DOE/SC-0140

Section:
Joint USDA-DOE Plant Feedstock Genomics for Bioenergy

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

Joint USDA-DOE Plant Feedstock Genomics for Bioenergy

119

Systems View of Root Hair Response to Abiotic Stress

Principal Investigator: **Gary Stacey[1]*** (staceyg@missouri.edu), **Co-Principal Investigators:** Jianlin Cheng,[2] Dong Xu,[2] David Koppenaal,[3] and Ljiljana Paša-Tolić[3]

[1]Center for Sustainable Energy, [2]University of Missouri, Columbia; [3]EMSL, Pacific Northwest National Laboratory, Richland, Wash.

http://staceylab.missouri.edu/
http://soyroothair.org/

Project Goals:

1. **Analyze the transcriptional response of root hair cells under conditions of abiotic stress (i.e.,heat and drought stress)**
2. **Analyze proteomic and metabolomic changes in root hair cells under conditions of abiotic stress (i.e., heat and drought stress)**
3. **Utilize the tools and information developed in Objectives 1-2 to develop descriptive models of root hair metabolic function.**

Over the last 100 years, the atmospheric concentration of carbon dioxide has dramatically increased, in major part due to the burning of fossil fuels, recent rapid industrialization, and land use changes. The predicted effects of continued climate change are complex but include effects on air and surface temperature, with coincident effects on water availability. Soil temperature can influence root growth, cell elongation, root length and extension, initiation of new lateral roots and root hairs, and root branching. These effects are likely manifestations of the variety of physiological effects brought about by temperature on plant roots; including changes in root respiration, nutrient uptake, as well as physicochemical effects on the soil environment (e.g., changes in nitrogen mineralization). Ambient temperature changes also affects other parts of the plant (e.g., photosynthetic rates), which also affects below ground growth and physiology. When we include in this discussion issues of plant genetic variation, as well as the effects of temperature on water availability, the full complexity of the effects of climate change on the plant root environment becomes clear.

In order to properly understand the effects of climate change, models must be developed that predict impacts across broad temporal (seconds to millennia) and spatial (microns to global) scales. In order to be useful, these models must draw upon accurate experimental data. Systems biology seeks to address these needs by providing a comprehensive, quantitative analysis of the manner in which all the components of a biological system interact functionally over time and space. The recent explosion of interest in systems biology is the result of the development of new tools for system-level analysis of cellular function and the availability of an increasing number of full genome sequences, which enables the full application of these new technologies. The ultimate goal is a new, predictive view of biological function, supplanting the older descriptive understanding. Hence, there is a need to integrate system approaches to understand the effects of climate change on molecules, cells, organisms and ecosystems.

However, the promise of this new 'predictive' science has yet to achieve its full potential. A number of challenges remain. For example, although the new tools do indeed provide for a full systems view of cellular function, integration of dissimilar data (e.g., proteomics, metabolomics, transcriptomics, etc.) remains a formidable challenge. Among the issues compounding the problems of data integration is the issue of "signal dilution", which results from the fact that most studies average the response of whole tissues, obscuring the actual cellular response. Hence, it is impossible to discern the difference, for example, of a gene that is expressed at a low level in all cells from a gene that is expressed at a very high level, but only in a few cells. Approaches are needed to conduct functional genomics on single cells.

This proposal will address the question of signal dilution by focusing, specifically, on soybean root hair cells, which represent a single, differentiated cell type. Over the past 5 years, we have established the soybean root hair cell as an excellent platform for plant systems biology studies. It is now arguably the best characterized cell type in plant biology, as exemplified by our various publications, databases and additional information yet to be published (see Libault et al., 2010).

Our vision is to utilize the soybean root hair system to explore, at a systems level, the biology of a single, differentiated plant cell type, while gaining novel insight into the impacts of temperature and water availability on a crucial root cell necessary for nutrient uptake. The proposed research should provide unambiguous measurements of the impact of these environmental factors on plant cell function, without the compounding effects of tissue dilution. The proposed research will focus on defining the transcriptional, metabolomic and proteomic response of the soybean root hair cell to variations in temperature and water availability. These data, in addition to other data available in our laboratory, will allow the development of computational models to examine regulatory networks that function at a single cell level to control the response to environmental change. The data obtained should provide a better understanding of the impacts of climate change (heat and water limitation) on plant root physiology.

The research team brings a wealth of experience and knowledge to the project, which is a collaboration between scientists at the University of Missouri and at the Environmental Molecular Sciences Laboratory at the DOE Pacific Northwest National Laboratory. The latter facility brings tremendous experience and instrumentation to conduct the metabolomic and proteomic studies described. In addition to the expected research outcomes, the project will also provide training for graduate and postdoctoral students to prepare them for their future careers.

Reference

1. Libault, Marc, Brechenmacher, Laurent, Cheng, Jianlin, Xu, Dong, Stacey, Gary (2010) Root hair systems biology. *Trends in Plant Science* 15: 641-650.

120

Night-Time Stomatal Conductance and Transpiration Negatively Impact Biomass Accumulation

C.E. Jahn,[1] L. Derose-Wilson,[1] **J.K. McKay**,[1] **D.R. Bush**,[1] **H. Leung**,[2] and **J.E. Leach**[1]* (Jan.Leach@colostate.edu)

[1]Program in Plant Molecular Biology, Colorado State University, Fort Collins; and [2]Plant Breeding, Genetics and Biotechnology Division, International Rice Research Institute (IRRI), Metro Manila, Philippines

Project Goals: Our goal is to provide the applied biomass research community and industry with information to allow exploitation of the genes and pathways relevant to biomass accumulation in grasses. The specific objectives for this project are to: Objective 1: Identify genes involved in biomass accumulation. Using genetic populations for rice lines that exhibit extremes in biomass accumulation, we will (a) map QTL using a comprehensive phenotyping/genotyping approach, (b) confirm the QTL location and phenotypic effect and (c) screen a deletion mutant collection to identify large deleted regions corresponding to biomass accumulation. Objective 2: Dissect the QTL using an integrated analysis. First, we will use an expedited approach to generating near isogenic lines containing each QTL. Second, we will integrate genome-wide data to refine the QTL regions including (a) association mapping, (b) expression profiling, (c) mutant analysis, (d) fine scale mapping, and (e) comprehensive sequencing.

Plant biomass accumulation is the culmination of many processes: carbon assimilation via photosynthesis, carbon losses due to respiration, efficiency of biosynthetic pathways, long-distance transport of assimilates, water balance, and mineral nutrition. Previously, we identified a negative correlation between total biomass and photosynthetic rates in diverse rice varieties (Jahn et al 2011). Although counterintuitive, this result is not unique to rice; strong negative correlations have been observed between photosynthetic rates and biomass accumulation in other crop and weed species. However, to date, there is no mechanistic explanation of this phenomenon. We took detailed physiological measures on a well replicated growth chamber experiment of 20 diverse rice varieties known to differ in biomass. We measured gas exchange in both day and night, chlorophyll fluorescence in light- and dark-adapted plants, and chlorophyll content. These data revealed significant genetic variation among the 20 lines in all traits measured. Trait variation was largely explained by genotype and breeding history (advanced vs landrace) but not by varietal groupings (indica, japonica, aus). Total biomass was negatively correlated with night-time stomatal conductance, transpiration and dark respiration. The 20 lines varied as much as 5.7-fold for night-time stomatal conductance. Among varieties, night-time transpiration rates were between 4 and 26% that of day-time and thus represent a substantial fraction of total daily water loss for some lines. Although little effort to date has been placed on directly improving photosynthesis, new initiatives to improve photosynthetic rates are considered a frontier for grain and biomass yields. However, our work indicates that simply increasing day-time carbon assimilation rates may not be sufficient. A more comprehensive approach will also target carbon metabolism, respiration and water losses due to night-time transpiration.

121

The Hunt for Green Every April: Phenotypic and Metabolomic Analysis of Nutrient Remobilization in Switchgrass

Gautam Sarath[1]* (Gautam.Sarath@ars.usda.gov), Nathan A. Palmer,[1] Aaron J. Saathoff,[1] Christian M. Tobias,[2] Soundararajan Madhavan,[3] Paul Twigg,[4] Kenneth P. Vogel,[1] and Jaehyoung Kim[5]

[1]GFBRU, USDA-ARS, East Campus, UNL, Lincoln; [2]GGDRU, WRRC, USDA-ARS, Albany, Calif.; [3]Department of Biochemistry, University of Nebraska, Lincoln; [4]Biology Department, University of Nebraska, Kearney; and [5]Core for Applied Genomics and Ecology, University of Nebraska, Lincoln, East Campus

Project Goals: Our interrelated project objectives are: (A) elucidation of the regulation of gene networks, proteins and metabolites for manipulation of plant feedstocks for improved productivity and sustainability, and improved water use efficiency and nutrient utilization; (B) Elucidation of the regulation of gene networks, proteins and metabolites for advanced understanding of carbon partitioning and nutrient cycling in plant feedstocks; and (C) comparative approaches to enhance fundamental knowledge of the structure, function, and organization of plant genomes leading to innovative strategies for feedstock characterization, breeding or manipulation. Our transcriptomic and marker studies will add to existing genomic resources available for this species. The proposed C and N recycling studies in diverse switchgrass genotypes will significantly improve our knowledge in this area, and has direct bearing on plant fitness and sustainability issues of other perennial grasses being developed for use as bioenergy. Genomic and physiological insights

obtained during this project will be utilized to better develop molecular markers using a defined population of switchgrass plants that will be exploited to characterize patterns of nucleotide diversity reflected through the processes of mutation, recombination, genetic drift and-directional selection.

This DOE-USDA funded research seeks to understand traits controlling winter hardiness in switchgrass (*Panicum virgatum* L.). We have planted five different populations with contrasting winter survival in replicated field plots. During the 2010 growing season, one group of replicates was labeled with ^{13}C to follow the label in aerial tissues over the growing season. In other experiments, crowns and rhizomes were harvested at boot, anthesis, mid-seed fill, late senescence, post-frost and early green up (2011). These harvest dates were based on stages of development of the cold-adapted cultivar "Summer." Crowns and rhizomes will be extracted for 454 and Illumina sequencing, metabolite analyses and other related physiological studies. Other objectives of this grant are (1) to use next-generation sequencing to query transcript abundance (levels of gene expression) in specific populations of switchgrass plants during regreening and dormancy; and (2) to study the genetic variation (extent of linkage disequilibrium in populations) using over 2000 plants from various genetic backgrounds that have been planted in the field for these analyses. In order to build a bioinformatic pipeline, we have recently acquired and analyzed ~1 million sequences obtained from pre-frost Summer crowns using the 454 platform. Data from this and related experiments will be presented.

