We see large changes in the fluorescence of our beads with fungal oxidation, as shown in 1. Ideally, we would be able to see the ROS gradient around a single hypha. We developed a silica gel culture system so that the fungus could not metabolize the water agar gel and multiply too quickly. We have succeeded in observing an ROS gradient around individual hypha, as shown in Figure 2. Individual beads show oxidation through a lower average ratio of red to green emission, and by displaying a pattern of oxidized pixels on the perimeter while the core of the bead has higher red to green ratios.



**Figure 1.** Fluorescent beads no fungus (left) and 3 days after inoculation with *Phanerochaete chrysosporium* (center). Note change in bead color with oxidation. Transmission image shows fungal hyphae (Right) Wood cellular structure is visible via auto-fluorescence (center).



**Figure 2.** Transmission image (left) of fungal hyphae growing over gel adjacent to wood. Oxidized bead is circled and unoxidized bead is squared. Unoxidized bead (center photo) has bright and dark pixels, representing degree of oxidation, randomly distributed. Oxidized bead (right) is most oxidized (darkest pixels) around perimeter while less oxidized light pixels are all in center. Note the average pixel value is also lower.

## Summary

We are able to measure oxidation produced by colonizing hyphae at a micron scale. We are currently generating micron scale oxidation maps to determine the ROS gradient as a function of distance. Next we plan to compare the oxidative profile of brown rot, white rot, and nondecay fungi.

## Stimulated Raman Scattering as an Imaging Tool for Lignocellulosic Biomass Conversion

Brian G. Saar, Yining Zeng, Christian W. Freudiger, Marcel G. Friedrich, Gary R. Holtom, Shi-you Ding, and **X. Sunney Xie\*** (xie@chemistry.harvard.edu)

Harvard University, Cambridge, Mass.

**Project Goals: To improve the conversion efficiency of ligno-cellulosic biomass to biofuels with SRL microscopy.**

Targeted lignin modification in bioenergy crops could improve conversion efficiency of lignocellulosic biomass to biofuels. In order to understand the practical consequences of genetic mutations of different crops and variations in the degradation procedures, real-time imaging of the two key chemical components involved in the process, lignin and cellulose, is needed. We demonstrate the use of simultaneous two color stimulated Raman loss (SRL) imaging to address this need. The SRL technique is capable of real-time chemical imaging of multiple components in plant cell wall samples without the use of labels. SRL imaging is background free and linearly dependent on the chemical concentration, allowing for straightforward and rapid quantification of both the lignin and cellulose in the plant samples with sub-micron spatial resolution and high time resolution. We show applications to a variety of plant samples that are candidates for biofuels, including engineered mutants of alfalfa and corn stover with varying lignin content. The superior sensitivity and image contrast in SRL microscopy make it an ideal tool for studying lignocellulosic degradation.

## Probing the Architecture of the Plant Cell Wall during Deconstruction in Single Cells from *Zinnia elegans*

C.I. Lacayo<sup>1\*</sup> (lacayo2@llnl.gov), H.-Y. Holman,<sup>2</sup> M. Hwang,<sup>1</sup> A. Hiddessen,<sup>1</sup> A.J. Malkin,<sup>1</sup> and **M.P. Thelen<sup>1</sup>** (mthelen@llnl.gov)

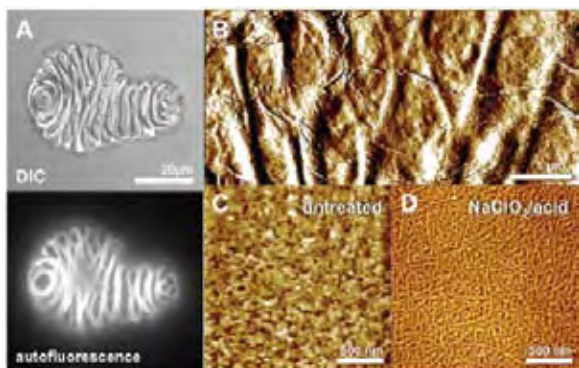
<sup>1</sup>Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, Calif.; and <sup>2</sup>Ecology Dept., Lawrence Berkeley National Laboratory, Berkeley, Calif.

**Project Goals: To use a combination of imaging approaches, including atomic force microscopy (AFM), synchrotron radiation Fourier-transform infrared spectroscopy (SR-FTIR), and fluorescence microscopy, to investigate the structural and dynamic features of plant cell wall biogenesis and degradation.**

The plant cell wall is primarily composed of lignocellulose, a non-food source of fermentable sugars, which can become

the starting material for the production of a variety of biofuels. The cell wall is also a highly complex and dynamic structure, which is not amenable to efficient commercial degradation using currently available technologies. Our goal in this project is to understand how the cell wall changes in response to chemical, enzymatic and microbial treatments, using a combination of imaging platforms, including atomic force microscopy (AFM), synchrotron radiation Fourier-transform infrared spectroscopy (SR-FTIR), and fluorescence microscopy. These techniques provide the capability to image changes in the organization of plant cell surfaces and reveal corresponding chemical composition, thus enabling a more comprehensive understanding of cell wall architecture.

To relate our imaging studies to bioenergy production from plant biomass with high lignocellulose content, we use a *Zinnia elegans* model culture system in which single cells develop lignocellulose-rich walls. *Zinnia* leaf mesophyll cells are induced to gradually differentiate in semi-synchrony into single tracheary elements (TEs), individual components of xylem tissue (panel A in the figure). Mature TEs possess large secondary cell wall thickenings, which are rich in lignocellulose and represent the bulk of woody biomass. These secondary cell wall thickenings become deposited below the primary cell wall and are highly ordered, creating helical or reticulate patterns (panel A). Cultured *Zinnia* TEs represent an advantage over plant tissues by providing a more homogeneous source of lignocellulosic material and by enabling the examination of both primary and secondary cell wall deconstruction in individual cells. In addition, the *Zinnia* cell culture system permits us to probe the cell wall at various stages of the TE differentiation process.



**Figure.** (A) Differential interference contrast (DIC) and autofluorescence images of a differentiated tracheary element from *Zinnia elegans* show a reticulate pattern of the secondary wall thickenings. Secondary cell wall autofluorescence is attributed to lignin. (B) A low-resolution AFM (amplitude) image of a tracheary element also reveals the presence of wall thickenings. (C) A high-resolution AFM (contrast-enhanced height) image shows granular structures on the surface of a cultured untreated tracheary element. (D) Cellulose fibrils become apparent after a chemical treatment.

We have observed that surface organization and composition of *Zinnia* TEs change dramatically after various chemical treatments. As illustrated in panel C in the figure, the cell walls of untreated TEs were found by AFM to have

pronounced granular structures, ranging in size from about 20 to 50 nm. Chemical treatments resulted in the removal of this outermost granular material, revealing the underlying network of cellulose fibrils, which ranged in diameter from approximately 10 to 20 nm (panel D). Distinctive changes in TE composition following chemical treatments were also examined by SR-FTIR.

We are currently developing experimental approaches for high-resolution topographical and structural characterization of the cell wall in TEs and their dynamic response to chemical, enzymatic and microbial treatments. We expect that these approaches will enable us to develop molecular-scale models of the architecture and mechanisms of degradation of the plant cell wall.

This work was funded by the U.S. Department of Energy GTL Program project number SCW0738 and performed under the auspices of the DOE by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

GTL

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

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

University of Florida, Gainesville, Fla.

**Project Goals:** The objectives of this proposed work are to 1) use advanced high resolution magnetic resonance microscopy (MRM), micro and nano x-ray computed tomography (x-ray CT), electron microscopy and imaging mass spectrometry (IMS) to quantify changes in the architecture, porosity, permeability, surface area, pore size, interconnectivity, chemical organization and composition of corn stover and particularly *Populus* and pine wood chips during pretreatment and enzymatic degradation, and; 2) integrate the results from these different quantitative imaging methods into a model for disassembly of the plant cell wall during pretreatment and bioconversion.

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



from these different quantitative imaging methods into a model for disassembly of the plant cell wall during pretreatment and bioconversion. We have standardized the methods for imaging biomass with MRM, x-ray CT and IMS.

IMS: Standardized method for sample preparation and mounting using cryosectioning has been developed for biomass. A number of matrices have been tested and work well with sections of *Populus* wood. Pretreatment of wood with dilute sulfuric acid leads to very complex MS spectra, creating challenges for peak identification. Pure cellulose, hemicelluloses and lignin are being tested to identify characteristic oligomer peaks to focus searches.

MRM: Excellent image quality is obtained from *Populus* wood and bagasse samples using T2 and diffusion weighted modes. In T2 images, vascular bundles appear dark consistent with the knowledge that lignified cells contain limited free water. A chemical shift has been found in some samples and the cause for this is under investigation.

X-ray CT: Excellent images have been obtained at high resolution from *Populus*, pine, and bagasse samples. In addition to the basic density, images are readily segmented and the material and airspace sizes can be quantified.

Current work aims to validate quantitative analysis of each imaging method with the same biomass samples.

GTL

## Single-Molecule Studies of Cellulose Degradation by Cellulosomes

**Haw Yang\*** (hyang@lbl.gov), Jamie Cate (jcate@lbl.gov), Chang-Li Chen, Padma Gunda, Dan Montiel, and Hu Cang

Lawrence Berkeley National Laboratory and University of California, Berkeley, Calif.

**Project Goals: Ethanol production from biomass is one of the few promising solutions to our current energy problem. As a renewable energy source, this approach is hindered by the rate-limiting step of lignocellulose degradation. In nature, this step is accomplished enzymatically by the extracellular cellulosomes from anaerobic microbes. The mechanism, however, has been nearly impossible to study at the biophysical level due to the multiple-scale nature of the problem. The single-molecule approach affords the direct observation of molecular processes within a heterogeneous distribution to shed light on this important issue. Yet, technologies to image individual molecular machines in temporally and spatially heterogeneous setting are not well established.**

**In order to probe function and dynamics of molecular machines in situ, we propose to develop a single-molecule spectrometer with the capability to track individual fluorescent particles as they move in three dimensions**

**(3D). This spectrometer will provide unprecedented time and spatial resolution for imaging single molecules in a complex environment. We will develop and optimize biomolecule tagging, single-molecule assays, and 3D single-particle tracking technologies. These technologies will help establish single-molecule spectroscopy with 3D positioning as a general approach to study molecular machines in complex environments.**

A cellulosome is a large extracellular supramolecular complex that is produced by anaerobic microbes to enzymatically decompose crystalline cellulosic polymers and plant cell walls. It consists of a scaffolding protein that accommodates other essential protein and enzyme components for cellulose degradation. They include carbohydrate-binding modules for attachment to the solid cellulose substrate, various glycoside hydrolases to efficiently hydrolyze a heterogeneous substrate, and in some cases, anchoring proteins to attach cellulosomes to the bacterial cell surface. The mechanism of its function is poorly understood due to the complexity of the cellulosome itself and the natural environment in which it functions. Outstanding issues include the location and the manner in which the cellulosome is assembled, the distribution in the cellulosomal composition, and the dynamic interactions between cellulases and the insoluble substrate, to name a few.

These questions are very difficult to address quantitatively using conventional, ensemble-based methods due to the multiple layers of complexity involved. The convoluted spatio-temporal dynamics in cell and cellulosome interaction with insoluble substrates make it very hard to quantitatively study the various molecular dynamics of a functioning cellulosome. We anticipate that the single-molecule approach, due to its capability of directly monitoring the individual processes from a distribution, will prove invaluable in efforts to unravel how microscopic, molecular interactions impact macroscopic behavior in plant cell wall degradation.

In order to study the various processes involved in lignocellulosic degradation by cellulosomes, we are developing a single-molecule spectrometer with the capability to track individual fluorescent particles in three dimensions. We have developed three-dimensional single-particle tracking spectroscopy instrument that allows us to follow the 3D movements of individual luminescent nanoprobe and at the same time perform spectroscopic investigation. The capabilities will be briefly outlined. We have also developed protocols for imaging individual cellulosomes acting on well prepared crystalline cellulose fibers. Preliminary data will be presented. The mechanistic insights are expected to have a direct impact on the improvement and engineering of tailored biomass depolymerization systems.

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

GTL

### Metabolic Modeling for Maximizing Photobiological H<sub>2</sub> Production in Cyanobacteria

**Alexander S. Beliaev**<sup>1\*</sup> (alex.beliaev@pnl.gov), Grigoriy E. Pinchuk,<sup>1</sup> Oleg V. Geydebrekht,<sup>1</sup> Michael H. Huesemann,<sup>2</sup> Elizabeth H. Burrows,<sup>3</sup> Trang Vu,<sup>4</sup> Jennifer L. Reed,<sup>4</sup> Andrei L. Osterman,<sup>5</sup> John R. Benemann,<sup>6</sup> and Jim K. Fredrickson<sup>1</sup>

<sup>1</sup>Pacific Northwest National Laboratory, Richland, Wash.; <sup>2</sup>Pacific Northwest National Laboratory Marine Science Laboratory, Sequim, Wash.; <sup>3</sup>Oregon State University, Corvallis, Ore.; <sup>4</sup>University of Wisconsin, Madison, Wis.; <sup>5</sup>Burnham Institute for Medical Research, La Jolla, Calif.; and <sup>6</sup>Benemann Associates, Walnut Creek, Calif.

