DOE/SC-0140

Sections:
Systems Biology for DOE Energy Missions: Bioenergy
Bioenergy Research Centers
BioEnergy Science Center (BESC)
Great Lakes Bioenergy Research Center (GLBRC)
Joint BioEnergy Institute (JBEI)

U.S. DEPARTMENT OF ENERGY
Office of Science

Joint Meeting 2011

Genomic Science Awardee Meeting IX

and

USDA-DOE Plant Feedstock Genomics for Bioenergy Awardee Meeting

[Revised: April 14, 2011]

Crystal City, Virginia
April 10-13, 2011

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

http://genomicscience.energy.gov

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

Systems Biology for DOE Energy Missions: Bioenergy

Bioenergy Research Centers

1

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

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

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

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

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

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

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

BioEnergy Science Center (BESC)

2

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

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

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

http://bioenergycenter.org

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

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

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

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

3

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

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

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

http://bioenergycenter.org

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

Background

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

Methods

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

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

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

Results

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

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

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

4

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

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

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

http://bioenergycenter.org

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

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

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

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

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

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

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

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

5

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

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

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

http://bioenergycenter.org

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

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

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

6

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

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

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

http://bioenergycenter.org

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

The thick recalcitrant secondary cell walls of woody tissues are principally composed of cellulose, glucuronoxylan (GX), and lignin. The GX present in hardwoods including poplar and in mature stems of the model plant *Arabidopsis thaliana* have a backbone composed of 1,4-linked β-D-xylosyl (Xyl) residues that are often substituted at *O*-2 with glucuronic acid (GlcA) or 4-*O*-methyl glucuronic acid (MeGlcA). *Arabidopsis* GX has approximately one uronic acid residue for every eight xylosyl residues and a GlcA:MeGlcA ratio of 1 to 3. We are using *Arabidopsis* as a model system to identify genes involved in xylan biosynthesis and modification, as the identities and functions of many of the enzymes involved in this process are poorly understood. The information gained with *Arabidopsis* and with model monocot species will provide a solid basis for decreasing recalcitrance by targeted modification of wall composition and architecture in biofuel crops including poplar and switchgrass.

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

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

7

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

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

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

http://bioenergycenter.org

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

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

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

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

Reference

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

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

8

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

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

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

http://bioenergycenter.org

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

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

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

9

Understanding Cellulose Structure

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

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

http://bioenergycenter.org

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

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

This work was supported by DOE BER BioEnergy Science Center (cellulose structure by NMR) and by the DOE ASCR SciDAC program (cellulose modeling)
The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

10

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

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

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

http://bioenergycenter.org

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

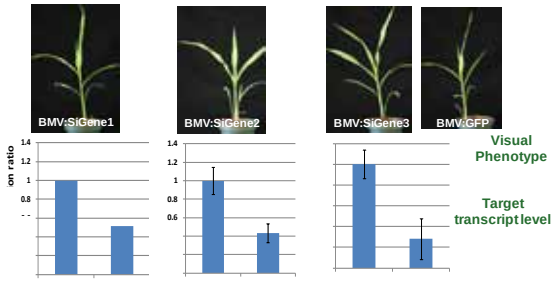
Number of gene fragments cloned into VIGS vector	51
Number of genes analyzed by VIGS	37
Number of genes showing silencing	18
Number of genes analyzed for sugar release after VIGS study	4
Number of remaining genes with NREL for sugar release analysis	14

Fig. 1: BESC VIGS progress to date

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

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

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



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

Sugar release and lignin S/G values in silenced plants

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

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

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

11

Strategy to Identify Plant Genes that Affect Biomass Recalcitrance

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

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

http://bioenergycenter.org

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

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

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

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

12

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

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

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

http://bioenergycenter.org

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

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

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

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

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

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

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

13

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

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

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

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

http://bioenergycenter.org

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

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

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

14

Development of Resources for Switchgrass Functional Genomics

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

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

http://bioenergycenter.org

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

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

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

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

15

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

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

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

[a]C. Fu and J. Mielenz contributed equally to this work.

http://bioenergycenter.org

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

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

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

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

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

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

16

Integrated Computational Biology Capabilities at BESC

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

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

http://bioenergycenter.org

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

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

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

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

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

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

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

17

High-Throughput Compositional Analysis of Lignocellulosic Biomass

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

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

http://bioenergycenter.org

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

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

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

18

BioEnergy Science Center Education and Outreach

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

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

http://bioenergycenter.org

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

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

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

19

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

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

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

http://bioenergycenter.org

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

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

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

20

Genetic Tool Development and Deployment in *Clostridium thermocellum*

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

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

http://bioenergycenter.org

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

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

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

21

Reconstruction of Transcription and Metabolic Networks in *Clostridium thermocellum*