DOE-USDA funded research (2010-2012; Office of Science (BER), U. S. Department of Energy grant number DE-AI02-09ER64829)

122

Functional Analysis of Regulatory Networks Linking Shoot Maturation, Stem Carbon Partitioning, and Nutrient Utilization in Sorghum

Stephen Moose[1]* (University of Illinois), Pat Brown,[1] and Max Moehs[2]

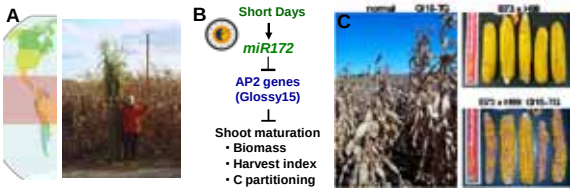
[1]University of Illinois and [2]Arcadia Biosciences

http://cropsci.illinois.edu/faculty/moose/lab/energy.html

Project Goals: The phenotypes conditioned by the Gl15-OX transgene in maize suggest simple molecular breeding strategies to modulate shoot maturation and enhance sustainable biomass yields of bioenergy grasses. The goals of this project are to increase our understanding of shoot maturation pathway genes in sorghum and test their utility in improving sorghum as a bioenergy feedstock. We are employing three complementary experimental approaches:

1. **Characterize allelic and expression variation for *miR172*, *AP2*, *SPL* and *miR156* genes within a core set of 12 diverse sorghum genotypes representing grain, sweet, and forage types.**
2. **Allelic variation in shoot maturation pathway genes will be tested for associations with phenotypic variation for flowering time, biomass yields, photoperiod-sensitivity, and stem sugar production in a panel of 500 diverse sorghum genotypes.**
3. **Transgenic sorghum lines with altered expression of *Glossy15* or *miR172* will be generated and evaluated for beneficial changes in flowering time, biomass yields, and nitrogen utilization. These transgenic lines will also be used to identify changes in the expression of down-stream target genes via Illumina RNA tag profiling.**

Phenotypic traits that are important to increasing sustainable biomass production and conversion efficiency from bioenergy crops include growth habit (perennial versus annual), flowering time, vegetative senescence, tillering, carbon partitioning and cell wall composition, and nutrient use efficiency. Each of these traits are impacted by the developmental process of shoot maturation or phase change, where shoot meristems and lateral organs progress through embryonic, vegetative, and reproductive growth. Studies conducted in both *Arabidopsis* and maize have identified a conserved regulatory network of two antagonistically acting microRNAs (*miR156* and *miR172*) and their target transcription factors (SPL and AP2 proteins) that control shoot maturation and phase change. Mutations and transgenic lines that alter the relative activities of *miR156*, *miR172*, *SPL* and *AP2* genes condition changesin growth habit, tillering, cell wall composition, response to photoperiod, flowering time, fertility, and seed size.



D	Grain hybrids		Tropical hybrids		Gl15-OX hybrids	
Yields (Mg/ha)	-N	+N	-N	+N	-N	+N
Total biomass	14.8	18.9	20.0	22.0	18.0	20.5
Grain	7.5	10.8	1.3	2.6	2.5	3.3
Stover	7.3	8.1	18.7	19.4	15.5	17.2
Sugar	0.7	0.2	3.0	3.6	2.7	3.0

Figure 1. Phenotypic effects of *Glossy15* overexpression in transgenic maize. (A) Photograph of tropical-adapted maize hybrid grown in temperate environment (foreground), taller and still green relative to commercial grain hybrid at physiological maturity. (B) Simplified model for photoperiodic regulation of shoot maturation, where short days increase miR172 activity, which downregulates AP2 genes that suppress phase change. (C) Enhanced plant height and reduced seed set in Gl15-OX lines relative to control. (D) Mean yields of total biomass, grain, stover, and sugar pressed from stalk sap from plants harvested at 145 days after sowing. Data from 2008 trial in N-responsive field plots at Champaign, Illinois, and includes 5 elite grain hybrids, 3 high biomass tropical hybrids, and 5 Gl15-OX hybrids where *Gl15* transgene was introgressed into same inbred parents of the 5 elite grain hybrids.

The Moose laboratory has developed and characterized transgenic maize hybrids that overexpress the maize

Glossy15 gene, which encodes an AP2 protein required for vegetative phase change. The *Glossy15*-overexpression (Gl15-OX) lines delay shoot maturation, leading to prolonged vegetative development, later flowering and vegetative senescence, reduced seed number, and the accumulation of greater amounts of total biomass and sugar in stem tissues relative to current commercial grain hybrids, (Figure 1). It is important to note that when grown without supplemental N fertilizer in field plots depleted for residual N, total biomass yields of Gl15-OX lines are nearly equal to those obtained for the elite grain hybrids when provided N fertilizer. These phenotypic effects mimic those observed when photoperiod-sensitive tropical maize germplasm is grown in temperate environments, in genotypes that are already well-adapted to temperate environments.

123

Role of Histone Modifications in the Regulation of Cell Wall Synthesis in Rice (*Oryza sativa*)

Feng Tan,[1] Hui Wang,[1] Guo-Liang Wang,[2] Pamela Ronald,[3] and **Zhaohua Peng**[1*]

[1]Department of Biochemistry and Molecular Biology, Mississippi State University; [2]Department of Plant Pathology, the Ohio State University; and [3]Department of Plant Pathology, University of California-Davis

Project Goals: Investigate the regulatory mechanism of cell wall synthesis and cell wall characteristics

Cell wall synthesis is subjected to precise temporal and spatial regulation. The multiple components, including cellulose, hemicellulose, lignin, pectin, proteins, etc, have to be synthesized and deposited coordinately in the cell wall. Therefore, it is conceivable that some global regulators must be presented in the cell to coordinate different pathways of cell wall synthesis. The goal of our project is to investigate the regulatory mechanism of cell wall synthesis and cell wall characteristics. In rice protoplasts, we find that cell wall removal and regeneration is associated with substantial chromatin reorganization and global histone modification changes. ChIP-Seq studies reveal that substantially more cell wall metabolic pathway genes subjected to the regulation of histone modifications than the average of the genome. Preliminary studies further show that mutations in some key histone modification genes lead to cell wall content change in rice. Our results suggest a critical role of histone modifications in the regulation of cell wall synthesis and the characteristics of cell wall.

124

Characterization of Nitrogen Use Efficiency in Sweet Sorghum

Ismail Dweikat[1*] (idweikat2@unl.edu), Tom Clements,[2] Don Weeks,[3] and Jianming Yu[4]

[1]Department of Agronomy and Horticulture, University of Nebraska-Lincoln; [2]Center for Biotechnology, Plant Science Initiative, and Department of Agronomy and Horticulture, University of Nebraska-Lincoln; [3]Dept. of Biochemistry, University of Nebraska-Lincoln; and [4]Dept. of Agronomy, Kansas State University, Manhattan

Project Goals: Objective 1: Conduct quantitative trait loci (QTLs) analysis and marker identification for nitrogen use efficiency (NUE) in advanced grain sorghum populations. Objective 2: Identify loci and specific alleles that control NUE using whole genome and candidate gene association mapping techniques across a diverse set of grain and sweet sorghum accessions.

Originated in the dry semi-arid region in Africa, sorghum has developed adaptive traits for stress environments and there is wide genetic variability for those traits including tolerance to low nitrogen supply and efficiency to utilize water and nutrients. These present opportunities for further enhancing stress tolerance and adaptability of the crop through identification of plant characteristics associated with the traits and exploiting them in breeding programs. Nitrogen is one of the essential mineral nutrients required for production of grain and biomass crops. The cost of chemical fertilizer in recent years has increased dramatically and significant amount of fertilizer nitrogen applied are lost in various ways. In sorghum it is estimated that less than fifty percent of nitrogen fertilizer added to the soil is taken up by the plant and converted to grain or biomass. This not only indicates the amount of dollars lost but also shows the yield benefit that would have been obtained if the fertilizer was properly utilized by the plant to produce grain/biomass. It is very important that this loss is reduced through developing nitrogen efficient genotypes. The objectives of this study are to identify quantitative trait loci associated with nitrogen use efficiency (NUE) using two developed recombinant inbred line population developed between a high N-efficient parents (Chin 17 and San Chi San) and the normal genotypes CK60, and to identify loci and specific alleles that control NUE using whole genome and candidate gene association mapping techniques across a diverse set of grain and sweet sorghum accessions. A total of 236 simple sequence repeat (SSR) markers have been mapped on the two populations with an average of 10 cM distance between markers. In addition to the SSR markers, 1530 single nucleotide polymorphisms (SNPs) are being mapped. A nested association mapping (NAM) population is already under development where 100 recombinant lines each were generated of BTx 623 inter-mated with 50 parental lines exhibiting wide variation high biomass producing, early germination and seedling cold tolerant, high NUE, stalk rot resistance, non-flowering, lodging resistance, high

tillering, and stay-green phenotypes. The SNPs and about 2000 SSRs will be used to screen the NAM populations to identify genomic that show association with NUE.

125

Phenomic Analysis of Natural and Induced Variation in *Brachypodium distachyon*

John Vogel*[1] (john.vogel@ars.usda.gov), Jennifer Bragg,[1] Richard Poiré,[2] Xavier Sirault,[2] Ludmila Tyler,[1,3] Vincent Chochois,[4] Michelle Watt,[4] and Robert Furbank[2]

[1]USDA-ARS, Western Regional Research Center, Albany, Calif.; [2]High Resolution Plant Phenomics Centre, CSIRO Plant Industry, Canberra, Australia; [3]University of California, Berkeley; and [4]CSIRO Plant Industry, Canberra, Australia

Project Goals:

1. **Assemble a collection of natural accessions and 2,000 homozygous T-DNA lines.**
2. **Conduct a detailed phenotypic characterization of the collection using a phenomic approach.**
3. **Begin detailed characterization of a select group of mutants and natural accessions.**

Herbaceous energy crops, especially grasses, are poised to become a major source of energy in the United States. Despite their increasing importance, we know little about the basic biology underlying the traits that control the utility of grasses as energy crops. Better knowledge of basic grass biology (e.g. identification of the genes that control cell wall composition, plant architecture, cell size, cell division, reproduction, nutrient uptake, carbon flux, etc.) could be used to design rational strategies for crop improvement and shorten the time required to domesticate these new crops. The model grass *Brachypodium distachyon* (*Brachypodium*) is an ideal system with which to acquire this knowledge. We are conducting high-throughput phenotypic analysis (phenomics) of homozygous T-DNA mutants and natural accessions of the model grass *Brachypodium distachyon* to accelerate the acquisition of this knowledge.