**Project Goals:** In this project, we seek to improve our understanding of H<sub>2</sub> production by a diazotrophic unicellular cyanobacterium *Cyanobacterium* *sp.* strain ATCCC 51142 using a metabolic modeling approach for simulating the fundamental metabolism of indirect biophotolysis, as well as identifying the main metabolic and regulatory controls in this organism. From this, the potential for H<sub>2</sub> production by indirect biophotolysis in this organism will be assessed based on imposing new constraints to redirect the low redox potential electron transport pathways from normal metabolism towards H<sub>2</sub> production from accumulated carbohydrates. As a result, it will provide an *in silico* tool for manipulating such microorganisms to act as catalysts for solar energy conversion to H<sub>2</sub> and potentially allow for a development of a highly efficient H<sub>2</sub> production process.

Advances in microbial genome sequencing and functional genomics are greatly improving the ability to construct accurate systems-level models of microbial metabolism and to use such models for metabolic engineering. With the increasing concerns over the reliance on fossil fuels, there is a revitalized interest in using biological systems for producing renewable fuels. Genomics and metabolic engineering hold great promise for the rational design and manipulation of biological systems to make such systems efficient and economically attractive.

In this project, we are developing a constraint-based metabolic model for marine diazotrophic cyanobacterium *Cyanobacterium* ATCC 51142. The initial set of reactions, used for the first draft of our metabolic model, was deduced from genomic annotations generated using the RAST (Rapid Annotation using Subsystems Technology) server. A draft

reconstruction was generated in SimPheny<sup>®</sup> by Genomatica (San Diego, CA) based on sequence comparisons between *Cyanobacterium* and organisms with genome-scale metabolic models. This reconstruction was a starting point, upon which organism specific pathways were added, and gene-protein-reaction associations were closely evaluated. The current model includes 569 metabolic and transport genes, 542 proteins, and 715 reactions. The following pathways are accounted for in the current model: central metabolism, fatty acids biosynthesis, amino acid synthesis, nucleotide synthesis, cofactor biosynthesis, oxidative phosphorylation, and photosynthesis. The number of metabolic gaps (metabolites which can only be produced or only consumed) has been reduced from 484 in the original draft reconstruction to 302 and is expected to decrease further as other pathways (such as nitrogen fixation and carbon fixation) are being incorporated into the current model. In addition, the biomass formation reaction is being constructed based on experimental measurements and the reaction will help significantly reduce the number of gaps in the model.

Conditions promoting H<sub>2</sub> production by *Cyanobacterium* ATCC 51142 are also being studied in order to develop a strategy for maximizing H<sub>2</sub> production using metabolic modeling approach. Preliminary results showed that this cyanobacterial strain is not capable of generating significant amounts of H<sub>2</sub> from stored glycogen under dark fermentative conditions. However, illumination of *Cyanobacterium* ATCC 51142 cultures exposed to Ar atmosphere and deprived of nitrogen source significantly increased the rate of light-driven H<sub>2</sub> production. No O<sub>2</sub> was produced in course of more than 60 hours hydrogen photoproduction. Interestingly, the addition of DCMU, a photosystem II inhibitor, did not affect this process, whereas presence of N<sub>2</sub> in atmosphere or ammonia in the medium effectively prevented H<sub>2</sub> appearance. Further biochemical analysis revealed that nitrogenase was the enzyme primarily responsible for light-driven H<sub>2</sub> production by *Cyanobacterium* ATCC 51142 under the experimental conditions. Current hypothesis of metabolic pathways involved in nitrogenase mediated H<sub>2</sub> production includes generation of ATP by cyclic electron transport in photosystem I, whereas stored carbohydrates are used as source of electrons. The predicted pathways and experimental data will be incorporated into the metabolic model of *Cyanobacterium* ATCC 51142 to identify the means for maximization of H<sub>2</sub> production by this organism.

## A High-Throughput Genetic Screen for Large Subunit Hydrogenase (*boxH*) Enzyme Presence in Cyanobacteria Using Whole-Cell Duplex PCR and Novel Amplification Parameters

Nicholas J. Skizim (nskizim@princeton.edu), Derrick R.J. Kolling,\* Diana Chien, and G. Charles Dismukes

Dept. of Chemistry, Princeton University, Princeton, N.J.

**Project Goals: The goal of this project is the development of a high-throughput screening method for the enzymatic presence of hydrogenase in cyanobacteria.**

In designing a screen for hydrogen production capabilities in cyanobacteria, it is advantageous to use a method which has high-throughput as well as high confidence in identifying positive hits. Here we describe degenerate primers amplifying a portion of the large subunit of the hydrogenase complex (*boxH*) designed from protein motifs which bind the [Ni-Fe] active site in the complex and are conserved throughout cyanobacteria. These oligonucleotides contain flanking restriction sites to increase target specificity as well as to facilitate subsequent cloning/sequencing. The screen relies on Duplex PCR to concurrently amplify fragments of both *psbA* (the photosystem II D1 protein) along with *boxH* to provide a positive control on the reaction conditions, enzyme activity, and whole-cell template. Touchdown PCR is in use to achieve the highest level of specificity from the degenerate primers and the PCR cocktail is tailored as well to (a) increase specificity, and (b) be compatible with whole-cell template methods. A preliminary screen of five cyanobacterial strains (*Cyanosphaera miamiensis* BG043511, *Synechocystis* PCC6803, *Synechococcus* PCC7002, *Arthrospira maxima*, and *Gloeobacter violaceus*) has verified the presence and absence of hydrogenase genes as expected, illustrating proof of method. Data is also presented showing the effect of various cellular pre-treatments on the amplification. Further, the possibilities of a fluorescent based detection instead of agarose gel electrophoresis and ethidium bromide stain are discussed.

This work has been supported by the Department of Energy – Genomics:GTL, DE-FG02-07ER64488, as well as the Air Force Office of Scientific Research, AFOSR-MURI- FA9550-05-1-0365.

## Probing the Interdependence of Hydrogenase- and Nitrogenase-Dependent Hydrogen Production in *Cyanosphaera* sp. Miami BG043511 by Genetic Deletion and Mutant Analysis

Nicholas J. Skizim<sup>1\*</sup> (nskizim@princeton.edu), Randall Kerstetter,<sup>2</sup> Todd Michael,<sup>2</sup> and G. Charles Dismukes<sup>1</sup>

<sup>1</sup>Dept. of Chemistry, Princeton University, Princeton, N.J.; and <sup>2</sup>Waksman Institute of Microbiology, Rutgers University, Piscataway, N.J.

**Project Goals: The goal of this project is for the development of quantitative tools to measure hydrogen production in photosynthetic bacteria.**

The unicellular diazotrophic cyanobacterium *Cyanosphaera* sp. Miami BG043511 is unique in being the sole reported cyanobacterium possessing nitrogenase and bidirectional hydrogenase but lacking an “uptake” hydrogenase. In addition to the bidirectional enzyme, the nitrogenase can produce hydrogen. Because of the complex anaerobic energy balance in this organism, including the ATP demand of nitrogen fixation, it is difficult to chemically inhibit one of the two hydrogen-producing enzymes without indirectly affecting the other. Therefore, we aim to create mutants deficient in each of the above named enzymes, including a double mutant in which both are deleted. Initial work relying on degenerate primers and inverse PCR has given us the complete sequence of the nitrogenase iron protein (*nifH*) as well as a partial sequence homologous to *hupL*, the large subunit of the uptake hydrogenase. We have collected sequence data from this organism using Applied Biosystem’s SOLiD sequencing techniques and are in the process of assembling the entire genome. A rough assembly of the genome along with Sanger sequencing to fill in the gaps has confirmed this presence of a complete *hupL*. Because this strain is claimed to lack activity of the uptake hydrogenase and this observation is in contradiction to that claim, we are in the process of assessing the claim of zero uptake hydrogenase function. A mutant ( $\Delta hupL$ ) is being constructed to determine whether the genetic presence of this enzyme influences the phenotype in any way. Hydrogen production from this strain has been measured on an ultra-sensitive Clark-type electrode revealing multiple kinetic phases of hydrogen production under fermentative conditions. This hydrogen production, as well as the excretion of other fermentative end-products, has been investigated both when the culture is concurrently fixing nitrogen as well when it is not. The data from this shall be presented.

This work has been supported by the Department of Energy – Genomics:GTL, DE-FG02-07ER64488, as well as the Air Force Office of Scientific Research, AFOSR-MURI- FA9550-05-1-0365.

## Microalgal Biomass as Feedstock for Production of Hydrogen and Methane Fuels

Mark B. Smith<sup>1\*</sup> (mbsmith@princeton.edu), Damian Carrieri<sup>1</sup> (carrieri@princeton.edu), Jun Cheng<sup>2</sup> (juncheng@zju.edu.cn), Gennady Ananyev<sup>1</sup> (anayev@princeton.edu), and **G. Charles Dismukes**<sup>1</sup> (dismukes@princeton.edu)

<sup>1</sup>Princeton University, Princeton, N.J.; and <sup>2</sup>Zhejiang University, Zhejiang Province, China

<http://www.princeton.edu/~biosolar/>

As an alternative to land-based crops for biofuels, Aquatic Microbial Oxygenic Phototrophs (AMOPs), such as algae, cyanobacteria and diatoms have several potential advantages. AMOPs are inherently more efficient solar collectors, use less or no land, and can be converted to usable fuels using simpler technologies than cellulose (1). One developed strategy for conversion of biofeedstocks to methane and hydrogen is anaerobic digestion using microbes from sewage sludge. This has been previously demonstrated for glucose and starch containing feedstocks from land-based crops (2-3). Here we illustrate the use of AMOPs as an alternative feedstock for such processes. AMOPs are buoyant cells that are completely devoid of the recalcitrant biopolymers that comprise terrestrial crops (cellulose, lignin, hemi-cellulose) and enriched in energy precursors. *Arthrospira maxima* is a filamentous cyanobacterium that grows to high biomass densities in open aquifers and is permissive of high salinity, high pH and high carbonate concentrations, conditions that suppress microbial contamination. *Arthrospira* species have been grown in large outdoor facilities for commercial purposes and demonstrated to grow on fertilized sea water (4). The high conversion efficiency of cyanobacterial biomass to methane and hydrogen using our process will be discussed.

### References

1. G.C. Dismukes, D. Carrieri, N Bennette, G. M. Ananyev, and M.C. Posewitz. (2008) "Aquatic phototrophs: efficient alternatives to land-based crops for biofuels" *Current Opinion in Biotechnology* 19: 235-240
2. Binfei Xie, Jun Cheng, Junhu Zhou, et al. Cogeneration of hydrogen and methane from glucose to improve the energy conversion efficiency. *International Journal of Hydrogen Energy*, 2008, 33(19): 5006-5011.
3. Binfei Xie, Jun Cheng, Junhu Zhou, et al. Production of hydrogen and methane from potatoes by two-phase anaerobic fermentation. *Bioresource Technology*, 2008, 99(13): 5942-5946.
4. M. R. Tredici, T. Papuzzo, and L. Tomaselli (1986) "Outdoor mass culture of *Spirulina maxima* in sea-water" *Applied and Microbiology and Biotechnology* 24: 47-50

Supported by AFOSR-MURI- FA9550-05-1-0365 and Princeton University.

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

Alexandra Dubini,<sup>1</sup> Matthew Wecker,<sup>2</sup> and **Maria L. Ghirardi**<sup>1\*</sup> (maria\_ghirardi@nrel.gov)

<sup>1</sup>National Renewable Energy Laboratory, Golden, Colo.;  
<sup>2</sup>GeneBiologics, LLC, Boulder, Colorado

Photobiological H<sub>2</sub> production from water is a clean, non-polluting and renewable technology. Although the potential light conversion efficiency to H<sub>2</sub> by biological organisms is theoretically high (about 10%), the system is currently limited by biochemical and engineering constraints. These limitations include (but are not restricted to) the extreme O<sub>2</sub> sensitivity of the biological hydrogenases and the low availability of reductants to the hydrogenase due to the existence of competing metabolic pathways.

Our research addresses the O<sub>2</sub> sensitivity issue by developing a new, biologically-based assay to screen large microbial populations for improved H<sub>2</sub>-production properties. This novel assay is based on the H<sub>2</sub>-sensing properties of systems found in nitrogenase-containing photosynthetic bacteria. We will validate the new assay by using it to screen mutants generated through directed-evolution techniques for O<sub>2</sub> tolerant [FeFe]-hydrogenases. The hydrogen-sensing system in *R. capsulatus* is being optimized as an assay of heterologous hydrogen production by the bacteria and will then be used to screen mutants generated through directed-evolution techniques for O<sub>2</sub> tolerant [FeFe]-hydrogenases. The hydrogenases of *Clostridium acetobutylicum* and *Bacteroides thetaiotaomicron*, along with their respective assembly proteins have been introduced into broad host range vectors and are being shuttled into the photosynthetic bacteria *Rhodospirillum rubrum*.