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

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

http://bioenergycenter.org

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

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

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

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

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

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

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

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

22

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

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

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

http://bioenergycenter.org

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

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

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

23

Understanding Cellulosomal Enzymes: Experiment and Modeling

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

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

http://bioenergycenter.org

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

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

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

24

Progress in Quantitative Assessment and Interpretation of Cellulosome Activity

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

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

http://bioenergycenter.org

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

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

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

25
Cellulosome Self Assembly

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

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

http://bioenergycenter.org

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

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

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

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

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

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

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

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

http://bioenergycenter.org

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

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

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

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

Great Lakes Bioenergy Research Center (GLBRC)

27

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

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

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

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

http://www.glbrc.org/

Project Goals:

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

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

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

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

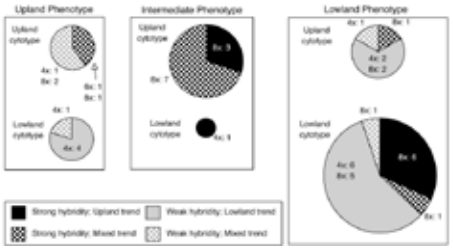


Figure 1. Pie diagrams showing the relationships among phenotype, cytotype (cpDNA sequence), and nuclear SSR marker profiles for 46 switchgrass plants that were identified as having marginal to high probabilities of hybrid descent (pH > 0.05) and presented in Tables 3 and 4. Each pie chart refers to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

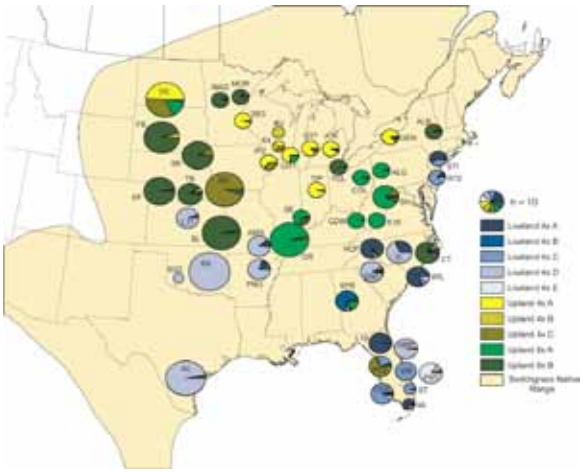


Figure 2. Partial map of the USA, showing the approximate location of each switchgrass cultivar or accession and the probabilities of accession membership in one of 10 groups, identified by STRUCTURE analysis. The size of each circle represents the sample size for each accession, with n = 10 shown in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

28

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

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

Great Lakes Bioenergy Research Center, University of Wisconsin

http://www.glbrc.org/

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

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

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

29

Systems Biology of *Cellvibrio japonicus* Lignocellulose Degradation

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

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

http://www.glbrc.org/

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

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

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

30

High Throughput Production and Characterization of Cellulytic and Hemicellulytic Enzymes

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

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

http://www.glbrc.org/

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

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

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

31

Improved Enzyme Cocktails for Biomass Conversion

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

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

http://www.glbrc.org/

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

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

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

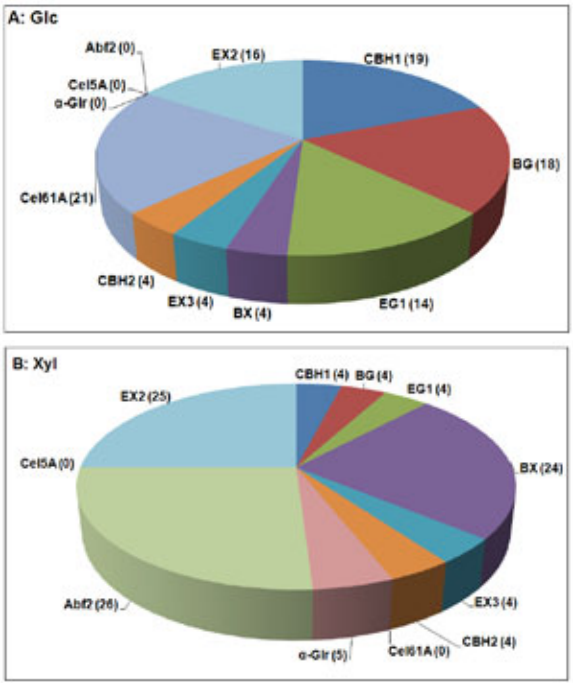


Figure 1. (A): Optimal proportions of 11 enzymes for release of Glc from alkaline peroxide-pretreated corn stover determined using GENPLAT. (B): Optimal proportions of the same enzymes for release of Xyl. Yields of Glc and Xyl were 85% and 65%, respectively, at an enzyme loading of 15 mg/g glucan and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