Accurate phenotypic characterization of large numbers of individuals under carefully controlled conditions is a costly and rate limiting step for functional genomic projects. These analyses typically involve destructive measurements, multiple replicate sets of plants and exhaustive manual labor. An additional problem is that measurements made at different times on different sets of plants are not directly comparable due to small variations in environmental conditions. The 'phenomic' approach to these problems is to perform multiple non-destructive phenotypic measurements in an automated high-throughput fashion on a large number of plants at one time and, where possible, to make repeated observations over time to provide more robust data. This approach achieves cost savings due to economies of scale and more reliable data due to standardization of environmental conditions, repeated observations and automation.

We are conducting high-throughput phenotypic analysis of homozygous *Brachypodium* T-DNA mutants and natural accessions. We plan to phenotype 2,000 homozygous T-DNA mutants and >100 inbred lines under defined environmental conditions at the state-of-the-art High Resolution Plant Phenomics Centre, part of the new Australian Plant Phenomics Facility. We have completed pilot experiments designed to identify the optimal growth conditions (light level, photoperiod, light quality, nutrient level, watering and soil type) prior to large-scale phenotyping. In addition, we have completed preliminary experiments to establish methods to examine root architecture. With our conditions and protocols optimized, we have initiated the first full set of experiments to phenotype a diverse collection of 100 natural accessions. These accessions were selected from a larger collection of over 1,000 accessions.

Concurrent with the initial phenotyping of the natural accessions, we are using a PCR-based approach to identify homozygous T-DNA lines containing insertions predicted to disrupt genes. We have optimized and validated our methods and have scaled up our genotyping efforts. To date, we have identified over 200 homozygous T-DNA lines and plan to have sufficient seed to phenotype this initial set beginning in 8 weeks. We anticipate that we will identify the full set of 2,000 homozygous lines within the coming year.

Student Oral Presentation–Monday

126

Beneficial Bacterial Endophyte *Burkholderia phytofirmans* Strain PsJN Significantly Promotes Switchgrass Alamo Growth

Scott Lowman,[1]* Seonhwa Kim,[1] Alejandra Lara-Chavez,[1] Guichuan Hou,[4] Barry Flinn,[1] John Seiler,[3] Jerzy Nowak,[2] and **Chuansheng Mei**[1]

[1]Institute for Sustainable and Renewable Resource, Institute for Advanced Learning and Research, Danville, Va.; [2]Department of Horticulture, [3]Department of Forest Resources and Environmental Conservation, Virginia Polytechnic Institute and State University, Blacksburg; and [4]The Dewel Microscopy Facility in the College of Arts and Sciences, Appalachian State University, Boone, N.C.

http://www.isrr.ialr.org/index.php/faculty/dr-chuansheng-mei

Switchgrass is one of the most promising bionergy crop candidates for the U.S. It gives relatively high biomass yields and can grow on marginal lands. However, the biomass yield varies from year to year and from location to location. Our goal is to develop a low input and sustainable switchgrass feedstock production system utilizing beneficial bacterial endophytes. Beneficial microbial endophytes, generally, promote plant growth, increase nutrient uptake, enhance host tolerance to environmental stresses, and inhibit the growth of plant pathogens and associated diseases. We have demonstrated that one plant growth-promoting bacterial endophyte, *Burkholderia phytofirmans* strain PsJN, is able to colonize and significantly promote the growth of

switchgrass cv. Alamo under *in vitro*, growth chamber, and greenhouse conditions. Using the strain PsJN containing a GFP tag, we were able to visualize bacterial cells inside roots under confocal microscope three days after inoculation (Fig. 1).

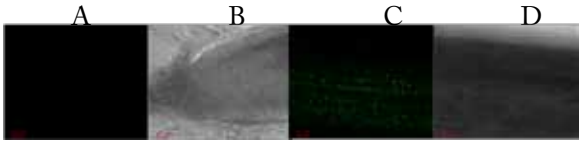


Figure 1. Root confocal images were taken 3 days after inoculation of imbibed switchgrass cv. Alamo seeds with PsJN-GFP (40X magnification). A and B are images of non-infected control root; C and D, images of the PsJN-infected root; A and C were taken under fluorescence; and B and D, under visible light.

In six independent *in vitro* experiments conducted on cv. Alamo, 10-day seedlings derived from seeds inoculated with the strain PsJN had on average 56.5% higher fresh weight than controls derived from seeds inoculated with buffer (PBS) alone. Figure 2 illustrates PsJN growth promotion of *in vitro* Alamo seedlings one month after inoculation.

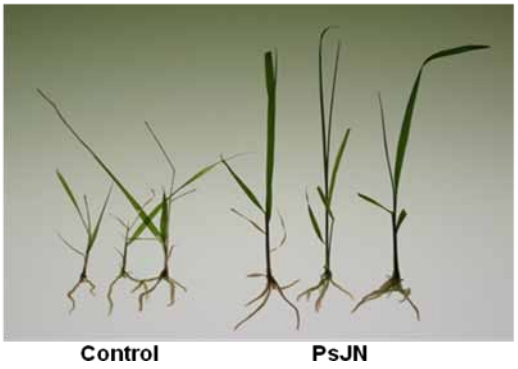


Figure 2. PsJN promoted Alamo growth.

When the one-month-old seedlings were transferred from tissue culture to cavity trays filled with 2/3 Micro Grow and 1/3 *Arabidopsis* soil mix and grown in a growth chamber (28/22˚C day/night temperature, 16-hour photoperiod) for 30 days, the PsJN-inoculated Alamo plants had significantly higher shoot and root biomass than the controls ($p<0.01$). The total dry weight averaged from five experiments was 54.1% higher in the inoculated treatment compared to non-inoculated control. Figure 3 shows one representative experiment indicating that fresh weight and dry weight in Alamo inoculated with the strain PsJN were significantly increased compared with non-inoculated control plants, with a p-value of 7.57E-08 for total dry weight.

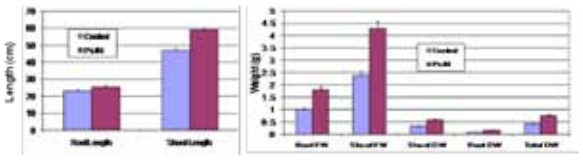


Figure 3. Strain PsJN promoted Alamo shoot and root growth in growth chamber conditions.

Similar results were obtained in greenhouse experiments with transplants grown in 6-gallon pots for two months. The total dry weight of the inoculated plants was 48.6% higher than controls. The inoculated plants formed more and earlier tillers than controls. When the PsJN-inoculated Alamo seedlings were grown *in vitro* for 25 days and transferred to 6-gallon pots containing the field soil with no fertilizer application and grown for 3 months in glasshouse under ambient conditions, they produced twice higher total dry weight than controls. The data indicate the potential benefit of switchgrass seed inoculation with PsJN for the production on marginal lands. In our preliminary data also show that the PsJN-inoculated plants had elevated lignin content compared to controls. We have also begun measuring leaf gas exchange in both PsJN-inoculated and control plants using a Li-COR 6400 photosynthesis system. Plans are to measure leaf physiological properties over the first 12 weeks of growth, starting with newly inoculated plants, to determine the physiological processes involved in the bacterium induced growth enhancement. We will also measure changes that occur during a simulated drought. In contrast to the observed stimulation of switchgrass growth responses caused by PsJN described for cv. Alamo, no beneficial responses were recorded with the upland cultivar Cave-in-Rock. In order to explore this genotype influence further, we are currently conducting a comparative global gene expression profiling in both cultivars following PsJN inoculation, using EST microarrays, in collaboration with Dr. Yuhong Tang of the Noble Foundation, Ardmore, OK. The generated information will be utilized in switchgrass breeding for low input production systems based on genetic compatibility between the host plant and inhabiting microflora. We have worked with Lynchburg Grows (Lynchburg, VA), a non-profit urban farm and environmental educational center that provides job training and educational programs for youth-at-risk and the Central Virginia Governor's school for Science and Technology to set up hydroponic tray units to educate and foster interests in the bioenergy field to the next generation of scientists.

127

Biomass Accumulation in Wide Crosses Between Wild and Domesticated Rice

A. Bordeos,[2] C.E. Jahn,[1] **J. McKay**,[1] **J.E. Leach**[1]* (Jan. Leach@colostate.edu), **D. Bush**,[1] and **H. Leung**[2]

[1]Program in Plant Molecular Biology, Colorado State University, Fort Collins; and [2]Plant Breeding, Genetics and Biotechnology Division, International Rice Research Institute (IRRI), Metro Manila, Philippines

Project Goals: Our goal is to provide the applied biomass research community and industry with information to allow exploitation of the genes and pathways relevant to biomass accumulation in grasses. The specific objectives for this project are to: Objective 1: Identify genes involved in biomass accumulation. Using genetic populations for rice lines that exhibit extremes in biomass accumulation,

we will (a) map QTL using a comprehensive phenotyping/genotyping approach, (b) confirm the QTL location and phenotypic effect and (c) screen a deletion mutant collection to identify large deleted regions corresponding to biomass accumulation. Objective 2: Dissect the QTL using an integrated analysis. First, we will use an expedited approach to generating near isogenic lines containing each QTL. Second, we will integrate genome-wide data to refine the QTL regions including (a) association mapping, (b) expression profiling, (c) mutant analysis, (d) fine scale mapping, and (e) comprehensive sequencing.

Developing a sustainable liquid fuel production from cellulosic feedstock is a major challenge and will require significant breeding efforts to maximize plant biomass production. Rice has an unusual depth of genetic and phenotypic variation due to years of domestication and selection under very diverse environments (e.g. well-watered, flooded, water limited, etc.). Even more variation is available by using wide crosses between very large, perennial wild rice species (*Oryza longistaminata*) and a very small domesticated species (*O. sativa*, cultivar IR64). We are exploiting these wide crosses to identify traits for biomass productivity. The *O. longistaminata* X *O. sativa* (IR64) population was advanced through backcrossing with the recurrent parent IR64 (currently BC_4F_2) to reduce negative characteristics including seed shattering, fertility issues, late flowering, and lodging while maintaining increased biomass. Ten independent families of BC_4F_2 generation were phenotyped for total biomass, height, number of productive tillers, and days to flowering. These plant families produced approximately 2-3 times more biomass, were 30 cm taller, and produced 15 more productive tillers than the parent *O. sativa* IR64. These plants will are being genotyped to determine the genetic loci donated from the wild parent that contribute to the increases in biomass in these families and may represent novel genes to increase biomass in energy crops.