To address the issue of competitive metabolic pathways with H<sub>2</sub> production, we will adapt the yeast two-hybrid assay to measure the interactions between different ferredoxin isoforms present in *Chlamydomonas reinhardtii* with different proteins known to accept electrons from ferredoxin in most photosynthetic and fermentative organisms. We started examining the direct interaction of the HYDA2 hydrogenase and other partners using a yeast two-hybrid system. The cDNA library has been prepared and a first screen has been done. Some prey clones have been picked and are being tested for interaction strength by growth tests and β-galactosidase assay.

This project will develop techniques that will drive a deeper understanding of algal H<sub>2</sub> metabolism and accelerate the development of future photobiological H<sub>2</sub>-production catalysts and organisms.

## Novel Hydrogen Production Systems Operative at Thermodynamic Extremes

Michael J. McInerney,<sup>1</sup> Jesscia Seiber,<sup>1</sup> Ralph S. Tanner,<sup>1</sup> and **Robert P. Gunsalus**<sup>2\*</sup> (robgs@microbio.ucla.edu)

<sup>1</sup>University of Oklahoma, Norman, Okla.; and <sup>2</sup>University of California, Los Angeles, Calif.

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

**Project Goals: The goals of our collaborative project are to develop new research strategies, model organisms, and research resources addressing the needs of the Genomics:GTL program in the area of biohydrogen production. We will apply a suite of molecular, bioinformatics, and biochemical tools to: i) interrogate the thermodynamically limiting steps of H<sub>2</sub> production from fatty and aromatic acids in syntrophic communities and in pure cultures, ii) develop a new microbial model system that generates high H<sub>2</sub> concentrations (over 17% of the gas phase) with high hydrogen yields (3+ mol H<sub>2</sub> per mole glucose), and iii), perform systems-based studies of biohydrogen production in model anaerobe consortia to identify key metabolic steps. The results of these studies will greatly expand our ability to predict and model systems for H<sub>2</sub> production in novel anaerobes that are currently very poorly understood.**

The goals of this collaborative project are to develop new research strategies to address the Genomics:GTL program needs in the area of bio-hydrogen production. This includes the delineation of the molecular machinery involved in hydrogen production from thermodynamically difficult substrates, as well as the characterization of new microbial model systems that generate high H<sub>2</sub> concentrations approaching 17% of the gas phase.

To identify the molecular machinery for hydrogen production in model syntrophic microorganisms we are performing genomic analysis with *Syntrophus wolfei*. This bacterium is representative of an important but poorly understood class of hydrogen producing organisms that are capable of syntrophic fatty and aromatic acid metabolism when co-cultured with suitable microbial partner(s). Their ability to produce H<sub>2</sub> requires reverse electron transport where energy input is needed.

Analysis of the *S. wolfei* genome predicts that it encodes multiple hydrogenase enzymes, two of which are predicted to be cytoplasmic hydrogenases, and one is predicted to an externally located hydrogenase. One of the soluble Fe-only hydrogenase genes is clustered with genes for NADH dehydrogenase. This arrangement is also found in two other sequenced organisms capable of syntrophy, *Pelotomaculum thermopropionicum* and *Syntrophobacter fumaroxidans*. Thus, one mechanism for hydrogen production during syntrophy maybe to form a complex that directly couples NADH oxidation with hydrogen production. The externally-oriented

is predicted to be membrane-associated, ferredoxin hydrogenase (E.C. 1.12.7.2). One of ORFs in this cluster was most similar to a gene for cytochrome b<sub>556</sub>, suggesting that this hydrogenase could produce hydrogen from a reduced quinone intermediate by reverse electron transfer. Cells of *Syntrophomonas wolfei*, each grown in coculture with *Methanospirillum hungatei*, were separated from the methanogenic partner by Percoll density centrifugation and the amounts of externally-oriented hydrogenase and total hydrogenase activities were measured. Most of the hydrogenase activity was externally-oriented. These data implicate the involvement of membrane-associated hydrogenases when thermodynamically difficult substrates such as fatty and aromatic acids are the substrates.

In another project we are characterizing a newly isolated bacterium called *Anaerobaculum* strain OS1 that can generate H<sub>2</sub> at concentrations up to 17%. This ability suggests the presence of a novel system for H<sub>2</sub> production from carbohydrates not involving NADH where we hypothesize that glucose metabolism involves ferredoxin-linked rather than NADH-linked dehydrogenases and membrane-bound hydrogenase systems. Initial microbiological characterization of strain OS1 reveals phenotypically distinguishable traits from *A. thermoterrenum* and *A. mobile*, and it is designated as a new species: "*Anaerobaculum hydrogeniformans*". In pure culture, strain OS1 produced 3.85 moles of H<sub>2</sub> per mole of glucose and is a much better H<sub>2</sub> producer than either of the two described *Anaerobaculum* species. OS1 also grew syntrophically in the presence of a H<sub>2</sub>-consuming methanogen, suggesting the ability for a more complex alternative lifestyle.

## Mechanism of Post-Translational Control of Nitrogenase is Revealed by Analysis of Constitutive Hydrogen-Producing Mutants of *Rhodospseudomonas palustris*

Erin K. Heiniger\* (erinkh@u.washington.edu), Yasuhiro Oda, and **Caroline S. Harwood**

Dept. of Microbiology, University of Washington, Seattle, Wash.

**Project Goals: The overall goal of this project is to develop and apply techniques in metabolic engineering to improve the biocatalytic potential of the bacterium *Rhodospseudomonas palustris* for nitrogenase-catalyzed hydrogen production. *R. palustris*, is an ideal platform to develop as a biocatalyst for hydrogen gas production because it is an extremely versatile microbe that produces copious amounts of hydrogen by drawing on abundant natural resources of sunlight and biomass. Anoxygenic photosynthetic bacteria, such as *R. palustris*, generate hydrogen and ammonia during nitrogen fixation. The applied use of nitrogenase for hydrogen production is attractive because hydrogen is an obligatory product of**

this enzyme and is formed as the only product when nitrogen gas is not supplied. Our challenge is to understand the systems biology of *R. palustris* sufficiently well to be able to engineer cells to produce hydrogen continuously, as fast as possible and with as high a conversion efficiency as possible of light and electron donating substrates. This abstract increases our understanding of how nitrogenase and hydrogen production are regulated in *R. palustris*.

*Rhodospseudomonas palustris* is a phototrophic proteobacterium that produces hydrogen gas via nitrogenase, the enzyme responsible for biological nitrogen fixation. Nitrogenase expression and activity are tightly controlled by the bacterium because nitrogen fixation is a costly reaction. Previously we described mutants that produce hydrogen constitutively, even when grown in the presence of ammonium, due to activating NifA\* mutations in the regulator NifA (1). Ammonium is a biologically available nitrogen source that normally prevents NifA from activating nitrogenase gene transcription. In many bacteria nitrogenase activity is also tightly controlled by DRAT and DRAG enzymes. DRAT inactivates nitrogenase by ADP-ribosylation in response to ammonia, while DRAG removes the modification. NifA does not control the expression of these enzymes in *R. palustris*. We constructed a DRAT mutant and confirmed that DRAT inactivates nitrogenase in *R. palustris*. We also confirmed that DRAT does not switch off nitrogenase activity or hydrogen production in our NifA\* mutants in response to ammonia addition. To explain this surprising finding we hypothesized that the hydrogen-producing mutants do not express a small regulatory PII protein that is required to activate DRAT activity in response to ammonia. We constructed mutations in the *glnB*, *glnK1* and *glnK2* genes that encode the PII proteins of *R. palustris*, and determined that GlnK2 is required for the activation of DRAT upon ammonium addition to cultures. In complementary microarray studies, we found that genes controlled by the *nrBC* regulatory system are not expressed in the hydrogen-producing mutants grown in the presence of ammonium. Among these is the *glnK2* gene. Thus, in addition to constitutively activating nitrogenase gene expression, NifA\* mutations also indirectly allow *R. palustris* cells to escape posttranslational inactivation of nitrogenase and produce hydrogen even when grown in the presence of ammonium.

## Reference

1. Rey, F. E., E. K. Heiniger and C. S. Harwood. 2007. Redirection of metabolism for biological hydrogen production. *Appl. Environ. Microbiol.* **73**:1665-1671

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

Erik L. Hendrickson,<sup>1</sup> Qiangwei Xia,<sup>1</sup> William B. Whitman,<sup>2</sup> Murray Hackett,<sup>1</sup> John A. Leigh<sup>3\*</sup> (leighj@u.washington.edu), and Nitin Baliga<sup>4</sup>

<sup>1</sup>Dept. of Chemical Engineering, University of Washington, Seattle, Wash.; <sup>2</sup>Dept. of Microbiology, University of Georgia, Athens, Ga.; <sup>3</sup>Dept. of Microbiology, University of Washington, Seattle, Wash.; and <sup>4</sup>Institute for Systems Biology, Seattle, Wash.

**Project Goals: Use transcriptomics, proteomics, and metabolomics to study the systems biology of H<sub>2</sub> metabolism, formate metabolism, nitrogen fixation, and carbon assimilation in *Methanococcus maripaludis*. Determine the mechanism of H<sub>2</sub> sensing and transcriptional regulation by H<sub>2</sub>.**

## Background

We are engaged in a long-term effort to understand regulatory networks in hydrogenotrophic methanogens, members of the Archaea whose energy metabolism specializes in the use of H<sub>2</sub> to reduce CO<sub>2</sub> to methane. (Many hydrogenotrophic methanogens can use formate as an alternative to H<sub>2</sub> and CO<sub>2</sub>). Our studies focus on *Methanococcus maripaludis*, a model species with good laboratory growth characteristics, facile genetic tools, and a tractable genome of 1722 annotated ORFs. Much of our work to date has focused on the response that occurs when supplies of essential nutrients are decreased to growth-limiting levels. Thus, we have studied the responses to H<sub>2</sub> limitation, nitrogen limitation, phosphate limitation, and leucine limitation (using a leucine auxotroph) [1, 2].

## Continuous culture of *M. maripaludis*

A key aspect of our approach is the use of continuous culture for maintaining defined nutrient conditions [3]. Chemostats are operated at a constant dilution rate, hence growth rate is constant. Cell density is held constant as well, determined by the supply of the growth-limiting nutrient.

## Effect of H<sub>2</sub>-limitation on the transcriptome

The effect of limitation by a given nutrient is determined by comparison with at least two other nutrient limitations. Thus, samples from H<sub>2</sub>-limited cultures were compared with samples from phosphate-limited cultures and samples from leucine-limited cultures. Array analysis revealed that the mRNA levels most highly affected by H<sub>2</sub> were for those encoding certain enzymes in the methanogenic pathway. Notably, these enzymes use coenzyme F<sub>420</sub> as electron donor or receptor, highlighting the importance of electron flow through F<sub>420</sub>.

## Effect of H<sub>2</sub>-limitation on the proteome

Proteomics was conducted by 2-D capillary HPLC coupled with tandem mass spectrometry. Samples from H<sub>2</sub>-limited cultures were compared with samples from phosphate-

limited cultures and samples from nitrogen-limited cultures. Eight percent of the proteome changed significantly with H<sub>2</sub> limitation. Many of the proteins involved in methanogenesis increased. One protein involved in methanogenesis decreased: a low-affinity [Fe] hydrogenase, which may dominate over a higher-affinity mechanism when H<sub>2</sub> is abundant.

#### Utility of high-coverage quantitative proteomics in *M. maripaludis*

With H<sub>2</sub> limitation, a wider variety of proteins involved in methanogenesis increased than was previously indicated for mRNAs. While a variety of factors are possible, it is likely that part of the explanation lies in the superior ability of the proteomics to discern statistically significant differences. Exhaustive sampling of the proteome was key. When sampling the proteome at or near saturation in terms of qualitative identifications from proteolytic digests, the statistical power to detect a defined fold change between two sets of growth conditions, at a defined level of significance, is quite high for many *M. maripaludis* proteins. How many of these statistically significant abundance changes are of biological relevance is a question we are actively investigating.

#### Future directions

We plan an extensive study of transcriptomic and proteomic correlates to changes in H<sub>2</sub> metabolism (utilization and production), electron flow, nitrogen fixation, and carbon assimilation, leading ultimately to modeling of regulatory networks. We will also investigate the mechanism of H<sub>2</sub> sensing and regulation.

#### References

1. Hendrickson EL, Haydock AK, Moore BC, Whitman WB, Leigh JA: Functionally distinct genes regulated by hydrogen limitation and growth rate in methanogenic Archaea. *Proc Natl Acad Sci U S A* 2007, 104(21):8930-8934.
2. Hendrickson EL, Liu Y, Rosas-Sandoval G, Porat I, Söll D, Whitman WB, Leigh JA: Global responses of *Methanococcus maripaludis* to specific nutrient limitations and growth rate. *J Bacteriol* 2008, 190(6):2198-2205.
3. Haydock AK, Porat I, Whitman WB, Leigh JA: Continuous culture of *Methanococcus maripaludis* under defined nutrient conditions. *FEMS Microbiol Lett* 2004, 238(1):85-91.