References

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

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

32

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

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

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

http://www.glbrc.org/

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

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

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

33

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

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

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

http://www.glbrc.org/

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

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

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

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

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

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

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

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

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

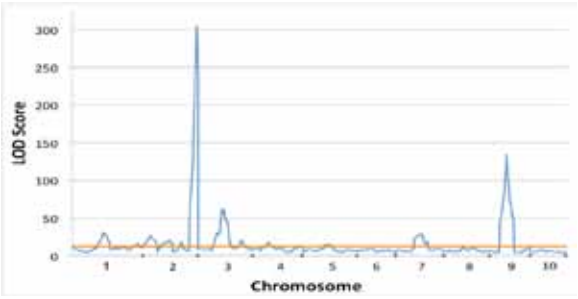


Figure 1. Joint quantitative trait loci (QTL) analysis of transition from juvenile to adult tissue on the Nested Association Mapping (NAM) population evaluated in 2008 and 2009 at the Arlington, WI Agricultural Research station.

34

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

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

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

http://www.glbrc.org/

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

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

35

Restructuring Crystalline Cellulose Hydrogen Bond Network Enhances its Depolymerization Rate

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

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

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

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

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

Student Oral Presentation–Monday

36

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

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

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

http://www.glbrc.org/

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

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

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

37

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

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

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

http://www.glbrc.org/

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

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

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

38

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

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

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

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

Project goals:

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

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

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

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

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

39

Improving the Sustainability of Bioenergy Crops Through Arbuscular Mycorrhization

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

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

http://www.glbrc.org/

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

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

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

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

40

A Lattice Monte Carlo Model of Substrate Channeling on Catalytic Scaffolds

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

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

http://www.glbrc.org/

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

41

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

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

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

http://www.glbrc.org/

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

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

Results

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

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

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

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

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

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

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

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

42

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

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

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

http://www.glbrc.org/

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

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

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

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

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

43

Biomass Trait Analyses of Diverse Wild Type *Brachypodium* Accessions

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

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

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

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

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

272‡

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

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

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

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

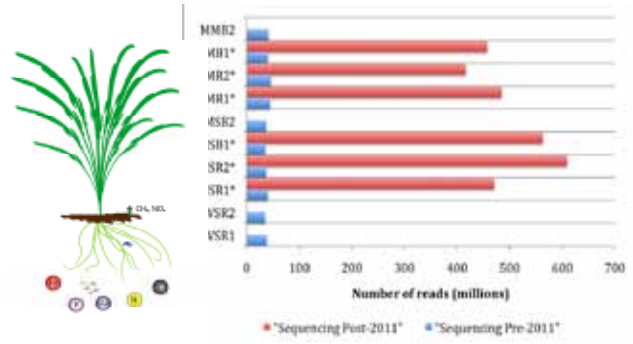


Figure 1a (left). Conceptual model of plant-microbe interactions related to bioenergy sustainability efforts. We target genes mostly related to phosphorus, nitrogen, and iron cycling, as well as plant growth promoting genes. Figure 1b (right). Number of metagenomic short reads sequenced from switchgrass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

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

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

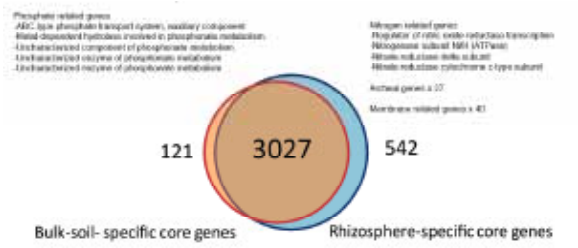


Figure 3. Comparison of COGs identified from 454 short reads from bulk and rhizosphere soil communities.