128

Transcription Regulatory Networks of Cell Wall Biosynthesis Revealed by Protein-DNA Interaction Mapping

Li Lin,[1] Naomi Young,[1] Pubudu Handakumbura,[1] Ghislain Breton,[2] Todd C. Mockler,[3] Steve A. Kay,[2] and Samuel P. Hazen[1] (hazen@bio.umass.edu)

[1]Biology Department, University of Massachusetts, Amherst; [2]Cell and Developmental Biology, University of California, La Jolla; and [3]Department of Botany and Plant Pathology, Oregon State University, Corvallis

Project Goals:

1. **Identify regulators of plant cell wall biosynthesis**
2. **Analyze transcription factor genomic binding sites**
3. **Determine effects of regulator perturbation on amenability to deconstruction and cell wall properties of monocots and dicots**

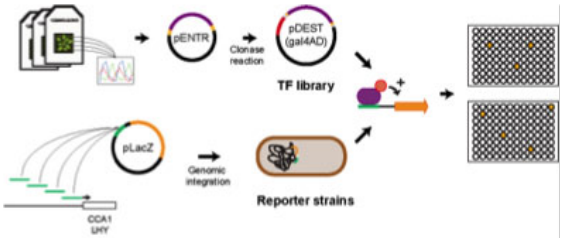


Figure 1. Schematic of the promoter hiking approach (Upper half represents the construction of the transcription factor library from identification of the clones using microarray, cloning in Gateway compatible vectors and transfer into gal4-AD vector for fusion protein expression in yeast. The lower half represents the cloning of tilled promoter fragments in fusion to the LacZ reporter and integration into yeast genome. The library and the yeast strain are then combined for B-galactosidase monitoring assays in 96-well plates.

One mechanism regulating cell wall biosynthesis is the activity of transcription factors that control higher order events of growth and differentiation; the likely direct regulation of processive and non-processive glycosyltransferases as well as the phenylpropanoid metabolic grid. We measured interactions among the promoters of *A. thaliana* cell wall genes and a nearly comprehensive transcription factor library using a high throughput yeast one-hybrid assay (Fig 1). In addition to several NAC and numerous MYB transcription factors some of which have been implicated in cell wall regulation, we measured interactions among over twenty other families with cell wall promoters (Fig 2A). While most (60%) bind just one, 48 many bind multiple promoters. Interestingly, cellulose, hemicellulose, and lignin cis-regulatory regions share several interactors.

Figure 2B describes in detail a protein-DNA interaction sub-network of 6 transcription factors: an AS2, bZIP, MYB, AP2, and two NAC proteins that bind cellulose, hemicellulose, and lignin promoters. Many of these interactions have been confirmed with other *in vitro* methods and *in planta*. Of partuclar interest is a NAC proteins that interacts only with cellulose genes; namely *CESA4/7/8*, *KORIGGAN*, and *COBRA-LIKE1*.

We further characterized this interaction in vivo by electrophoretic mobility shift assay using the fragments. Ultimately, we hope to determine the effects of regulator perturbation on amenability to deconstruction and cell wall properties.

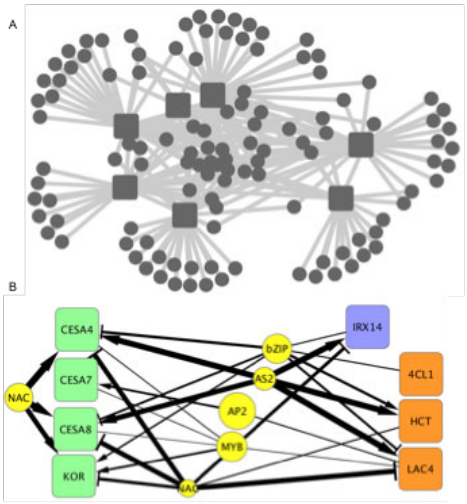


Figure 2. **(A)** An *Arabidopsis thaliana* protein-DNA interaction network of seven cell wall genes and 117 transcription factors. **(B)** A sub-network of three lignin (orange), four cellulose (green), and one hemicellulose (purple) promoter. Squares represent promoter regions and circles the transcription factors (yellow). The size of the circle is proportional to expression level in stem and the thickness of the connector (a.k.a. edge) is proportional to the correlation with the target gene in various microarray experiments. Arrows indicate a positive correlation and a circle a negative correlation.

129

Accelerating the Domestication of *Miscanthus* for Biofuel Production

Changsoo Kim[1]* (changkim@uga.edu), Dong Zhang,[1] Susan Auckland,[1] Lisa Rainville,[1] Katrin Jakob,[2] Jeff Klingenberg,[2] Erik Sacks,[3] Jack Juvik,[3] Cora Muennich,[4] Martin Deuter,[4] Zili Yi,[5] Jianxiong Jiang,[5] and **Andrew H. Paterson**[1]

[1]Plant Genome Mapping Laboratory, University of Georgia, Athens; [2]Mendel Biotechnology, Hayward, Calif.; [3]Department of Crop Sciences, University of Illinois at Urbana-Champaign; [4]Tinplant, Klein-Wanzleben, Germany; and [5]Department of Biotechnology, Hunan Agricultural University, Hunan, Changsha, China

Project Goals: The objectives of this project are to provide genomic tools and resources for an efficient biomass accumulator, *Miscanthus*. A well-saturated molecular linkage map is a prerequisite for enabling effective breeding of high yielding, locally adapted *Miscanthus* varieties for biomass production. In this project, EST-derived SSRs and SNPs are being implemented to construct a genetic map of *Miscanthus*, facilitating comparison between *Miscanthus* and the fully-sequenced sorghum genome by virtue of marker sequence information. The comparative maps are expected to be useful for transferring information from model species (such as sorghum), and for efficiently increasing marker density near genetic loci of specific interest. The comparative map will permit us to infer the locations in *Miscanthus* corresponding to a variety of previously mapped domestication or biomass-determining genes/QTLs from sorghum and other cereals. Detailed map and sequence information for *Miscanthus* will also provide clues to answer practical and fundamental questions about *Miscanthus* genome structure and organization, such as the discrepancy of the basal chromosome numbers between *Miscanthus* and much of the Saccharinae, the levels and patterns of homoeologous gene duplication, and the types and frequencies of genetic polymorphism in both diploids and polyploids that are important to *Miscanthus* improvement.

Research specialization: Tropical grasses included in the Saccharinae have gained attention as biofuel feedstocks because of their high biomass production in part due to their C4 photosynthetic system. However, their complex genomes with large DNA contents and high ploidies complicate genetics and genomics research. *Miscanthus*, a promising cellulosic feedstock owing to perenniality, a longer growing season, greater leaf area, and higher carbon storage per unit of leaf area than some Saccharinae grasses, currently has only limited genetic maps, which are largely based on dominant-type semi-arbitrary markers such as random amplified polymorphic DNA (RAPD). The current research is focusing on building a framework to leverage sequence information from other taxa in *Miscanthus* improvement by integrating simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs). The outcome of this project will clarify genome structure in *Miscanthus* and its evolution since divergence from a common ancestor shared with sorghum (a botanical model for the Saccharinae), as well as provide breeding resources for sustainable biomass production by identifying specific QTL regions and diagnostic markers.

Funding source: DOE-USDA Plant Feedstock Program, Project grant number: 112786

130

Genetic Variation Among Sorghum and *B. distachyon* Accessions for Biological Conversion Quality

Scott J. Lee,[1] Thomas A. Warnick,[2] Naomi F. Young,[1] Jeffrey F. Pedersen,[3] Susan B. Leschine,[2] and Samuel P. Hazen[1]

[1]Biology Department, [2]Department of Microbiology, University of Massachusetts, Amherst; and [3]USDA-ARS, Department of Agronomy, University of Nebraska-Lincoln

Project Goals:

1. **Development of a high throughput translational bioassay for plant biofuel properties**
2. **Assay variation for biological conversion rates among energy crop germplasm**

We developed an assay that provides the ability to measure the impact of pretreatment, conversion processes, and microbial and plant genetic diversity of digestibility,

and thereby determine the potential effects of numerous variables in biofuel production. In contrast to other established methods, the *C. phytofermentans* bioassay provides a direct and quantitative means of assessing feedstock quality, both in terms of digestibility and conversion. The use of *C. phytofermentans* takes into consideration specific organismal interactions, which will be critical in single stage fermentation or consolidated bioprocessing. In this assay, stem tissue from completely senesced plants is pulverized, inoculated with *C. phytofermentans*, and allowed to grow anaerobically. Total sample requirements are as low as 25mg, making multiple measurements of stems from very small plants such as *B. distachyon* and *A. thaliana* possible. Supernatant ethanol concentration measured by HPLC (high performance liquid chromatography) is used as the metric to estimate feedstock quality and conversion efficiency. Ethanol concentration was directly proportional to the concentration of energy crop sorghum feedstock in the culture (Fig panel A) and increased in a linear manner over time (Fig panel B). The assay is capable of detecting significant differences in ethanol production between wild-type sorghum and *brown midrib* (*bmr*) mutants with one or two mutations in the lignin biosynthesis pathway (Fig panel C). We sieved ground biomass and assayed only feedstock ranging from 53 to 62.4 μm and confirmed that differences in ethanol yield are not the result differential grinding among genotypes. We have also detected significant genetic diversity among *B. distachyon* accessions (Fig panel D) sorghum landraces and *A. thaliana* accessions (data not shown). The sorghum accession most amenable to conversion yielded 30% more ethanol than the most recalcitrant, a range we observed among *B. distachyon* and *Arabidopsis thaliana* accessions as well. The genetic differences measured among accessions within several species are similar to the range measured between the *bmr* lignin mutants and wild-type sorghum. While mutations effecting lignin have a deleterious effect on plant architecture, vigor, and yield in many crop species, the accessions we observed significant variation for ethanol yield exhibit no such differences in overall plant architecture. This suggests that gain from selection and transgenic modification for feedstock quality need not be pleiotropic for low biomass yield.