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

Harry Christman,<sup>1</sup> Elsie Campbell,<sup>1</sup> Brett Phinney,<sup>2</sup> Wan-Ling Chiu,<sup>3</sup> and **John C. Meeks**<sup>1\*</sup> (jcmeeks@ucdavis.edu)

<sup>1</sup>Dept. of Microbiology and <sup>2</sup>Genome Center, University of California, Davis, Calif.; and <sup>3</sup>Dept. of Biology, Virginia Commonwealth University, Richmond, Va.

**Project Goals: Heterocyst-forming cyanobacteria, such as *Nostoc punctiforme*, are applicable for cost-effective biohydrogen production, providing the frequency of heterocysts in the filaments can be increased by genetic manipulations and the metabolic end product, H<sub>2</sub>, is uncoupled from growth.**

A rationale for this project is that heterocyst-forming cyanobacteria, such as *Nostoc punctiforme*, are applicable for cost-effective biohydrogen production, providing the frequency of heterocysts in the filaments can be increased by genetic manipulations and the metabolic end product, H<sub>2</sub>, is uncoupled from growth. Heterocysts are the sites of nitrogenase and hydrogenase activities in these oxygenic photoautotrophs. Two keys to the genetic manipulation of heterocyst differentiation are detailed knowledge of the regulatory pathway for the initiation and maintenance of the approximately 10% heterocyst frequency in the free-living growth state, and identification of the regulatory targets of plant signals in establishment of a nitrogen-fixing symbiosis with *N. punctiforme*, wherein the heterocyst frequency increases to 30-40% of the cells and ammonium is released to the plant. The *N. punctiforme* symbiotic growth state is additionally characterized by an unbalanced metabolic physiology and a slow growth rate; these characteristics are also essential to duplicate for bioreactor H<sub>2</sub> production. Our working hypothesis is that heterocyst differentiation is modulated by a regulatory system operating in a cascade manner and that plant partners have evolved signals to co-opt the developmental regulatory pathway, as well as to control *Nostoc* growth and metabolism.

The experimental objectives are to apply transcriptomic and proteomic time course assays to wild-type and mutants of free-living cultures defective in stages of heterocyst differentiation in order to define epistatic relations. Expression patterns and potential protein modification of *N. punctiforme* in symbiotic growth with the bryophyte hornwort *Anthoceros punctatus* and the angiosperm *Gunnera manicata* will then be pursued.

Cluster analysis of 0.5 to 24 h time course experiments in free-living cultures have identified three distinct temporal clusters of 344 total up-regulated genes and one 28 member cluster of down-regulated genes. Comparisons of the cluster

members to steady state dinitrogen grown cultures indicate that by 24 h after induction of heterocyst differentiation, the cultures have not reached the steady state levels of gene expression. The results of this latter analysis allow us to identify four unique transcriptional patterns of stress related genes; 328 genes up-regulated following nitrogen starvation that decline to steady state values after 24 h, 147 genes that do not immediately change expression which are up-regulated after 24 h, 75 genes that do not change expression which are down-regulate after 24 h and 389 genes that are down-regulated following nitrogen starvation which are up-regulated after 24 h. The latter group of genes is enriched in those encoding core metabolic functions.

GTL

### *Thermotoga maritima* Sugar Kinome

Irina Rodionva,<sup>1</sup> Xiaoqing Li,<sup>1</sup> Chen Yang,<sup>1</sup> Ying Zhang,<sup>1</sup> Ines Thiele,<sup>2</sup> Bernhard O. Palsson,<sup>2</sup> Dmitry Rodionov<sup>1</sup> (rodionov@burnham.org), and **Andrei Osterman**<sup>1,3\*</sup> (osterman@burnham.org)

<sup>1</sup>Burnham Institute for Medical Research, La Jolla, Calif.; <sup>2</sup>Dept. of Bioengineering, University of California San Diego, La Jolla, Calif.; and <sup>3</sup>Fellowship for Interpretation of Genomes, Burr Ridge, Ill.

**Project Goals:** This project is focused on a systems-level understanding of biological hydrogen production using *Thermotoga maritima* as a model organism. We will address the basic science required to improve our understanding of hydrogen production from various carbon sources including glucose, cellulose, starch and xylan by this thermophilic microorganism. The overall goal is 1) to reconstruct the regulatory and metabolic network in *T. maritima* using various sets of “omics” data, 2) to integrate regulatory and metabolic networks into one “integrated” genome-scale model, 3) to confirm and validate the ability of the integrated model to predict processing of various environmental signals.

The marine hyperthermophilic bacterium *Thermotoga maritima* has extensive and highly diversified carbohydrate utilization machinery. A detailed reconstruction of this machinery including uptake mechanisms, biochemical transformations and transcriptional regulation is of key importance for the scope of our DOE-GTL-sponsored project “Systems-level understanding of hydrogen production by *Thermotoga maritima*” (see the accompanying poster by V. Portnoy et al.). Accurate functional assignment of carbohydrate utilization genes is challenging due to substantial variations of the respective pathways between species including frequent nonorthologous gene displacements and functionally divergent paralogs. Therefore, their homology-based annotations in various genomic databases are often incomplete and imprecise. To address this challenge we combine a subsystems-based approach to pathway analysis (implemented in the SEED genomic platform) with the experimental characterization of *signature* genes.

In this study we applied this integrated approach to infer and experimentally assess substrate specificities within the *T. maritima* sugar kinome (SK) represented by at least 20 sugar kinases involved in a variety of carbohydrate utilization pathways. Sugar phosphorylation is an essential step in the overwhelming majority of such pathways. Whereas in *E. coli*, this step is often performed by uptake-associated phosphotransferases (PTS), in *T. maritima* it appears to be fully delegated to the members of its highly extended and diversified SK. Using genome context analysis (conserved operons and regulons) we were able to tentatively assign nearly all members of SK specific physiological roles in reconstructed pathways. We used a “matrix” approach to experimentally test these assignments and to explore the relationship between the inferred physiological roles and *in vitro* substrate preferences of respective enzymes. Purified recombinant proteins (obtained using expression strains from the Joint Center for Structural Genomics, www.jcsg.org) were tested for their kinase activity versus a panel of >40 different mono- and disaccharides. Remarkably, nearly all of the 15 experimentally characterized enzymes (from 4 structurally distinct superfamilies) displayed a strong preference towards a single physiological substrate. The results of this analysis provide the direct validation of reconstructed pathways in *T. maritima* and improve our ability to accurately annotate SK in other genomic and metagenomic datasets.

Most of the analyzed sugar utilization pathways are controlled by committed transcription factors (TF). *T. maritima* has an unusually large number of TFs of the so-called ROK-family that are composed of two domains, a N-terminal kinase-like effector domain and a C-terminal DNA-binding HTH-domain. We used a comparative genomic approach for the *ab initio* prediction of regulons tentatively controlled by all six members of this family. Five of them were implicated in the regulation of carbohydrate utilization pathways. Gel-shift mobility assays were used to confirm all of the predicted TF binding sites and to assess possible small molecule effectors, most of them, mono- or disaccharides involved in respective pathways. This study illustrates the power of the subsystems-based approach to comparative genomic reconstruction of metabolic and regulatory networks that will be further extended for the assessment of biohydrogen production by *T. maritima*.

GTL

### Conservation and Variations in *Shewanella* Transcriptional Regulatory Network

Dmitry Rodionov<sup>1\*</sup> (rodionov@burnham.org), Irina Rodionova,<sup>1</sup> Xiaoqing Li,<sup>1</sup> Pavel Novichkov,<sup>2</sup> Margaret Romine,<sup>3</sup> Mikhail Gelfand,<sup>4</sup> and **Andrei Osterman**<sup>1,5</sup> (osterman@burnham.org)

<sup>1</sup>Burnham Institute for Medical Research, La Jolla, Calif.; <sup>2</sup>Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; <sup>3</sup>Pacific Northwest



National Laboratory, Richland, Wash.; <sup>4</sup>Institute for Information Transmission Problems RAS, Moscow, Russia; and <sup>5</sup>Fellowship for Interpretation of Genomes, Burr Ridge, Ill.

**Project Goals: The project “Integrated Genome-Based studies of *Shewanella* Ecophysiology” is a component of the *Shewanella* Federation studies aimed at better understanding of the ecophysiology and speciation of this important genus.**

Comparative genomics approach was used to infer transcriptional regulatory networks in *Shewanella oneidensis* MR-1 and fourteen other species of *Shewanella* with sequenced genomes. To accomplish this goal, we combined the identification of transcription factors (TF) and candidate TF-binding sites with the cross-genome comparison of regulons and the genome context analysis of candidate target genes. The identification of TF repertoire (termed regulome) in *Shewanella* points to the existence of between 113 and 199 regulons per genome. Overall, the pan-regulome of 15 *Shewanella* strains includes 66 groups of orthologous TFs present in all strains (the core regulome), 26 unique to single strains, and 242 present in two or more strains.

Using the combined comparative genomic approach we described 78 TF regulons that control the central metabolism, production of energy and biomass, metal ion homeostasis and stress response in the *Shewanella* lineage that split into two groups. TFs and respective DNA motifs in the first group of 37 regulons are conserved between *E. coli* and *Shewanella*. This group includes regulons associated with metabolism of nitrogen, amino acids, fatty acids, carbohydrates, and cofactors that are largely conserved among *Shewanella* spp. It also includes several global regulons (such as Crp), for which the conservation of respective regulatory sites is much weaker. Although some variations in the regulon content are observed between different members of the *Shewanella* group, the most striking differences in the overall regulatory strategy are revealed in comparison with *E. coli*. Among multiple trends in diversification and adaptive evolution of regulatory interactions in two lineages of gamma-proteobacteria are regulon “shrinking”, “expansion”, “mergers”, and “split-ups”. The second group of 41 regulons identified in *Shewanella* utilizes TFs that do not have orthologs in *E. coli*. Among them are novel regulons that are predicted to control degradation of branch chain amino acids (LiuR), fatty acids (PsrA), and sugar catabolism (NagR, ScrR, AraR, and BglR tentatively implicated in the control of utilization of N-acetylglucosamine, sucrose, arabinose and  $\beta$ -glucosides, respectively). Remarkably, a number of physiologically equivalent pathways in *E. coli* and *Shewanella* appear to be under control of nonorthologous TFs. For example, two regulators FruR and Crp that control a large number of catabolic genes in *E. coli* are functionally replaced in this capacity by distinct TFs, HexR and PdhR, in *Shewanella*. At the same time, the content and functional role of the Crp regulon in *Shewanella* is significantly shifted towards controlling the anaerobic respiration.

Experimental validation of novel predicted regulons in *Shewanella* species is underway. It includes combination of several approaches. We use gel-shift mobility assays to

confirm TF-binding sites and effectors for selected novel TFs (e.g. NagR). The analysis of correlations between multiple microarray expression profiles provides additional information for the assessment of regulon predictions. We are planning to use systematic gene expression profiling in a panel of TF knockout mutants for validation and further refinement of the genomic reconstruction of the transcriptional regulatory network in *Shewanella* spp. This project is a component of the *Shewanella* Federation studies aimed at better understanding of the ecophysiology and speciation of this important genus.

GTL

## Development of *Cyanotheca* as a New Model Organism for Biological Hydrogen Production

Anindita Banerjee,<sup>1</sup> Michelle Liberton,<sup>1</sup> Hanayo Sato,<sup>1</sup> Jana Stöckel,<sup>1</sup> Lawrence Page,<sup>1</sup> Rick Huang,<sup>2</sup> Yinjie Tang,<sup>2</sup> Hogtao Min,<sup>3</sup> Louis A. Sherman,<sup>3</sup> Stephen J. Callister,<sup>4</sup> Richard D. Smith,<sup>4</sup> and Himadri B. Pakrasi<sup>1,2\*</sup>

<sup>1</sup>Depts. of Biology and <sup>2</sup>Energy, Environmental and Chemical Engineering, Washington University, St. Louis, Mo.; <sup>3</sup>Dept. of Biological Sciences, Purdue University, West Lafayette, Ind.; and <sup>4</sup>Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash.

**Project Goals: The objective of this proposal is to develop the cyanobacterium *Cyanotheca* as a model organism for photobiological hydrogen production. Members of the genus *Cyanotheca* are unicellular oxygenic prokaryotes with the ability to fix atmospheric nitrogen. Our long-term goal is to develop a deep understanding of the metabolism of these microbes as it pertains to H<sub>2</sub> evolution. Specifically, we will use genome sequencing, microarrays, proteomics, mutagenesis, biochemical analysis and physiology, all of which are encased in a systems biology framework.**

This project is focused on developing *Cyanotheca*, a cyanobacterium, as a new model microbe to study biological hydrogen production. Cyanobacteria are oxygen evolving photosynthetic prokaryotes that live under wide ranges of environments and are key organisms in the harvesting of solar energy at a global level. Among them, unicellular species such as *Cyanotheca* have the additional ability of nitrogen fixation, a process that is exquisitely sensitive to oxygen. To accommodate such incompatible processes in a single cell, *Cyanotheca* produces oxygen and stores carbon in the form of glycogen during the day, and subsequently creates an O<sub>2</sub>-limited intracellular environment to perform oxygen-sensitive processes such as N<sub>2</sub>-fixation and H<sub>2</sub> production during the night. Thus, *Cyanotheca* cells are natural bioreactors for the storage of captured solar energy with subsequent utilization at a different time during a diurnal cycle.