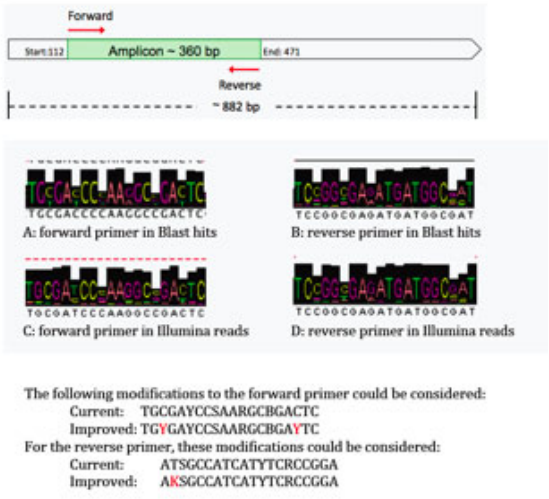


Figure 4. Primers used in 454 Titanium sequencing of *nifH* gene, using *Trichodesmium thiebautii* as an example. Consensus sequences of the primer region from top Blast NifH hits and Illumina reads suggest improvements to current primers.

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

Reference:

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

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

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

Joint BioEnergy Institute (JBEI)

44

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

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

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

http://www.jbei.org/

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

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

45

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

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

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

http://www.jbei.org/

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

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

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

46

Identification and Characterization of Monolignol Transporters in *Arabidopsis*

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

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

http://www.jbei.org

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

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

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

47

Isolation and Proteomic Characterization of the *Arabidopsis* Golgi Apparatus

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

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

http://www.jbei.org

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

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

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

48

Feedstock-Adapted Anaerobic Consortia Derived from Tropical Forest Soils

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

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

http://www.jbei.org

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

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

49

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

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

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

http://www.jbei.org/

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

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

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

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

50

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

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

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

http://www.jbei.org/

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

Keyword: Metagenomics; lignocellulosic biomass; biofuels

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

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

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

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

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

51

Strategies for Improving Production and Resistance Phenotypes in Engineered Microbes

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

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

http://www.jbei.org/

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

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

52

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

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

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

http://www.jbei.org/

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

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

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

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

53

Microbial Production of a Terpene Based Advanced Biodiesel

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

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

http://www.jbei.org/

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

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

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

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

54

Re-Engineering Secondary Cell Wall Deposition in *Arabidopsis*

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

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

http://www.jbei.org

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

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

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

55

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

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

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

http://www.jbei.org

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

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

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

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

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

References

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

56

Thermophilic Cellulase Cocktail for the Saccharification of Ionic Liquid Pretreated Biomass

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

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

http://www.jbei.org

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

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

57

Discovering and Engineering Ionic Liquid Tolerant Cellulases

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

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

http://www.jbei.org

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

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

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

58

Lignocellulosic Biomass Degradation via Targeted Glycoside Hydrolases

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

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

http://www.jbei.org/

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

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

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

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

59

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

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

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

http://www.jbei.org/

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

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

60

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

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

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

http://www.jbei.org

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

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

61

Targeted Proteomics for Metabolic Pathway Optimization: Application to Terpene Production

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

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

http://www.jbei.org

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

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

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

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

62

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

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

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

http://www.jbei.org/

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

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

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

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

63

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

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

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

http://www.jbei.org/

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

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

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

64

Encoding Substrates with Mass Tags to Resolve Stereospecific Reactions

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

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

http://www.jbei.org/

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

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

65

Microfluidic Platform for Synthetic Biology Applications

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

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

http://www.jbei.org

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

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

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

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

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

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

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

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

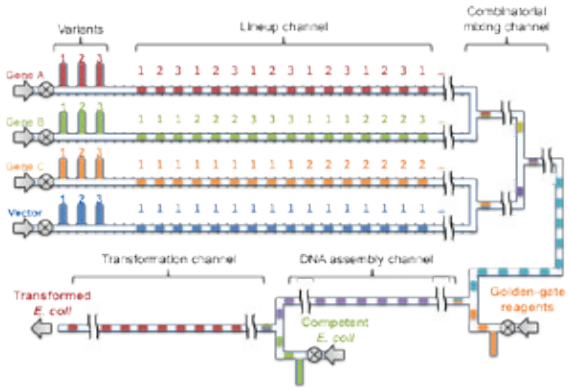


Figure 1. Schematic of microfluidic platform for synthetic biology applications.

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

66

The Joint BioEnergy Institute (JBEI) Computational Biology Core

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

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

http://www.jbei.org/

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

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

67

Electronic Laboratory Notebook Usage at JBEI

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

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

http://www.jbei.org/

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

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

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

68

DNA Assembly Design with j5 and DeviceEditor Biocad Tools

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

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

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

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

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

Biofuels: Analytical and Imaging Technologies, Engineering, and Production

69

Engineering Bacterial Surfaces to Display Cellulosomes for Biofuel Production

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

University of California, Los Angeles

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

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