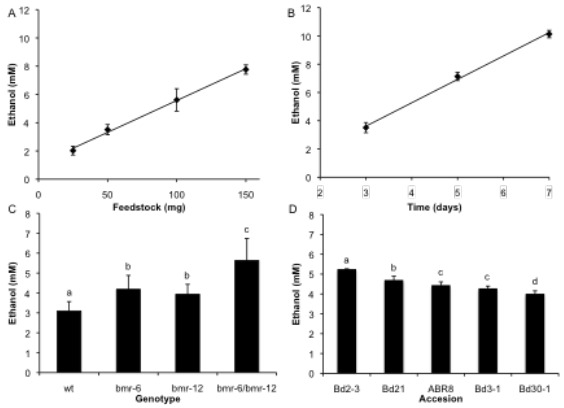


Figure. Bioassay for digestibility using *C. phytofermentans* ethanol production as a measure of feedstock quality. Ethanol production on sorghum is proportional to (A) feedstock concentration and (B) time after inoculation. (C) Diversity of ethanol production by *C. phytofermentans* grown on wild type and single (*bmr-6* and *bmr-12*) and double *brown midrib* (*bmr-6/bmr-12*) lignin biosynthesis sorghum mutant feedstocks. (D) Diversity of ethanol production by *C. phytofermentans* grown on five accessions of *B. distachyon* as feedstock. Values followed by the same letter are not significantly different at $P < 0.05$ based on Duncan's Multiple Range test.

131

Regulation of Root Development in *Populus* in Response to Nitrogen Deficiency and Drought

Victor Busov[1]* (vbusov@mtu.edu), Hairong Wei,[1] Erik Lilleskov,[2] and Yordan Yordanov[1]

[1]Michigan Technological University; and [2]Forest Service

Project Goals: see below

To avoid competition with food production and curb greenhouse gas emissions, lignocellulosic bioenergy crops will be primarily grown on marginal lands. Intermittent, patchy, and limited availabilities of resources like nitrogen (N) and water will severely compromise productivity under such environments. Therefore, sustainable production of bioenergy crops will entail developing varieties that can maintain high levels of biomass productivity under sub-optimal N and water conditions. Root architecture is essential in sensing, absorbing, and transporting both nitrogen and water. The genetic bases of these root traits are poorly understood and thus difficult to manipulate. We use a systems biology approach to understand how nitrogen and water shape root architecture in the lignocellulosic bioenergy crop *Populus*.

The goal of this project is to generate a systems-level knowledge and identify key regulators of root architecture in relation to nitrogen and water use in the bioenergy crop *Populus*. Objectives are to:

1. Perform comprehensive microarray profiling on poplar roots during response to N and water limitations.
2. Construct genetic networks inferred from microarray analysis and generate a systems-level knowledge of the underlying mechanisms.
3. Generate transgenic modifications of key regulators identified in Objective 2 and characterize their response to nitrogen and water.
4. Identify activation-tagged mutants affecting water and nitrogen response, isolate the candidate tagged genes, and recapitulate phenotypes via retransformation in a subset.
5. Use advanced stereo x-ray imaging to characterize root architecture under soil-like conditions in transgenic plants generated in Objectives 3 and 4.
6. Develop web portal for access to, analysis, and dissemination of project data.

To date we have produced a comprehensive transcriptional profile of roots' response to N and water deficiencies. Building on this resource we have used advanced genetic network analyses to generate systems-level knowledge of the underlying molecular mechanisms. We have identified key regulators and have begun modifying their expression in transgenic plants to test the biological significance of the identified network mechanisms. In a parallel and complementary approach taking advantage of the poplar genome

sequence and efficient transformation system, we use activation tagging as a forward genetics approach to discover novel genes or corroborate the effect of genes identified via the genetic network analyses. To this end we have generated 2,000 activation tagged lines and have screened 1,000 for modified response to nitrogen deficiency or drought stress. We have identified and validated 53 mutants which display enhanced growth characteristics under the nitrogen and water stress conditions. For approximately half of these we have already positioned the tag, validated the behavior of the tagged genes and for a limited number initiated recapitulation experiments. A web site which will facilitate access and dissemination of data and germplasm generated through the project is under construction.

132

Genomics of Energy Sorghum Biomass Accumulation

John Mullet[1]* (jmullet@tamu.edu), **Patricia Klein**,[2] and **William Rooney**[3]

[1]Department of Biochemistry and Biophysics, Texas A&M University, College Station; [2]Institute for Plant Biotechnology and Genomics, Texas A&M University, College Station; and [3]Department of Soil and Crop Science, Texas A&M University, College Station

Project Goals: The overall goal of the proposed research is to identify the genetic and biochemical basis for increasing the yield and improving the composition of high biomass cellulosic energy sorghum. The specific objectives of the proposed research are to; (1) characterize the molecular diversity of 750 photoperiod sensitive (late flowering) energy sorghum germplasm accessions and use this information to select 250 diverse accessions for analysis of variation in stem biomass yield, structure, and composition, (2) map QTL for stem biomass yield, structure, and composition in three energy sorghum populations derived from diverse sorghum parental genotypes, and (3) develop information and biological resources that will enable positional cloning of QTL/genes and analysis of gene regulatory networks that modulate energy sorghum biomass yield, stem structure, and composition.

High biomass energy sorghum (*Sorghum bicolor* L. Moench) has excellent potential as a bioenergy crop and research on this species will also provide fundamental information about the genomics of C4 grass energy crop design for the following reasons; (1) sorghum, like *Miscanthus* and energy cane, is a highly productive C4 grass that under optimum conditions can produce ~20dT of biomass per acre making it one of the most productive bioenergy crops currently under development, (2) energy sorghum has excellent drought tolerance and high water use efficiency, critical attributes for production of bioenergy crops in marginal environments and where irrigation is either too expensive or would deplete water reserves, (3) energy sorghum has wide adaptation and is highly amenable to production and cultivation systems currently used in the U.S. facilitating rapid adoption by producers at low risk, (4) energy sorghum is a resilient low risk annual hybrid crop that can be used in normal crop rotations to maintain soil fertility, reduce pest pressures, and meet annual variation in demand for biomass, (5) the extensive and diverse sorghum germplasm collection (~40,000 accessions) contains useful genetic variation for an array of bioenergy traits including biomass yield, composition, and drought tolerance that can be mined and exploited for further improvement of energy sorghum, (6) sorghum's good genetics, relatively small genome size (~800Mbp) and complete genome sequence provides an excellent technology platform for conducting genome-scale research into pathways that influence biomass yield and composition, and (7) information gained through analysis of energy sorghum will be useful for the design of perennial bioenergy C4 grass species such as switchgrass, *Miscanthus* and energy cane that are more complex in genetics and breeding. Information and biological resources generated by this project will be used to create improved versions of high biomass energy sorghum and other C4 bioenergy grasses in order to minimize acreage used for biomass production, reduce food vs. biofuels competition, and reduce the cost of feedstock, while increasing the carbon balance of biofuels and creating a sustainable source of biomass feedstock for large scale biofuels production in the U.S.

133

Genome and Developmental Variation in DNA Methylation in Poplar

Kelly Vining,[1] Kyle R. Pomraning,[2,3] Larry Wilhelm,[4] Henry D. Priest,[4] Cathleen Ma,[1] Ruoqing Zhu,[1] Elizabeth Etherington,[1] Matteo Pellegrini,[5] **Todd Mockler**,[4] **Michael Freitag**,[3] and **Steven Strauss**[1]* (steve.strauss@oregonstate.edu)

[1]Department of Forest Ecosystems and Society, Center for Genome Research and Biocomputing, [2]Molecular and Cellular Biology Program, [3]Dept. of Biochemistry and Biophysics, [4]Department of Botany and Plant Pathology, Oregon State University, Corvallis; and [5]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles

Project Goals:

1. **Produce and characterize DNA methylation-deficient poplars generated by RNA-induced silencing of the poplar homologs of the *Arabidopsis* Decreased DNA Methylation 1 (DDM1) gene, PtDDM1-1 and PtDDM1-2.**
2. **Prepare genomic DNA from tissues of the sequenced *Populus* trichocarpa clone and use commercially-available antibodies against 5-methylcytidine to isolate methylated DNA segments.**
3. **Prepare chromatin from tissues of the sequenced *P. trichocarpa* clone and use commercially available antibodies to methylated lysine residues K9 and K27**

of histone H3 to produce DNA fractions enriched for these modifications.

4. **Sequence the methylated and enriched DNA on an Illumina GAII analyzer at the OSU CGRB and conduct bioinformatics and statistical studies to test the following hypotheses: a. DNA and/or histone methylation differs between differentiation states. b. DNA and/or histone modifications are associated with RNA transcription levels. c. Chromatin states and gene expression are associated.**

We are conducting research to determine the role of epigenetic modifications during tree development using poplar (*Populus trichocarpa*), a reference woody feedstock species. Using methylated DNA immunoprecipitation followed by high-throughput sequencing (MeDIP-seq), we have analyzed DNA methylation patterns in the *P. trichocarpa* genome in relation to four biological processes: bud dormancy and release, mature organ maintenance, *in vitro* organogenesis, and methylation suppression.

We sequenced methylated DNA from eleven target tissues in wildtype *P. trichocarpa*: leaves, roots, xylem, phloem, fall buds, winter buds, spring buds, stem explants, callus and *in vitro* regenerated plants. A total of 64 Illumina sequencing lanes represented 1 to 3 biological replicates for each sampled tissue, 16M – 125M sequencing reads per tissue type. Reads aligning to unique positions in the reference genome covered ~30% of genome space. Average sequence depth within covered regions varied by tissue type, ranging from 4 to 12 reads/nucleotide. Numbers of significantly methylated tiled 1Kb genome windows called by RPKM calculations at a 1% false discovery rate varied by tissue type, ranging from approximately 2,000 (xylem) to 40,000 (pooled bud data). In all tissues, transposons and other repeat elements were enriched relative to their overall representation in the genome, with LTR-gypsy retroelements being the most highly enriched transposable element type. Gene methylation exhibited a pattern of higher methylation at promoters, middle of coding region, and 3' UTRs relative to 5' and 3' ends of coding regions. Numbers of methylated genes varied by tissue type and gene region considered, and represented 3-5% of the genes in the genome. We performed bisulfite sequencing of nine selected target regions with varying MeDIP-seq signal. Results confirmed MeDIP-seq results, and allowed a higher-resolution view of methylation at selected genes.