Our long-term goal is to gain a broad understanding of the capacity of hydrogen production by *Cyanotheca* cells, using a systems biology approach. We are studying seven *Cyanotheca*

strains with sequenced genomes and choose one with the best metabolic properties as the model organism. Our objective is to develop a two-stage photobiological hydrogen production process. The first stage is the production of biomass (glycogen) during photosynthesis (day time). In the second stage, H<sub>2</sub> will be produced from glycogen during O<sub>2</sub>-limited dark fermentation (night time). We plan to use highly controlled and monitored cultivation methods combined with advanced analytical techniques to optimize both processes. Furthermore we will use genome sequence-based insights and advanced omics-level tools to optimize H<sub>2</sub>-production during O<sub>2</sub>-limited dark fermentation. These studies are expected to provide potential solutions to many of the existing challenges in hydrogen production by photosynthetic microbes, including the strategy for periodic production of an oxygen-limited environment in an oxygen-evolving cell under aerobic conditions

GTL

## Systems-Level Understanding of Hydrogen Production by *Thermotoga maritima*

Vasiliy Portnoy<sup>1\*</sup> (vasiliy.portnoy@gmail.com), Karsten Zengler,<sup>1</sup> Ines Thiele,<sup>1</sup> Ying Zhang,<sup>2</sup> Andrei Osterman,<sup>2</sup> and **Bernhard O. Palsson**<sup>1</sup> (palsson@ucsd.edu)

<sup>1</sup>Dept. of Bioengineering, University of California San Diego and <sup>2</sup>Burnham Institute for Medical Research, La Jolla, Calif.

**Project Goals: The project addresses the basic science required to improve our understanding of hydrogen production from various carbon sources including glucose, cellulose, starch and xylan by *Thermotoga maritima*. The goals of the project are 1) to reconstruct the regulatory and metabolic network in *T. maritima* using various sets of “omics” data, 2) to integrate regulatory and metabolic networks into one “integrated” genome-scale model, 3) to confirm and validate the ability of the integrated model to predict processing of various environmental signals and hydrogen production. This project combines the state of the art computational and experimental procedures and will significantly advance our systems-level understanding of hydrogen production in *T. maritima* through the generation of new data sets, and their integration with existing knowledge in the form of a genome-scale network reconstruction that can be used to compute physiological functions and states in hydrogen producing organisms.**

This project is focused on a systems-level understanding of biological hydrogen production using *Thermotoga maritima* as a model organism. We will address the basic science required to improve our understanding of hydrogen production from various carbon sources including glucose, cellulose, starch and xylan by this thermophilic microorganism. The overall goal is 1) to reconstruct the regulatory and metabolic network in *T. maritima* using various sets of “omics” data, 2) to integrate regulatory and metabolic networks into one “integrated” genome-scale model, 3) to confirm and validate

the ability of the integrated model to predict processing of various environmental signals.

The current metabolic reconstruction of *T. maritima* contains 479 metabolic genes, 565 metabolites (non-unique) and 646 internal and external metabolic reactions. About 93% of these reactions (519) could be associated with a gene product. The total of 479 genes considered in this reconstruction corresponds to 25% of *T. maritima*'s genome. This reconstruction is based on the integration of the data from the literature supplemented by the extensive use of genomic techniques including a subsystem-based approach implemented in the SEED database. This high gene coverage also illustrates the amount of information being available for major metabolic pathways in *T. maritima*. The *T. maritima*'s metabolic reconstruction covers reactions and pathways of central metabolism, amino acid, nucleotide, lipid, and carbohydrate metabolism. The reconstruction accounts for the metabolism of 45 carbohydrates including some highly complex polysaccharides. Furthermore, the *T. maritima* model is able to produce hydrogen using various carbon sources.

The overall scope, content and quality is comparable with organisms for which more publications are available. Partially, this stage of development can be contributed to the availability of structural and detailed physiological studies. In fact, this reconstruction is the first that used extensively structural information for reaction/gene evidence during the reconstruction process. For example, crystallography data provided annotation evidence in some cases, confirming the function of the annotated gene as well as the candidate substrate and coenzyme utilization specific for the *T. maritima* enzyme.

Building on this metabolic reconstruction, we intend to identify genes and therefore reactions that are “missing” using competitive genomics approaches and computational algorithms. In order to improve the predictive abilities of the current model, we will also re-define the current biomass function by directly measuring the cellular composition of *T. maritima*. We will further reconstruct the transcriptional regulatory network (TRN) of *T. maritima*. TRN reconstruction will be based on a computational approach as well as on experimental methods using chromatin immunoprecipitation combined with whole-genome microarrays (ChIP-chip) or sequencing (ChIP-seq) to experimentally define genome-wide binding patterns of RNA polymerase, major sigma factors, and selected transcription factors in *T. maritima*. These experiments will be complemented with high-resolution gene expression profiles using tiled microarrays. The most important bioinformatics predictions of novel genes and pathways will be supported by focused validation experiments. We will then integrate the metabolic and regulatory networks to generate a comprehensive model of *T. maritima* that will allow for prediction of optimal production of hydrogen. Some examples illustrating our integrated approach to genomic reconstruction and experimental validation of metabolic pathways and regulatory circuits related to carbohydrate utilization are provided in the accompanying poster “Sugar kinome of *T. maritima*”.

GTL

## Systems-Level Kinetic Flux Profiling Elucidates a Bifurcated TCA Cycle in *C. acetobutylicum*

Daniel Amador-Noguez\* (damador@princeton.edu), Jing Fan (jingfan@princeton.edu), Nathaniel Roquet (nroquet@Princeton.EDU), Xiao-Jiang Feng, Herschel Rabitz, and **Joshua D. Rabinowitz** (josh@princeton.edu)

Dept. of Chemistry and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, N.J.

**Project Goals: Microbial hydrogen (H<sub>2</sub>) production holds great promise as a source of renewable clean energy. A critical step towards more efficient biohydrogen production is improved understanding of the regulation of H<sub>2</sub>-related metabolism. While a diverse range of microorganisms are capable of producing H<sub>2</sub>, the complete intracellular pathways involved in H<sub>2</sub> production are known for only a few of them, and the full regulation of these pathways for none. Advances in systems biology, analytical chemistry, and computer science are beginning to provide the tools required for effective quantitative modeling of complex biological networks. A critical goal is to develop models that are sufficiently accurate to enable rational control of the network behavior. The cellular metabolic pathways for H<sub>2</sub> production form a biological network whose rational control would have profound value. With the long term aim of enabling such control, we propose to develop integrated experimental-computational technologies for quantitative dissection of microbial hydrogen-producing metabolism. These tools will be broadly applicable to many microbial H<sub>2</sub> producers. We plan to illustrate them with two organisms: *Clostridium acetobutylicum* (possessing the fastest and highest yielding hexose fermentation pathway to H<sub>2</sub> of any microbe yet reported), and a novel class of thermophilic cyanobacteria that lack hydrogenase and produce H<sub>2</sub> at 62°C via a nitrogenase-dependent pathway (*Synechococcus* species).**

*Clostridium acetobutylicum* is an organism with great potential for the commercial production of butanol and hydrogen. Much work has been done to elucidate the metabolic pathways by which solvents are produced in this anaerobic bacterium. However, there are still key pathways of primary metabolism that remain unresolved, including the TCA cycle and amino acid biosynthesis.

Annotation of the genome sequence of *C. acetobutylicum* failed to identify obvious homologues of the enzymes of the right-side of the TCA cycle: citrate synthase, aconitase, and isocitrate dehydrogenase. These enzymes are traditionally required to synthesize  $\alpha$ -ketoglutarate, the carbon skeleton of the glutamate-family of amino acids. To explain *C. acetobutylicum*'s ability to synthesize glutamate and grow on minimal media, it was proposed that the TCA cycle

could function in the reductive (counterclockwise) direction to produce  $\alpha$ -ketoglutarate. Alternatively, it was suggested that glutamate is synthesized from ornithine by the arginine biosynthesis pathway running in reverse.

To elucidate the actual pathway that leads to  $\alpha$ -ketoglutarate and glutamate, we studied the incorporation of various isotope-labeled nutrients into intracellular metabolites using liquid chromatography-tandem mass spectrometry (LC-MS/MS). These studies quickly ruled out both of the above hypotheses: feeding U-<sup>13</sup>C-glucose labeled both succinate and  $\alpha$ -ketoglutarate with similar kinetics, but the labeling patterns were inconsistent with carbon exchange between the compounds; feeding U-<sup>13</sup>C ornithine labeled downstream arginine pathway compounds but not upstream ones or glutamate.

The actual route of  $\alpha$ -ketoglutarate production was suggested by the observation that (despite the putative lack of citrate synthase), citrate was labeled faster than  $\alpha$ -ketoglutarate upon feeding either U-<sup>13</sup>C-glucose or U-<sup>13</sup>C-acetate, with the labeling patterns of citrate, aconitate, and  $\alpha$ -ketoglutarate consistent with turning of the right side of the TCA cycle in the oxidative direction. Interestingly, there was no passage of carbon from acetate into succinate, confirming the bifurcation of the TCA cycle.

In most organisms, the pro-chiral center of citrate has *S* stereochemistry. The position of <sup>13</sup>C atoms in glutamate and proline obtained from the U-<sup>13</sup>C glucose and U-<sup>13</sup>C acetate labeling experiments indicated, however, the production of citrate with *R* stereochemistry at the prochiral center. A homology search revealed a gene in *C. acetobutylicum* with ~60% sequence similarity to the *Re*-citrate synthase present in *C. kluyveri*. Efforts are ongoing to directly demonstrate the citrate synthase activity of the encoded protein biochemically.

Our results demonstrate that *C. acetobutylicum* has a bifurcated TCA cycle where  $\alpha$ -ketoglutarate is produced in the oxidative direction from oxaloacetate and acetyl-CoA via citrate, and succinate is produced in the reductive direction from oxaloacetate via malate and fumarate. This observation is essential for construction of an accurate genome-scale model of *C. acetobutylicum* metabolism and lays the groundwork for better understanding of integration of biosynthetic metabolism with solvent and hydrogen production.

GTL

## Systems Approach to Probing Hydrogen Regulation (SAPHyRe): The SurR Redox-Switched Transcriptional Regulator Controlling Hydrogen Production in *Pyrococcus furiosus*

Gina L. Lipscomb,<sup>1\*</sup> Hua Yang,<sup>1,3</sup> Annette Keese,<sup>2</sup> Gerrit J. Schut,<sup>1</sup> B.C. Wang,<sup>1</sup> Michael Thomm,<sup>2</sup> Janet Westpheling,<sup>1</sup> Michael W.W. Adams,<sup>1</sup> and **Robert A. Scott**<sup>1</sup> (rscott@uga.edu)

<sup>1</sup>Dept. of Biochemistry & Molecular Biology, University of Georgia, Athens, Ga.; <sup>2</sup>Dept. of Microbiology, University of Regensburg, Germany; and <sup>3</sup>Centers for Disease Control and Prevention, Atlanta, Ga.

**Project Goals: The goal of SAPHyRe (Systems Approach to Probing Hydrogen Regulation) is to develop a detailed systems-level description of the regulatory and metabolic networks controlling hydrogen production in the hyperthermophilic archaeon *Pyrococcus furiosus* (Pf). Pf will be used as the model organism to investigate its response to various environmental conditions relevant to all hydrogen-producing microorganisms, such as carbon and nitrogen sources, metal availability, and oxidative and reductive stresses. The outcome of this project will serve two purposes: 1) it will bring us one step closer to utilizing *P. furiosus* in development of H<sub>2</sub> as an alternative energy source and 2) it will serve as a model methodology for investigating the regulatory pathways of hydrogen production in other organisms.**

*Pyrococcus furiosus* has the potential to produce hydrogen (H<sub>2</sub>) efficiently from biomass, making it an excellent target organism for development of this alternative energy source. SAPHyRe (Systems Approach to Probing Hydrogen Regulation) explores the regulatory and metabolic networks controlling H<sub>2</sub> production in *P. furiosus* and will provide a model for investigating such networks in other H<sub>2</sub>-producing organisms. A major control point for H<sub>2</sub> production in *P. furiosus* involves sulfur (S<sup>0</sup>), which when present, causes a shift in metabolism from production of H<sub>2</sub> to H<sub>2</sub>S. A key transcriptional regulator involved in this shift is the S<sup>0</sup> Response Regulator, SurR, which activates the hydrogenase operons along with related genes and represses genes involved in S<sup>0</sup> metabolism such as the recently described NAD(P)H Sulfur Reductase. SurR was discovered by affinity capture from cell extract with a bait DNA promoter region of the membrane-bound hydrogenase operon which is significantly down-regulated during the primary response to S<sup>0</sup>. Biophysical, functional and structural characterization of SurR led to the discovery of an internal disulfide switch that controls its DNA-binding activity. The switch is an imbedded CxxC motif within the N-terminal helix-turn-helix DNA-binding domain, which when in the reduced thiol state, causes SurR to bind DNA in a sequence-specific manner. In the presence of S<sup>0</sup>, the CxxC motif becomes

oxidized to a disulfide, and SurR can no longer exert transcriptional control, resulting in deactivation of hydrogenase operons and derepression of S<sup>0</sup>-metabolizing genes. This exemplary regulatory system mediating the interplay of H<sub>2</sub> and S<sup>0</sup> metabolism will be used in SAPHyRe as a guide to further our understanding of how H<sub>2</sub> production is affected by other key factors such as carbon and nitrogen sources, metal availability, and oxidative and reductive stresses.

GTL

## Filling Knowledge Gaps in Biological Networks: Integrated Global Approaches to Understand H<sub>2</sub> Metabolism in *Chlamydomonas reinhardtii*

Christopher H. Chang,<sup>1</sup> Alexandra Dubini,<sup>1</sup> Florence Mus,<sup>2</sup> Venkataraman Subramanian,<sup>1,3</sup> David M. Alber,<sup>1</sup> Peter A. Graf,<sup>1</sup> Glenn Murray,<sup>3</sup> Kwiseon Kim,<sup>1</sup> Arthur R. Grossman,<sup>2</sup> Matthew C. Posewitz,<sup>3</sup> and **Michael Seibert**<sup>1\*</sup> (mike\_seibert@nrel.gov)

<sup>1</sup>National Renewable Energy Laboratory, Golden, Colo.; <sup>2</sup>Carnegie Institution of Washington, Stanford, Calif.; and <sup>3</sup>Colorado School of Mines, Golden, Colo.

**Project Goals: Development of photobiological H<sub>2</sub>-production processes, a key component in pursuit of DOE's renewable energy mission, would be substantially accelerated by increasing our understanding of the extremely complex underlying biology. Although systems biology has evolved rapidly in recent years, the lack of comprehensive experimental data for a given organism prevents reliable predictive modeling based on biophysical representations. We will therefore employ petascale computing to address this issue by (1) computational parameter estimation to delimit the space of stable solutions for experimentally constrained metabolic models and (2) de novo numerical experimentation to characterize network capabilities under assumed but easily reconfigured kinetic models. The response will be characterized at the level of enzyme kinetic differential equation parameters. Through development of scalable software tools, iterative model building, and incorporation of experimental constraints generated by high-throughput "omics" technologies, a model of metabolism linked to H<sub>2</sub> production in the green alga, *Chlamydomonas reinhardtii*, will be constructed and tested biologically by systems biology approaches. Once a set of acceptable kinetic parameters has been computed, the model will then be used for high-performance optimization of H<sub>2</sub> output in the space of enzyme expression levels, subject to limitations on cell viability. Integration of a graphical job configuration interface into the popular Systems Biology Workbench will make our tools accessible to a broader user community. The work is envisioned as an important contribution toward long-term development of a complete in silico cell.**

The goal of this project, jointly funded by the DOE GTL and SciDAC Programs, is to develop the means to globally map *in silico* all biological pathways in *Chlamydomonas reinhardtii* that can affect H<sub>2</sub> production by the organism. *C. reinhardtii* was the first alga with a sequenced genome (released in 2003 by the JGI), and it has recently emerged as a prototype organism for investigating fermentative processes and their regulation. *Chlamydomonas* has a complex anaerobic metabolic network that can be induced under dark, anaerobic conditions, where it can produce H<sub>2</sub> along with other fermentation products such as formate, acetate, ethanol, and CO<sub>2</sub>. Previous studies focused on determining genes (by microarray and RT-PCR) that were differentially regulated as the result of shifting cultures of the parental strain CC-425 from aerobic growth to dark, anaerobiosis [1]. Indeed, anoxia led to differential expression of genes involved in fermentation and more specifically upregulation of the pyruvate formate lyase (PFL1) and pyruvate:ferredoxin oxidoreductase (PFR1) genes. Moreover, *C. reinhardtii* synthesized formate, acetate, and ethanol in the ratio 1:1:0.5. More recent experiments showed that in a strain lacking active hydrogenases, the ratio shifted to 2:1:1, respectively, and that succinate was produced instead of H<sub>2</sub>. Interestingly the levels of transcripts encoding several proteins involved in fermentation also changed. Under this condition PFL1 was more upregulated and PFR1 was downregulated relative to the wild-type strain [2]. Together these results allowed us to establish a fermentative pathway model that provides information on the metabolic flexibility of the organism and H<sub>2</sub> metabolism.

In order to apply this and other experimental results discussed below, we have developed a computational model that integrates metabolic parameters obtained experimentally. We have encoded central carbon metabolism (glycolysis and the tricarboxylic acid cycle), oxidative phosphorylation, and fermentative metabolism known to occur in *C. reinhardtii* in the community-standard Systems Biology Markup Language. The model spans the cytosol; the mitochondrial intermembrane space, membranes, and matrix; and the chloroplast stroma. More explicitly, the modeling is based on a thermodynamic/kinetic paradigm, in which we seek to incorporate basic physical constraints and experimentally observed kinetic relationships directly. Merging of component pathway models, parsing of the merged model to C++ encoding the relevant ordinary differential equations, compiling, and linking to form a model-specific, high-performance executable program was accomplished through our software package, the *High-Performance Systems Biology Toolkit* (HiPer SBTK) [3]. A simple “make” process results in an executable program capable of simultaneous sampling and optimizing metabolite levels, fluxes, or sensitivities within the space of kinetic parameters or enzyme concentrations, with good scaling observed to 128 processing cores so far. Both local and global optimization is possible. Job configuration is possible through an auxiliary graphical interface or by direct editing of simple text. Stable model integration depends critically on the initial guess for metabolite levels, which in the absence of experimental data is generated by small manual trials. Key developments going forward include partial automation of initial guess genera-

tion, model expansion to include starch metabolism and photosynthetic light and dark reactions, standardized file format output for advanced visualization, and simultaneous sampling, fitting, and extremization of metabolites, fluxes, and sensitivities.

Our current biological research, which is supporting the computational studies, is focused on understanding the regulation of relevant pathways using proteomics-based approaches. Parental *C. reinhardtii* strain CC-425 was again grown under conditions where the cells were shifted from light-aerobic growth to dark-anoxic, H<sub>2</sub>-producing conditions. Whole-cell protein extracts from these cultures were separated on 2D gels. The first dimension (isoelectric focusing) used a 24-cm, non-linear Immobiline pH Gradient (IPG) strip (pH 3-10), and the second dimension used a 12-14% polyacrylamide gradient gel. Proteins were then visualized using SYPRO Ruby staining, and analyzed with ImageMaster 2D Platinum 6.0 software to identify proteins differentially regulated under the two conditions. Preliminary analysis revealed that approximately 500 proteins were expressed under dark, anoxic conditions, compared to 600 under light, oxic condition. Further analysis to identify the differentially expressed proteins is in progress using a Micromass Quantitative-Time-of-Flight (Q-Tof) mass spectrometer.

In summary, high-throughput ‘omics’ techniques are examining the metabolic flexibility of *Chlamydomonas* and are being used to build metabolic computational models of the alga. Information obtained from the *in silico* modelling is expected to improve understanding of complexity of the metabolic network system, how cells adjust to changes in metabolite fluxes, what happens when specific metabolic reactions are blocked, and how the organism might be engineered to improve H<sub>2</sub>-production yields.

## References

1. Mus, F., Dubini, A., Seibert M., Posewitz M.C., and Grossman A.R. (2007) “Anaerobic acclimation in *Chlamydomonas reinhardtii*: Anoxic gene expression, hydrogenase induction and metabolic pathways”, *J. Biol. Chem.* **282** (35), 25475-25486.
2. Dubini, A., Mus, F., Seibert M., Grossman, A.R., and Posewitz M.C., (2008) “Flexibility in anaerobic metabolism as revealed in a mutant of *Chlamydomonas reinhardtii* lacking hydrogenase activity”, *J. Biol. Chem.* *in press*.
3. Chang, C.H., Graf, P., Alber, D.M., Kim, K., Murray, G., Posewitz, M., and Seibert M., (2008) “Photons, Photosynthesis, and High-Performance Computing: Challenges, Progress, and Promise of Modeling Metabolism in Green Algae”, *J. Phys. Conf. Ser.* **125**: 012048.

## Examining the Molecular Basis for the Utilization of Alternative Redox Systems to Maximize Hydrogen Production in RubisCO-Compromised Mutants of Nonsulfur Purple Bacteria

Rick A. Laguna\* (laguna.2@osu.edu) and F. Robert Tabita

Dept. of Microbiology, Ohio State University, Columbus, Ohio

**Project Goals:** The use of hydrogen as an energy source is attractive because the end-product is water, as opposed to carbon dioxide, a green house gas generated by the burning of fossil fuels. Hydrogen is normally produced by nonsulfur purple (NSP) photosynthetic bacteria during nitrogen limiting conditions by the nitrogenase complex. However, the presence of ammonia in the environment normally repressed this process. We have shown that NSP photosynthetic bacteria possess an array of metabolic and regulatory capabilities that allow for the utilization of alternative redox sinks when the primary electron sink, carbon dioxide, is nullified via the inactivation or deletion of the RubisCO genes. These RubisCO-compromised mutant strains are able to derepress the synthesis of the nitrogenase complex under normal repressive conditions, thus allowing for the production of copious quantities of hydrogen gas. Such gain-of-function adaptive mutant strains have been obtained from *Rhodobacter capsulatus*, *Rhodobacter sphaeroides*, *Rhodospirillum rubrum*, and *Rhodopseudomonas palustris*. We are in the process of gaining a greater understanding of the molecular basis that allows for the utilization of the nitrogenase complex under normal repressive conditions, in order to maximize hydrogen production in RubisCO-compromised mutants of nonsulfur purple bacteria.

The Calvin-Benson-Bassham (CBB) reductive pentose phosphate pathway is responsible for the incorporation of CO<sub>2</sub> into cellular carbon under autotrophic growth conditions. Under photoheterotrophic conditions, CO<sub>2</sub> is primarily used as an electron acceptor via the CBB cycle in order to maintain redox poise within the cell. The key enzyme of the CBB pathway, RubisCO, catalyzes the actual CO<sub>2</sub> reduction step. Over the years (1-4), we have shown that nonsulfur purple (NSP) photosynthetic bacteria possess an array of metabolic and regulatory capabilities (5) that allow for the utilization of alternative redox sinks when the primary electron sink, CO<sub>2</sub>, is nullified via the inactivation or deletion of the RubisCO genes. In order to grow photoheterotrophically, such RubisCO-compromised strains develop interesting strategies and alter their basic metabolic profile. For example, in many instances the derepression of nitrogenase synthesis occurs under normal repressive conditions. Such gain-of-function adaptive mutant strains have been obtained from *Rhodobacter capsulatus*, *Rhodobacter sphaeroides*, *Rhodospirillum rubrum*, and *Rhodopseudomonas*

*palustris*, whereby such strains balance their redox potential via nitrogenase-catalyzed reduction of protons to hydrogen gas (1-4). Moreover, over the years we have shown that nitrogenase-derepressed mutant strains produce copious quantities of hydrogen gas by virtue of using the nitrogenase enzyme complex exclusively as a hydrogenase. We now show that knocking out competing redox balancing processes such as the CBB pathway plays a crucial role in maximizing hydrogen production in nitrogenase-derepressed strains.

We are in the process of gaining a greater understanding of the molecular basis that allows for the utilization of the nitrogenase complex under normal repressive conditions, particularly in RubisCO-compromised strains. A single point mutation in the *nifA* gene was shown to be important for nitrogenase derepression in RubisCO-compromised mutant strains of *R. capsulatus*, *R. sphaeroides*, and *R. palustris*. *NifA* is a key transcriptional activator of the structural genes encoding the nitrogenase complex (*nif*/HDK). While current experiments suggest that a mutant *NifA* protein appeared to be responsible for derepression of the nitrogenase complex in *R. palustris*, in *R. sphaeroides* an additional thus far unidentified mutation appears to be involved in the derepression of the nitrogenase complex. Interestingly, no such *nifA* mutation was found in a nitrogenase-derepressed strain of a RubisCO knockout strain of *R. rubrum*. These results suggest that these organisms utilize different mechanisms to derepress nitrogenase synthesis in CBB-compromised mutant strains. Moreover, derepression of the nitrogenase complex is not the only way in which these organisms may balance their redox potential in the absence of a functional CBB pathway; e.g., some strains may derepress the synthesis of a novel sulfate reduction pathway or use other means to dissipate excess reducing equivalents. In all cases, the alternative redox balancing pathways are under exquisite control and are not maximally expressed unless the CBB pathway is knocked out, suggesting regulatory cross talk between these key pathways.