We have produced summary data for genome methylation in *P. trichocarpa*, including distribution of methylation across chromosomes and in and around genes. This process has been driven by the development and adaptation of bioinformatic and statistical methods. Further, we have analyzed similarities and differences in methylation patterns among tissue types from four biological processes. We have developed a customized genome browser (Gbrowse version 1.69), compatible with the most recent (v2*) P. trichocarpa* genome assembly, at which our data can be explored: http://poplar-dev.cgrb.oregonstate.edu/cgi-bin/gbrowse/poplar_v2/.

134

Poplar Biomass Interactome

Eric Beers,[1] **Amy Brunner**[2]* (abrunner@vt.edu), **Allan Dickerman**,[3] Mingzhe Zhao,[1] Xiaoyan Jia,[1,2] Chengsong Zhao,[1] and Xiaoyan Sheng[2]

[1]Virginia Tech Departments of Horticulture and [2]Forest Resources and Environmental Conservation; and [3]Virginia Bioinformatics Institute at Virginia Tech

http://xylome.vbi.vt.edu/

Project Goals:

1. **Clone approximately 400 Gateway-compatible ORFs corresponding to the poplar xylem gene set (PxORF)**
2. **Identify poplar proteins that co-purify with selected TAP-tagged PxORF proteins expressed in poplar**
3. **Identify high-confidence Y2H interactions for a subset of approximately 60 PxORFs comprising putative regulators of lignocellulose synthesis screened against a poplar xylem cDNA library**
4. **Identify Y2H interactions resulting from a matrix of pair-wise assays between all PxORF proteins**
5. **Produce a protein-protein interaction map that incorporates interactions identified from the three screens**
6. **Maintain a web site to make results available and facilitate distribution of clones**

Proteins are molecular machines that are required for nearly all biological functions based on interactions with other molecules such as carbohydrates, lipids, other low molecular weight molecules, nucleic acids and other proteins. We are mapping protein-protein interactions relevant to biomass production by focusing on proteins coexpressed in poplar secondary xylem. In combination with transcriptomic and metabolomic data, this high-confidence wood interactome will provide a solid foundation for identifying key regulators of wood formation and biomass accumulation.

To date, 374 PxORFs were cloned into Gateway-compatible pENTR vectors, of which 323 were subsequently cloned into pDBdest and 335 into pADdest vectors respectively (http://xylome.vbi.vt.edu/ORF_List). In addition, we prepared transgenic poplar overexpressing 11 PxORFs as TAPa-tagged fusions for co-purification of interacting proteins. We established a replicated field trial of all TAPa-tagged overexpression lines, which will permit phenotypic analysis of the effects of overexpression of these TAPa-tagged proteins and ensure production of sufficient wood for extraction and identification of co-purified proteins. A 108,205 (323DB x 335AD) Y2H binary screen identified 11 PxORF interacting pairs. We are also screening a xylem cDNA prey library for interactors with a subset of PxORFs. For the 26 PxORFs that are completely through the library screen, we have identified 44 unique interacting sequences. Selected interactions have been or will be confirmed by other methods including bimolecular fluorescence complementation and co-immunoprecipitation using plant transient expression systems. The proportional yield from our

binary screen is similar to that represented by the current preliminary binary screen data from the *Arabidopsis* interactome project. In contrast, the proportional yield from our library screen is much higher. Additionally, in most cases, the proportional yield for enzymatic/structural proteins catalyzing metabolic reactions (such as cellulose synthase PB138) is much lower than that of regulatory proteins (such as NIMA kinase PB223). We have begun to integrate our findings for poplar xylem protein interactions with other protein-protein interaction data to produce a preliminary network by whereby poplar proteins are represented by their putative *Arabidopsis* orthologs.

Functional analyses of selected interacting proteins should provide valuable insight regarding new strategies for regulating woody biomass production. Hence, we have begun to functionally characterize select interacting pairs in both poplar and *Arabidopsis* by ectopically expressing or suppressing genes singly and in combination. For example, one interacting pair we are studying is PB15 (ROP-GTPase) and PB129 (DUF620). Co-overexpression of PB15 and PB129 in *Arabidopsis* resulted in expanded interfascicular regions containing enlarged fibers compared to fibers in normal interfascicular regions of the inflorescence stem. However, this phenotype was not observed in transgenics overexpressing just one of these genes, showing the potential of interactome data to be translated into alteration of wood phenotypes.

135

Strategies for Using Molecular Markers to Simultaneously Improve Corn Grain Yield and Stover Quality for Ethanol Production

Rex Bernardo,[1] Hans-Joachim G. Jung,[2] Jon M. Massman[1]* (massm016@umn.edu), Magan F. Lewis,[3] and Robenzon E. Lorenzana[4]

[1]Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul; [2]USDA-ARS, Plant Science Research Unit, St. Paul, Minn.; [3]Department of Plant Sciences, North Dakota State University, Fargo; and [4]AgReliant Genetics, Ivesdale, Ill.

Project Goals: Our objective was to optimize the use of DNA markers to simultaneously breed for high corn grain yield (for non-energy and energy uses) and high stover quality for ethanol production. Specifically, we aimed to (1) determine the prospects of and identify challenges in marker-assisted breeding for both corn grain yield and stover quality traits important for ethanol production; and (2) determine if genomewide selection (which does not require finding markers with significant effects) is superior to the usual approach of selecting only for significant markers, with the goal of simultaneously improving corn grain yield and stover quality.

About 235 million metric tons of corn (*Zea mays* L.) stover (i.e., stalks, leaves, cobs, husks, and tassels) are left unharvested in U.S. corn fields each year. This stover represents a most abundant source of lignocellulosic substrate that can be converted to ethanol biofuel. But while today's corn hybrids have been aggressively bred for grain yield, they have not been bred for stover-quality traits important for ethanol production. Here, our objective was to optimize the use of DNA markers to simultaneously breed for high corn grain yield (for non-energy and energy uses) and high stover quality for ethanol production.

We used SNP markers combined with classical quantitative-trait analyses to study the extent to which grain yield, agronomic traits, and stover quality can be simultaneously improved in the B73 x Mo17 corn population. Three stover-quality traits were measured: concentration of cell wall glucose in dry stover ("Glucose"); cell wall glucose released from the stover by thermochemical pretreatment and enzymatic saccharification ("Glucose Release"); and concentration of lignin on a cell-wall basis ("Lignin"). Genetic variances were significant for grain yield, moisture, stalk and root lodging, plant height, and all three stover-quality traits. Heritabilities of the stover quality traits were 0.57 for Glucose, 0.63 for Glucose Release, and 0.68 for Lignin. Genetic and phenotypic correlations among traits were generally favorable but also reflected the complexity of corn stover cell wall composition. We found 152 QTL, mostly with small effects, for Glucose Release and cell wall components on both a dry matter and cell wall basis. Because no major QTL were found, we expected that methods that predict performance based on markers, such as genomewide selection, would be appropriate in marker-assisted breeding for these traits. Responses to three cycles of selection for Glucose, Glucose Release, and Lignin were higher with genomewide selection (which utilized all markers rather than only those with significant effects) than with selection based only on significant markers. These responses were determined from NIRS predictions, and we are conducting wet chemistry tests to measure Glucose, Glucose Release, and Lignin in the populations. To our knowledge, this work represents the first report of the usefulness of genomewide selection based on empirical data in plants.

We conclude that current corn-breeding programs should be able to incorporate stover quality for cellulosic ethanol as a breeding objective, without having to use unadapted or exotic germplasm. Given the absence of major QTL and the complexity of the traits, we recommend genomewide selection for the improvement of stover-quality traits for cellulosic ethanol in corn.

136

Alfalfa Transcript Sequencing and SNP Discovery to Improve Biomass Composition

Xuehui Li,[1] Jiqing Guo,[1] John A. Crow,[3] Andrew D. Farmer,[3] Yuanhong Han,[1] Ji He,[1] Ryan Kim,[3] Mingyi Wang,[1] Yanling Wei,[1] **Gregory D. May**,[3] **Maria J. Monteros**,[1,2] and **E. Charles Brummer**[1,2*] (ecbrummer@noble.org)

[1]The Samuel Roberts Noble Foundation, Ardmore, Okla.; [2]Institute of Plant Breeding, University of Georgia, Athens; and [3]National Center for Genome Resources, Santa Fe, N.M.

Project Goals:

1. **Sequence alfalfa (*Medicago sativa L.*) transcriptomes to identify common SNP variants and develop a high throughput GoldenGate SNP array, and**
2. **Identify SNP markers associated with biofeedstock-related traits in diploid and tetraploid mapping populations.**

Alfalfa is an important forage crop in temperate and dry tropical regions in the world; it also is a potential biofuel crop. Alfalfa has the advantages of high yield, a high lignocellulose concentration in stems, and low input costs due to high levels of biological nitrogen fixation. Marker assisted selection (MAS) and/or genomic selection (GS) could enhance alfalfa improvement, but large numbers of markers are needed to map important agronomic traits and predict genomic breeding values. The main objective of this study is to identify single nucleotide polymorphism (SNP) for alfalfa using Illumina transcriptome sequencing and to develop SNP assays in candidate genes for biomass yield and composition. We have sequenced 27 alfalfa transcriptomes, including elite genotypes from four major alfalfa breeding companies in the U.S. A pilot sequencing of three genotypes resulted in a total of 80.7 million reads, assembling of which generated 155,484 contigs with a total length of 51.3 Mbp and an average length of 330 bp, giving an average read depth of 36-fold for each genotype. The realignment of reads to the contigs enabled the detection of 260,848 putative SNPs (one SNP per 197 bp) and 11,090 InDels among the three genotypes. Over 95% of the SNPs have coverage of 5 or more reads for each of the three genotypes. Of all contigs, 55.6% were aligned to *M. truncatula* coding sequence, build 3.0, which carry about 190,000 SNPs. The distribution of these SNPs along eight chromosomes is roughly even. We are developing marker assays for selected SNPs in order to map important agronomic traits and assess linkage disequilibrium (LD) and population structure in breeding populations.