### References

1. Joshi, H. M., and F. R. Tabita. A global two component signal transduction system that integrates the control of photosynthesis, carbon dioxide assimilation, and nitrogen fixation. *Proc. Natl. Acad. Sci. USA.* 93 (1996), 14515-14520
2. Tichi, M. A., and Tabita, F.R. Maintenance and control of redox poise in *Rhodobacter capsulatus* strains deficient in the Calvin-Benson-Bassham pathway. *Arch. Microbiol.* 174 (2000) 322-333.
3. Tichi, M. A. and Tabita, F.R. Interactive control of *Rhodobacter capsulatus* redox balancing systems during phototrophic metabolism. *J. Bacteriol.* 183 (2001) 6344-6354.
4. Dubbs, J. M., and Tabita, F. R. Regulators of nonsulfur purple phototrophic bacteria and the interactive control of CO<sub>2</sub> assimilation, nitrogen fixation, hydrogen metabolism and energy generation. *FEMS Microbiol. Rev.* 28 (2004) 353-376.
5. Romagnoli, S., and Tabita, F. R. Carbon dioxide metabolism and its regulation in nonsulfur purple photosynthetic bacteria. In, Hunter C.N., Daldal F., Thurnauer M.C., and Beatty J. T. (2008) *The Purple Phototrophic Bacteria. Advances in Photosynthesis and Respiration*, Vol. 28, Springer, Dordrecht, The Netherlands, pp. 563-576

GTL

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

Jonathan Millen<sup>1\*</sup> (jonathan\_millen@hotmail.com), Matthew Heckman,<sup>1</sup> Hsin-Yi Tsai,<sup>1</sup> Pin-Ching Maness,<sup>2</sup> Mary Lipton,<sup>3</sup> and J.H. David Wu<sup>1</sup> (davidwu@che.rochester.edu)

<sup>1</sup>University of Rochester, Rochester, N.Y.; <sup>2</sup>National Renewable Energy Laboratory, Golden, Colo.; and <sup>3</sup>Pacific Northwest National Laboratory, Richland, Wash.

**Project Goals: The overall objective of this research is to understand the H<sub>2</sub> metabolic pathway in this cellulose-degrading bacterium and its regulatory network at the molecular and systems levels. We hypothesize that: multiple hydrogenases work concertedly to contribute to hydrogen evolution; transcription factors control hydrogenase expression at the transcription level; and hydrogen metabolism is coupled to cellulolysis and fermentative pathways in this bacterium.**

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

The overall objective of this research is to understand the H<sub>2</sub> metabolic pathway in this cellulose-degrading bacterium and its regulatory network at the molecular and systems levels. We hypothesize that: multiple hydrogenases work concertedly to contribute to hydrogen evolution; transcription factors control hydrogenase expression at the transcription level; and hydrogen metabolism is coupled to cellulolysis and fermentative pathways in this bacterium.

We are determining the hydrogenase expression and metabolic network nodes employing Real-Time RT-PCR, DNA microarray, and proteomic analyses on cells subjected to different culture conditions and metabolic pathway inhibitors, to probe differential expression of the various hydrogenases and their interrelationship with other cellular metabolic pathways. EMSA (electrophoretic mobility shift assay)

coupled with mass spectrometry as well as the bacterial one-hybrid method are employed to identify transcription factors that bind to the promoters of the hydrogenase gene clusters. To extend quantitative analysis of genome expression to the protein level, we have initiated development of an Accurate Mass and Time (AMT) tag database for *C. thermocellum*. We have also established the methodology to profile system level changes in cellular protein expression in *C. thermocellum* using two-dimension separation and analysis on the ProteomeLab™ PF2D platform, a 2D chromatographic technique.

We will carry out ChIP-(chromatin immuno-precipitation-) on-chip assays targeting at transcription factors thus identified to map connections in the transcription factor network controlling linked metabolic pathways. Finally, we will purify FeFe-hydrogenases from its native producer and expressed recombinantly *E. coli* to determine their subunit compositions, endogenous redox partners, and the direction of reaction (hydrogen production vs. uptake) to shed light on their roles in hydrogen metabolism

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

GTL

## Inference and Integration of Regulatory Dynamics in Metabolic Network Models

William J. Riehl<sup>1\*</sup> (briehl@bu.edu), Hsuan-Chao Chiu<sup>1\*</sup> (hcchiu@bu.edu), and Daniel Segre<sup>1,2</sup> (dsegre@bu.edu)

<sup>1</sup>Graduate Program in Bioinformatics and <sup>2</sup>Depts. of Biology and Biomedical Engineering, Boston University, Boston, Mass.

<http://prelude.bu.edu>

**Project Goals: The goal of this project is to develop new mathematical approaches for integrating regulatory information into metabolic network models. In addition to using experimental data for the inference of dynamically changing objective functions, we implement an ab initio approach to generate putative regulatory networks capable of optimal control.**

Cellular metabolism dynamically adapts to changing environmental conditions and biological events. Maintenance of metabolic homeostasis is achieved through the regulation of metabolic enzymes in a complex interplay of transcriptional

and post-transcriptional mechanisms. In the context of genome-scale steady state models of metabolic networks, it has been shown that evolution may drive metabolic networks towards reaching computationally predictable optimal states, such as maximal growth capacity. However, the accuracy of these *in silico* models may vary significantly depending on environmental conditions, genetic perturbations, and the state of complex, and often unknown regulatory constraints. Many of these details, including the values of most kinetic parameters and the regulatory mechanisms that reprogram the metabolic pathways, have proven difficult to elucidate. Thus, understanding the principles that underlie metabolic regulation, and identifying approachable computational strategies for integrated modeling of metabolic and regulatory networks constitute an important open challenge, relevant for effectively modeling microbial systems towards increased production of valuable resources, such as biohydrogen.

Here, we propose two complementary approaches to this challenge, both extending current genome-scale steady-state models of metabolism. In the first approach [1], we develop a data-driven method based on linear optimization to infer time- or condition-dependent metabolic objectives, in the form of dynamically changing biomass compositions. Specifically, we seek to use experimental data (such as flux measurements, or mRNA expression levels) to infer best matching stoichiometrically balanced fluxes and metabolite sinks that maximally describe the metabolic objectives of multiple physiological states in an organism. We tested our approach by analyzing *E. coli* central carbon metabolism single gene mutants, and identified changes of metabolic objectives and potential compensation for reducing power in the pentose phosphate pathway. In the second, theory-driven approach [2], we develop a new mathematical method for predicting the regulatory network that controls a metabolic pathway, based on optimality criteria similar to the ones used in Flux Balance Analysis. The model hypothesizes that, once a steady-state metabolic system is perturbed, regulatory feedback based on the flux imbalance induced by the perturbation will gradually restore homeostasis. This feedback can either bring the system back to the original steady-state or, in a switch-like behavior, take it to a different state. This model has been applied to predicting optimal regulatory responses both in a simple toy network and in a fragment of the glycolysis pathway.

## References

1. Hsuan-Chao Chiu and Daniel Segrè: Comparative determination of biomass composition in differentially active metabolic states, *Genome Informatics*, 2008, *In press*.
2. William J. Riehl and Daniel Segrè: Optimal metabolic regulation using a constraint-based model, *Genome Informatics*, 2008, *In press*.

## Effects of Mutational Modification of Electron Pathways of *Desulfovibrio* Strains

Kimberly L. Keller<sup>1,2\*</sup> (kellerkl@missouri.edu), Steve D. Brown,<sup>3</sup> Iris Porat,<sup>3</sup> and Judy D. Wall<sup>1,2</sup> (WallJ@missouri.edu)

<sup>1</sup>University of Missouri, Columbia, Mo.; <sup>2</sup>VIMSS (Virtual Institute for Microbial Stress and Survival), <http://vimss.lbl.gov>, Berkeley, Calif.; and <sup>3</sup>Oak Ridge National Laboratory, Oak Ridge, Tenn.

**Project Goals: We propose to examine the capacity of the abundant soil anaerobes, sulfate-reducing bacteria, for hydrogen generation from organic acids. A comparison of *D. vulgaris* Hildenborough and *Desulfovibrio* G20 should provide insights into the limitations and potential of hydrogen production by members of this large group of strictly anaerobic bacteria. To explore this metabolic capacity, we will: 1) create a deletion of the gene encoding acetate kinase to confirm that this enzyme is essential for substrate-level phosphorylation during pyruvate fermentation, and determine the effect of this deletion on the efficiency of pyruvate and lactate respiration; 2) determine the enzyme(s) responsible for oxidizing pyruvate during fermentation and the role of formate, if any, in pyruvate fermentation; and 3) channel electrons from alternative sinks to hydrogen during fermentation and determine the effects of removal of those sinks on the fermentation efficiency. The total metabolism, flux through the pathways, and regulation are likely to be limiting factors which we can elucidate in our experiments.**

Fermentative hydrogen generation provides a mechanism for anaerobic microbes to release electrons in a neutral fashion during oxidation of vast quantities of organic matter. We are examining the capacity of the abundant soil anaerobes, sulfate-reducing bacteria, for hydrogen generation from organic acids. These apparently simple pathways have yet to be clearly established. Because the tools for genetic and molecular manipulation of sulfate-reducing bacteria of the genus *Desulfovibrio* are functional, our efforts are focusing on two strains, *D. vulgaris* Hildenborough and *Desulfovibrio* G20. A markerless deletion protocol has already been established in *D. vulgaris* Hildenborough and preliminary experiments with G20 show that the counterselectable marker on which the protocol is based is quite effective. Since G20 is able to grow reproducibly in defined medium in the absence of sulfate as a terminal electron acceptor on a number of substrates, we are using microarray analysis to compare the changes in gene expression between wild-type G20 and various G20 mutants that may be affected in fermentation. We have already begun analysis of hydrogen production and metabolite changes of key G20 Tn5 mutants (generously made available through the ESPP2 collaboration from A. Deutschbauer and A. Arkin). Ultimately, markerless deletions of any genes found to have an impact on hydrogen production will be constructed, allowing us to make sequen-



tial gene deletions to fully explore the metabolic processes that may channel reductant to hydrogen.

GTL

## Modeling Electron Flow in *Rhodobacter sphaeroides* for the Identification of Potential Approaches to Maximize Hydrogen Production

L. Safak Yilmaz<sup>1\*</sup> (yilmaz@cae.wisc.edu), Alison P. Sanders, Rodolfo Perez,<sup>1</sup> Wayne S. Kontur,<sup>2</sup> Timothy J. Donohue<sup>2</sup> (tdonohue@bact.wisc.edu), and Daniel R. Noguera<sup>1</sup> (noguera@engr.wisc.edu)

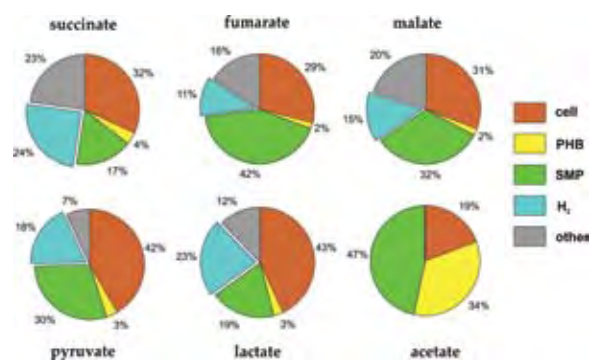
<sup>1</sup>Dept. of Civil and Environmental Engineering and  
<sup>2</sup>Dept. of Bacteriology, University of Wisconsin, Madison, Wis.

**Project Goals:** To develop electron balance and metabolic network models for the prediction of potential electron sinks in *Rhodobacter sphaeroides* which can be modified to maximize solar-powered hydrogen production by this organism.

**Introduction and Objective:** The metabolically versatile organism *Rhodobacter sphaeroides* produces hydrogen while using light as an energy source and organic substrates as electron donor. This project aims at optimizing hydrogen production from *R. sphaeroides* by identifying the main electron sinks and genetically modifying the organisms to divert as high a fraction of electrons as possible to hydrogen generation.

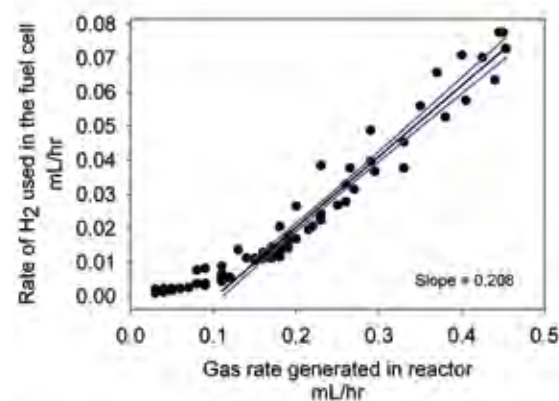
**Methodology:** Cultures were grown in illuminated anaerobic batch reactors until stationary phase was reached. The overall electron flux to biomass, soluble microbial products (SMPs), and polyhydroxybutyrate (PHB) were quantified by chemical oxygen demand measurements of liquid and solid phases. Gas production was measured with respirometry and the hydrogen content of the biogas was determined by gas chromatography from endpoint headspace samples. Furthermore, a fuel cell connected to the respirometer was used as an additional hydrogen sensor.