We have evaluated the LightScanner high resolution melting (HRM) technology from Idaho Technologies for use with autotetraploid alfalfa. The technology enables clear assessment of allele dosage and is reasonably high throughput. However, assessment of hundreds or thousands of loci is beyond the routine capacity of this technology. Therefore, we are discussing array design with Illumina. Illumina now offers the ability to distinguish among multiple heterozygote classes (AAAT, AATT, and ATTT, for example) making this platform will be able to generate useful data for alfalfa breeding applications. Using short read sequence data generated in this and other experiments, we have identified SNP in several genes in the lignin biosynthetic pathway and mapped them on the alfalfa genetic map. We will preferentially identify SNP in other cell-wall related genes as we analyze the transcriptomes of the 27 genotypes to include on the array.

We have developed genetic maps in three diploid alfalfa populations in addition to maps developed previously in two tetraploid populations. We are in the process of mapping QTL for cell wall composition on the tetraploid populations and are growing two diploid populations in the field, which will be analyzed for yield and composition. SNP markers developed as part of this project will be used to locate candidate genes on the linkage maps for comparison to QTL locations.

137

The Regeneration and Transformation of Foxtail Millet (*Setaria italica*), A Model Biofuel Crop

Xinlu Chen* (xchen24@utk.edu), **Janice Zale,** and **Feng Chen**

Department of Plant Sciences, University of Tennessee, Knoxville

Project Goals: The goal of this project is to establish genetic transformation systems for foxtail millet (*Setaria italica*), a newly established model plant for genetic improvement of other biofuel crops, such as switchgrass (*Panicum virgatum*).

Foxtail millet (*Setaria italica* L.) is a warm-season C4 annual crop commonly grown for grain and forage production worldwide. Recently, foxtail millet was established as a model plant for the genetic improvement of other biofuel crops, especially switchgrass (*Panicum virgatum* L.). A number of genomic tools have been established for foxtail millet, including a fully sequenced genome. The goal of this project is to establish genetic transformation systems for foxtail millet. Seeds and immature inflorescences are used as explants. Protocols for callus initiation, somatic embryo formation, and plantlet regeneration from explants have been developed for the foxtail millet genotype Yugu1. Optimal media for the induction of callus and somatic embryos from immature inflorescence explants is determined to be Murashige and Skoog (MS) medium containing 2.5 mgl-1 2,4-dichlorophenoxyacetic acid (2,4-D), 0.6 mgl-1 6-benzylaminopurine (BAP), and 3% sucrose. Calli induction from explants of immature inflorescences was significantly higher (76%) in comparison to the other two media tested (Table 1), whereas calli induction from seed explants was also relatively high

on this medium compared to six others tested (68.5%; data not shown). Callus induction from this medium is used for Agrobacterium-mediated transformation using AGL1 strain harboring a binary vector pCAMBIA1305.1. Transient expression of the reporter GUS gene is improved with treatments by sonication, vacuum, centrifugation and addition of L-cysteine and DTT. The higher centrifugation speed increased the number of GUS foci (Figure1). The Calli inoculated are cultured in the dark at room temperature. After three days, the calli are transferred onto callus induction medium containing 150 mg l-1 cefotaxime and 150 mg l-1 timentin. After two weeks, the calli are subcultured onto callus induction medium containing 150 mg l-1 cefotaxime, 150 mg l-1 timentin and 1.5 mg l-1 hygromycin. After two to three months, the calli are transferred to regeneration medium (MS medium, 0.2 mg l-1 kinetin, 3 % sucrose, pH 5.8) containing 150 mg l-1 cefotaxime, 150 mg l-1 timentin and 1.5 mg l-1 hygromycin. Finally Putative transgenic foxtail millet plants have been regenerated from callis.

Table 1. Media Comparison for Induction of Calli From Immature Inflorescences

Media	Explants Plated	Calli Induced	Percentage of Calli Induced	Mean Percentage[a] ± S.E.[b]
MS, 2.5 mgl^{-1} 2,4-D, 0.6 mgl^{-1} BAP	560	426	76.1	76.1 ± 5.3 a
MS, 5.0 mgl^{-1} 2,4-D, 1.1 mgl^{-1} BAP	210	0	0.0	0.0 ± 0.0 b
N6E	210	8	3.8	4.0 ± 0.0 b

[a]Mean percentages followed by the same letter are not significantly different at the 5% level as determined by Tukey's Multiple Comparison; there were four replicates of each treatment. [b]S.E. = standard error.

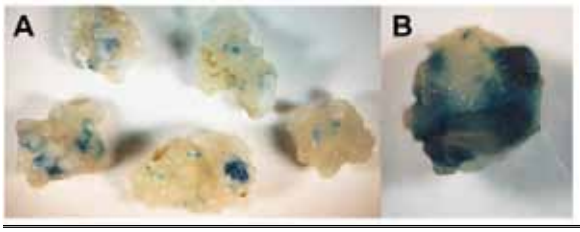


Figure 1. Transient GUSPlus™ expression in foxtail millet calli after treatments with sonication, vacuum and centrifugation. Gus foci showed after centrifugation for one minute at (A) 7,000 rpm and (B) 13,000 rpm.

138

Ploidy Variation and Reproductive Pathways in Upland Switchgrass

Edward S. Buckler,[1] **Michael D. Casler**,[2] **Jerome H. Cherney**,[3] Patrick Redmond,[4] and Denise Costich[1]* (dc58@cornell.edu)

[1]USDA-ARS; [2]USDA-ARS U.S. Dairy Forage Research Center, Madison, Wis.; [3]Department of Crop and Soil Sciences, Cornell University; and [4]Cornell University

http://www.maizegenetics.net

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

Many of the plant species that have been targeted for biofeedstock development have complex, polyploid genomes with limited prior research devoted to them. Switchgrass is a prime example of this. We employed a combination of approaches, including flow cytometry, classic cytology and molecular cytogenetics, to gain a better understanding of the extreme variation in chromosome numbers found in this species. Knowledge of chromosome-number variation in our key germplasm will be critical for the interpretation of genetic marker data, genetic mapping and breeding efforts that rely on marker-assisted selection.

In a recently published study (Costich et al. 2010. Plant Genome), our flow cytometric survey of a core set of 11 primarily upland polyploidy switchgrass accessions indicated that there was considerable variation in genome size within each accession, particularly at the octoploid ($2n$ = 8X = 72 chromosome) ploidy level. Highly variable chromosome counts in mitotic cell preparations indicated that aneuploidy was more common in octoploids (86.3%) than tetraploids (23.2%). The incidence of hyper- versus hypoaneuploidy is equivalent in tetraploids, however, this is clearly not the case in octoploids where close to 90% of the aneuploid counts are lower than the euploid number. Fluorescent in situ hybridization (FISH) revealed an unexpected degree of variation in chromosome structure underlying the apparent genomic instability at the octoploid level.

Polyploidy and reproductive biology are linked by the underlying mechanism of unreduced gamete formation, that is, the production of eggs and/or sperm with the somatic chromosome number. This one alteration in the outcome of meiosis can have profound effects on the reproductive success of the individual plant and on the overall population structure, affecting gene flow and the distribution of genetic diversity. As a follow-up to the research described above on

switchgrass ploidy and aneuploidy, we have initiated a study of the reproductive pathways in tetraploids and octoploids, examining the ploidies and genetic relatedness of maternal parent plants (both 4X and 8X) and their offspring (seeds). A flow-cytometric seed screen (FCSS; Matzke et al. 2000. Plant Journal) was carried out to compare the ploidies of the embryo and endosperm cell populations in seeds with the ploidy of the parent. Sets of seed were germinated and grown up to confirm the FCSS analysis and will be genotyped to examine the apomictic versus sexual nature of their origin. A better understanding of the reproductive biology of this species will provide the foundation for more efficient breeding programs, as well as, improved analysis and interpretation of the sequence data being generated by ongoing genomics projects.

139

Genomics of Wood Formation and Cellulosic Biomass Traits in Sunflower

Jessica G. Barb[1]* (jgbarb@uga.edu), Sukhpreet Sandhu,[1] John E. Bowers,[1,2] Angela Ziebell,[3] Crissa Doeppke,[3] Robert Sykes,[3] **Jeffrey F. Dean**,[4] **C. Joseph Nairn**,[4] **Mark Davis**,[3] **Laura F. Marek**,[5] **Steven J. Knapp**,[1] and **John M. Burke**[2] (jmburke@uga.edu)

[1]Center for Applied Genetic Technologies, Univ of Georgia, Athens; [2]Dept of Plant Biology, Univ of Georgia, Athens; [3]US Dept of Energy, National Renewable Energy Laboratory, Golden, Colo.; [4]School of Forestry and Natural Resources, Univ of Georgia, Athens; [5]USDA-ARS and Dept of Agronomy, Iowa State Univ, Ames

Project Goals: see below

The long-term goals of this project are to: (a) develop woody, high biomass cultivars of common sunflower (*Helianthus annuus* L.) for biofuel production; (b) gain insights into genetic and non-genetic factors affecting xylogenesis, secondary cell wall differentiation, lignocellulose accumulation, wood formation, biomass yield, and other cellulosic biomass traits in sunflower; and (c) gain an understanding of the chemical, physical, and biofuel properties of sunflower 'wood'. The oilseed sunflower hybrids grown around the world today produce pithy stems, and wood-forming ecotypes have not been discovered in *H. annuus*. In contrast, several wild sunflower species produce woody stems, are interfertile with *H. annuus*, and have the potential to supply genetic diversity for developing woody cultivars and enhancing cellulosic biomass yield in sunflower. Xylogenesis, wood formation, and lignocellulose accumulation have not been studied, selection for cellulosic biomass traits has not been done, and wood-producing cultivars have not been developed in sunflower. This project thus focuses on the development of the resources and knowledge needed for manipulating cellulosic biomass traits in hybrid sunflower breeding programs, and for tapping genetic diversity for wood formation, cellulosic biomass yield, and other traits in two drought-tolerant, wood-forming, wild species: silverleaf sunflower (*H. argophyllus* L.) and Algodones dune sunflower (*H. niveus* subsp. *tephrodes* (Gray) Heiser).

The specific objectives of this project are to:

1. Identify anatomical, developmental, physical, and chemical differences in the stems of woody and non-woody ecotypes of sunflower.
2. Develop an EST database from relevant tissues for use in SNP discovery and microarray development; perform comparative transcriptomic analyses of woody and non-woody sunflower ecotypes to identify genes responsible for wood development and cellulosic biomass traits in sunflower.
3. Characterize the variation present within *H. annuus* and *H. argophyllus* in terms of genetic diversity, wood chemistry properties, biomass traits, and other agronomic traits.
4. Investigate the genetic architecture of biomass traits and wood chemistry in sunflower using a genetic map-based approach to identify genomic regions harboring QTL for wood formation and other cellulosic biomass traits; develop QTL-NILs for wood formation, cellulosic biomass, and other traits.
5. Construct comparative genetic maps of *H. annuus*, *H. argophyllus*, and *H. niveus* ssp. *tephrodes* to characterize chromosomal differences that might limit gene introgression from the wild species into the cultivated sunflower gene pool.