Two models are being developed. First, an overall electron flow model is used to understand the global distribution of electrons, and second, a metabolic model of *R. sphaeroides* is being developed for detailed analyses of metabolic pathways and their potential contribution to hydrogen production. So far, a central metabolic network (CMN) that involves carbon metabolism, electron transport chain, and flow of electrons to hydrogen has been established. Computational flux balance analyses (FBA) have been performed to qualitatively test the model.



**Figure 1.** Distribution of substrate electrons in photosynthetically grown *R. sphaeroides*. Glutamate was used as a nitrogen source to promote hydrogen production in all tests. The carbon source was varied as indicated.

**Results:** The electron distribution from the growth of wild type cells with six different organic acids as carbon sources is given in Figure 1. Generally, an electron balance accounting for more than 80% of the reducing power was achieved. Hydrogen production efficiency was best from lactate and succinate with about a quarter of electrons flowing to hydrogen. Only acetate did not support detectable hydrogen production, possibly because of increased flow of electrons to PHB. The fraction of PHB was small for other substrates, while SMPs turned out to be an important electron sink in most experiments. Although hydrogen quantification was based on headspace measurements, the potential of using small-scale fuel cells as on-line sensors was tested. A linear relationship between gas production rate and voltage output in the fuel cell was observed in a wide range of gas rates (see Figure 2).



**Figure 2.** Conversion of biohydrogen to electricity in a proton exchange membrane-based fuel cell. The y-axis shows hydrogen used by the fuel cell, back-calculated from the voltage readout. The linear relationship suggests that hydrogen can be quantified by the fuel cell in a certain range of production rate. The slope indicates that around 20% of the hydrogen gas generated in the reactor was converted to electricity in the fuel cell.

The metabolic modeling effort has so far centered on creating the CMN model. At present, it correctly simulates growth (flux to biomass) coupled to hydrogen production

when an organic acid is used as the substrate, light is the energy source, and no external electron acceptors are provided. Consistent with experimental observations, the model predicts no hydrogen production when cells are grown under aerobic conditions in the absence of light.

**Future Work:** The electron balance modeling approach will continue to be used as a tool to quantify electron fate on wild type cultures under different growth conditions (e.g., different substrates and nitrogen sources and different growth stages) and on mutants generated to investigate the metabolic networks that impact hydrogen production. Experimental results will also be used to continue the development of the metabolic network model, which will be expanded to include metabolic networks necessary to simulate experimental observations with wild type and mutant strains.

## SBIR Bioenergy Research

SBIR

### Consolidated Bioelectrochemical Processing of Cellulosic Biomass to Ethanol and Hydrogen

T. Shimotori<sup>1\*</sup> (shimo@mfctech.net), M. Nelson,<sup>1</sup> and H. May<sup>2</sup>

<sup>1</sup>MFC Technologies, LLC, Mount Pleasant, S.C.; and

<sup>2</sup>Dept. of Microbiology and Immunology, Medical University of South Carolina, Charleston, S.C.

**Project Goal: Enhance cellulosic ethanol fermentation through the use of an ethanol and electricity generating microbial consortium in a consolidated bioprocess.**

Consolidated bioprocessing of cellulosic biomass, as originally proposed by Dr. Lynd, leverages the catalytic activity of cellulolytic and ethanologenic bacteria to produce ethanol.<sup>1,2</sup> Thermophilic bacteria such as *Clostridium thermocellum* and *Thermoanaerobacter thermosaccharolyticum* are typically used in this process due to their rapid and effective ability to metabolize cellulose and hemicellulose. However, the formation of by-products, especially fatty acids, is common. The by-product formation not only reduces the ethanol yield but also inhibits the overall fermentation. As a solution to this problem, Dr. Lynd's group recently reported a method to reduce by-product yields by genetic engineering of a cellulolytic bacterium.<sup>3</sup>

As an alternative way to tackle this problem, MFC Tech proposed to use a consolidated bioelectrochemical process to convert cellulosic biomass into ethanol or hydrogen. Electricigenic bacteria capable of consuming fatty acids are incubated with cellulolytic bacteria in thermophilic microbial fuel cells (MFCs), where fatty acids are converted to carbon dioxide and electricity is generated. MFCs have been extensively studied for years as summarized in Dr. Logan's

review.<sup>4</sup> Only recently, however, a few research groups reported such thermophilic electricigenic bacteria.<sup>5-7</sup> Dr. May and colleagues at MUSC successfully enriched a mixed community of thermophilic electricigenic bacteria from marine sediment that consume acetate in MFCs at 60°C.<sup>6</sup> Analysis of the 16S rRNA genes revealed a community that includes bacteria most closely related to *Deferribacter* spp.<sup>8-10</sup> and Gram positive Firmicutes, particularly of *Thermincola* spp.<sup>11,12</sup> We used this mixed culture and an isolate of *Thermincola ferriacetica* as thermophilic electricigenic bacteria in combination with cellulolytic bacteria in cellulose-fed MFCs.

We operated air-cathode MFCs with the thermophilic electricigenic bacteria at 60°C. Once the MFCs started producing steady current using acetate, the anode chamber was filled with cellulose-containing media and inoculated with *C. thermocellum*. Although these MFCs produced ethanol and electricity, ethanol production was lower and acetate production was higher compared with cellulosic fermentation without MFC. This is because the reducing potential necessary for ethanol production was channeled to electricity production.

Due to the failure of a consolidated bioelectrochemical process in a MFC, we adapted it to a microbial electrolysis cell (MEC). A MEC, although similar to a MFC, is entirely anaerobic and can produce hydrogen from fatty acids with the help of added electricity. For a review of MECs, see Dr. Logan's recent report.<sup>13</sup> Consolidated bioelectrochemical processing of cellulose in thermophilic MECs using the electricigenic bacteria and *C. thermocellum* resulted in higher ethanol/acetate ratios compared with those by fermentation alone. An even better result was obtained when the electricigenic bacteria and *C. thermocellum* were co-cultured with *T. thermosaccharolyticum* in a MEC. We also demonstrated hydrogen production from cellulosic biomass in a similar system.

### References

1. Lynd, L.R. *et al.* Consolidated bioprocessing of cellulosic biomass: an update. *Curr. Opin. Biotechnol.* **16**, 577-583 (2005).
2. Lynd, L.R. *et al.* Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* **66**, 506-577 (2002).
3. Shaw, A. J. *et al.* Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. *PNAS.* **105**, 13760-13774 (2008).
4. Logan, B.E. *Microbial fuel cells.* Wiley-Interscience (2008).
5. Jong, B. *et al.* Enrichment, performance and microbial diversity of a thermophilic mediatorless microbial fuel cell. *Environ. Sci. Technol.* **40**:6449-6454 (2006).
6. Mathis, B. J. *et al.* Electricity generation by thermophilic microorganisms from marine sediment. *Appl. Microbiol. Biotechnol.* **78**:147-155 (2008).
7. Wrighton, K. C. *et al.* A novel ecological role of the Firmicutes identified in thermophilic microbial fuel cells. *ISME J.* **2**:1146-1156 (2008).
8. Miroshnichenko, M. L. *et al.* *Deferribacter abyssi* sp. nov., an anaerobic thermophile from deep-sea hydrothermal

- vents of the Mid-Atlantic Ridge. *Inter. J. Syst. Evol. Microbiol.* **53**, 1637-1641 (2003).
9. Takai, K. *et al.* *Deferribacter desulfuricans* sp. nov., a novel sulfur-, nitrate- and arsenate-reducing thermophile isolated from a deep-sea hydrothermal vent. *Inter. J. Syst. Evol. Microbiol.* **53**, 839-846 (2003).
  10. Greene, A. C. *et al.* *Deferribacter thermophilus* gen. nov., sp. nov., a novel thermophilic manganese- and iron-reducing bacterium isolated from a petroleum reservoir. *Int. J. Syst. Bact.* **47**, 505-509 (1997).
  11. Sokolova, T. G. *et al.* *Thermincola carboxydiphila* gen. nov., sp. nov., a novel anaerobic, carboxydophilic, hydrogenogenic bacterium from a hot spring of the Lake Baikal area. *Int. J. Syst. Evol. Microbiol.* **55**, 2069-2073 (2005).
  12. Zavarzina, D. G. *et al.* *Thermincola ferriacetica* sp. nov., a new anaerobic, thermophilic, facultatively chemolithoautotrophic bacterium capable of dissimilatory Fe(III) reduction. *Extremophiles* **11**, 1-7 (2007).
  13. Logan, B. E. *et al.* Microbial electrolysis cells for high yield hydrogen gas production from organic matter. *Environ. Sci. Technol.* **42**:8630-8640 (2008).

SBIR

## ***Clostridium phytofermentans*: Genome Sequence of a Model System for the Direct Conversion of Plant Biomass to Fuels**

E. Petit,<sup>1</sup> T. Warnick,<sup>1</sup> W. Latouf,<sup>1,3</sup> A. Tolonen,<sup>4</sup> N. Ivanova,<sup>5</sup> A. Lykidis,<sup>5</sup> M. Land,<sup>6</sup> L. Hauser,<sup>6</sup> N. Kyrpides,<sup>5</sup> B. Henrissat,<sup>7</sup> G. Church,<sup>4</sup> D. Schnell,<sup>2</sup> S. Leschine,<sup>1</sup> and **J. Blanchard<sup>1\*</sup>** (blanchard@microbio.umass.edu)

<sup>1</sup>Dept. of Microbiology and <sup>2</sup>Dept. of Biochemistry and Molecular Biology, University of Massachusetts, Amherst, Mass.; <sup>3</sup>Qteros Inc., Amherst, Mass.; <sup>4</sup>Dept. of Genetics, Harvard Medical School, Boston, Mass.; <sup>5</sup>DOE Joint Genome Institute, Genome Biology Program, Production Genomics Facility, Walnut Creek, Calif.; <sup>6</sup>Oak Ridge National Laboratory, Life Sciences Division, Oak Ridge, Tenn.; and <sup>7</sup>Architecture et Fonction des Macromolécules Biologiques, CNRS, and Universités d'Aix-Marseille I and II, France

**Project Goals: Our long-term goal is to replace gasoline with fuels derived from ecologically and economically sustainable plant feedstocks. The objective of this project is to use an integrative systems level approach to characterize and manipulate the consortium of genes required to turn plant biomass into ethanol using *Clostridium phytofermentans*.**

*Clostridium phytofermentans* produces biofuels directly from a broad range of industrially relevant feedstocks. Thereby, alleviating the need for a separate industrial process that first breaks down the cellulose components into sugars before fermenting them to biofuels. The *C. phytofermentans* genome contains; (1) more and a greater diversity of glycoside hydrolases than any known biofuel-related bacteria; (2) a cornucopia of transport systems for assimilating solubilized

mono- and oligomeric carbohydrates; (3) no evidence of dockerin, cohesin and/or scaffoldin-like proteins or other cellulosome features; (4) three proteinaceous microcompartment loci involved in alcohol production, and; (5) seven putative hydrogenase that may code for novel signaling and metabolic pathways. The influx of biofuels-related genes via horizontal transfer from other microbes may explain why *C. phytofermentans* is capable of fermenting all major carbohydrate components of biomass. These and other attributes make *C. phytofermentans* an excellent model organism for understanding the direct conversion of plant biomass to biofuels.

SBIR

## **A Genome-Wide Perspective on the Regulation of Plant Carbohydrate Conversion to Biofuels in *Clostridium phytofermentans***

E. Petit,\* T. Warnick, S. Leschine, and **J. Blanchard** (blanchard@microbio.umass.edu)

Dept. of Microbiology, University of Massachusetts, Amherst, Mass.

**Project Goals: Our long-term goal is to replace gasoline with fuels derived from ecologically and economically sustainable plant feedstocks. The objective of this project is to use an integrative systems level approach to characterize and manipulate the consortium of genes required to turn plant biomass into ethanol using *Clostridium phytofermentans*.**

The economic costs of degrading cellulosic biomass currently hold back widespread use of plant biomass from agricultural and forestry wastes as a domestic renewable alternative to gasoline. *Clostridium phytofermentans* ferments all major component of the plant cell wall to ethanol and is an emerging model organism for understanding the direct conversion of plant biomass to fuels. The *C. phytofermentans* genome contains more carbohydrate degradation enzymes, more carbohydrate transporters and more transcriptional regulatory factors than any other biofuels-related microorganism. In order to build a foundation for engineering improvements on industrially relevant feedstocks, microarray experiments were carried out on a number of purified plant cell wall carbohydrates. The results demonstrate that *C. phytofermentans* regulates the stoichiometry of the plant degradative and assimilatory machinery in response to substrate availability. Because of the modularity of the sugar degradation and assimilation systems and the lack of a cellulosome, further improvement in biomass degradation through genetic manipulation is relatively straightforward.