140

Developing Genomic Selection (GS) and Genome-Wide Association Studies (GWAS) for Upland Switchgrass

PIs: **Edward S. Buckler**[1]* (esb33@cornell.edu), **Michael D. Casler**,[2] and **Jerome H. Cherney,**[3]
Co-Authors: Denise Costich,[1] Elhan Ersoz,[6] Moria Sheehan,[4] Fei Lu,[5] and Alexandra Lipka[5]

[1]USDA-ARS; [2]USDA-ARS U.S. Dairy Forage Research Center, Madison, Wis.; [3]Department of Crop and Soil Sciences, Cornell University; [4]Nature Source Genetics; [5]Cornell University; and [6]Sygenta

http://www.maizegenetics.net

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

The ability to predict and model trait variation with genomic markers has tremendous opportunities for the breeding of perennial crops. In the standard breeding cycle of perennials, it can be five or more years between cycles of advancement, while genomic selection permits advancement every time a cross can be made, for example, every 1-2 years. To enable genomic selection and genome wide association analysis in switchgrass, we have developed germplasm resources from upland switchgrass, which includes bi-parental mapping populations and association mapping populations. Replicated phenotypic trials in both Wisconsin and New York have shown that while there is tremendous phenotypic variation for a wide range of traits, this species has substantial problems with establishment, which is exacerbated by cold winters and soil conditions. We have combined these trials with two studies on the genome of switchgrass—chromosome biology and high throughput genotyping. Analysis of the chromosome biology found substantial instability in chromosome number, with tetraploids, hexaploids, and octoploids all found in our germplasm. There were also many accessions showing a gain or loss of a few chromosomes in mitotic cells (aneuploidy). While the chromosome variation is tractable, it must be considered in any breeding effort, and it strongly favors using the tetraploids. Finally, genotyping-by-sequencing approaches have been applied to switchgrass in order to identify thousands of variable regions of the genome. These studies are enabling genomic selection models, which will be evaluated in the coming months. Overall, upland switchgrass has tremendous variation, but whether it is bred directly using molecular markers, or traits from upland switchgrass are introgressed with the aid of markers into lowland switchgrass, this study provides a foundation for advancement and identifies the challenges.

141

Transcriptional Genomics in Maize for Improvement of Bioenergy Grasses

Bryan W. Penning,[1,2] Christopher K. Dugard,[1] Nicholas Babcock,[3] John F. Klimek,[1] Robert Sykes,[4] Mark Davis,[4] Clifford Weil,[3] Maureen C. McCann,[2] and **Nicholas C. Carpita**[1]* (carpita@purdue.edu)

[1]Departments of Botany and Plant Pathology, [2]Biological Sciences, and [3]Agronomy, Purdue University, West Lafayette, Ind.; and [4]National Bioenergy Center, National Renewable Energy Laboratory, Golden, Colo.

Project Goals: The goals of our project are to use bioinformatics and high throughput sequencing technologies to identify and classify the genes involved in cell wall formation in maize for translation to other bioenergy grasses. Targets for genetic modification will be identified and tested for enhanced processing ability.

The backbone genome sequences of *Arabidopsis* and rice provided functional annotations for an estimated 1500 genes in maize that are implicated in cell wall development in grass species [1; http://cellwall.genomics.purdue.edu/]. Highly parallel pyrosequencing was used to profile gene expression at defined stages of development for the maize internode characteristic of cell division, cell enlargement, fiber differentiation, secondary wall cellulose deposition, and lignification. Algorithms were developed to determine co-expression of suites of genes associated with these stages of development and specifically for primary and secondary cell wall cellulose synthase isoforms, mixed-linkage b-glucan synthases, xylan synthases, and enzymes of the phenylpropanoid and monolignol synthesis pathways. In parallel, we have used Pyrolysis Molecular Beam MS as a high throughput means to classify cell wall architectures of maize stover from the Intermated B73 x Mo17 (IBM) recombinant inbred population. High-density mapping of these diverse populations permitted determination of several quantitative trait loci (QTL) for traits of hexose and pentose, *p*-coumaric acid, guaiacyl and syringyl lignin abundance. We are currently employing comparative expression analysis in B73, Mo17 and target IBM lines to refine candidate genes in the B73/Mo17 genomes responsible for biochemical contributions to diverse architectures. Another line of experiments is defining small RNA populations during different developmental stages. From these populations, potential regulatory sequences derived from naturally occurring antisense transcripts of candidate genes are being elucidated by strand-specific PCR.

This work is supported by the Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under Contract No. DE–FG02–08ER64702

Reference

1. Penning et al. (2009) Genetic resources for maize cell wall biology. *Plant Physiol.* 151, 1703-1727

142

Genomic Analysis of miRNAs and Target RNAs of *Brachypodium*

Skye A. Schmidt*[1] (sschmidt@udel.edu), Dong-Hoon Jeong,[1] Linda Rymarquis,[1] Monica Accerbi,[1] Sunhee Park,[1] Marcelo A. German,[2] Matthias Ganssmann,[1] Jixian Zhai,[1] Emanuele DePaoli,[3] Sai Guna Ranjan Gurazada,[4] Blake C. Meyers,[1] and **Pamela J. Green**[1]

[1]Delaware Biotechnology Institute, University of Delaware, Newark; [2]Present address: Dow AgroSciences LLC, Portand, Ore.; [3]Present address: Istituto Agrario San Michele all'Adige, Trento, Italy; [4]Present address: DuPont Crop Genetics, Wilmington, Del.

Project Goals: In this study, we use next-generation sequencing technology to identify miRNAs in different tissues and following abiotic stress treatments of *Brachypodium distachyon*, a rapidly developing model system for temperate grasses and bioenergy crops.

miRNAs are small, endogenous RNAs that post-transcriptionally regulate gene expression in nearly all eukaryotic systems. In plants, miRNAs can serve as major regulators of development, stress responses, metabolism, and other processes through the miRNA-guided cleavage of specific target RNAs. While miRNAs and their target interactions are well-characterized in systems such as *Arabidopsis* and rice, little is known about their roles in others including temperate grasses and potential bioenergy crops. In this study, we use next-generation sequencing technology to identify miRNAs in different tissues and following abiotic stress treatments of *Brachypodium distachyon,* a rapidly developing model system for temperate grasses and bioenergy crops. A total of more than 64 million reads were obtained from 12 small RNA libraries, resulting in an average of more than 1.4 million distinct genome-matched small RNA sequences per library, from which both conserved and new miRNAs have been identified. To identify the targets of these miRNA on a global scale, we use an approach called Parallel Analysis of RNA Ends (PARE) that facilitates the sequencing of 3' products of miRNA-guided target RNA cleavage. Because miRNAs and their targets can form missing links in many important gene regulatory networks, the identification of miRNA and target RNA pairs in *Brachypodium* will help to better understand how small RNAs contribute to the regulation of genes and genomes.

This work was funded by grant # DE-FG02-07ER64450 of the Plant Feedstock Genomics for Bioenergy Program of the DOE.

143

Insertional Mutagenesis of *Brachypodium distachyon*

John Vogel*[1] (john.vogel@ars.usda.gov), Jennifer Bragg,[1] Jiajie Wu,[1,2] Yong Gu,[1] Gerard Lazo,[1] and Olin Anderson[1]

[1]USDA-ARS, Western Regional Research Center, Albany, Calif.; and [2]Shandong Agriculture University, Taian, Shandong, China

Project Goals:
(renewal project).

1. **Generate 30,000 insertional mutants.**
2. **Sequence DNA flanking the insertion sites of all mutants to place the insertions in a genomic context.**
3. **Collaborate with other groups creating T-DNA lines to integrate their lines into our database.**

Herbaceous energy crops, especially grasses, are poised to become a major source of energy in the United States. Despite their increasing importance, we know little about the basic biology underlying the traits that control the utility of grasses as energy crops. Better knowledge of basic grass biology (e.g. identification of the genes that control cell wall composition, plant architecture, cell size, cell division, reproduction, nutrient uptake, carbon flux, etc.) could be used to design rational strategies for crop improvement and shorten the time required to domesticate these species. The use of an appropriate model system is an efficient way to gain this knowledge. Unfortunately, due to its distant relationship to monocots, *Arabidopsis* is not suitable to study biological features unique to the grasses (e.g. cell wall composition). *Brachypodium distachyon* (*Brachypodium*) is a small annual grass with all the attributes needed to be a modern model organism including simple growth requirements, fast generation time, small stature, small genome size and self-fertility. These attributes led to the recommendation in the DOE's "Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda" report to propose developing and using *Brachypodium* as a model for energy crops to accelerate their domestication. Strategic investments in *Brachypodium* by the DOE are now bearing fruit and *Brachypodium* is rapidly being adopted as the grass model of choice by hundreds of laboratories worldwide. The DOE-JGI has recently sequenced the entire *Brachypodium* genome. The genome sequence and annotation are of unprecedented quality for a draft plant genome and will serve as a firm foundation for a host of functional genomic tools. *Brachypodium* is also readily transformed by *Agrobacterium tumefaciens.* Indeed, with average transformation efficiencies around 50% *Brachypodium* is now, arguably, the easiest grass to transform. Other resources available to *Brachypodium* researchers include a large germplasm collection, resequenced genomes, microarrays, molecular markers, a high-density genetic map, BAC libraries and physical maps. In addition key protocols have been optimized for *Brachypodium* including: efficient crossing, chemical mutagenesis, radiation mutagenesis and live root imaging.

Sequence indexed insertional mutants are an extremely powerful tool for both forward and reverse genetics. We have begun to create a collection of *Brachypodium* T-DNA mutants. Our T-DNA tagging project was initially funded in 2007 and was renewed in 2010. Using our high-efficiency *Agrobacterium tumefaciens*-mediated transformation method, we have generated 8,700 *Brachypodium* T_0 lines. We sequenced the DNA flanking the insertion sites in 7,111 lines and assigned 8,107 of the resulting flanking sequence tags (FSTs) in 4,393 insertional mutant lines to 5,348 unique locations in the *Brachypodium* genome. These include 1,559 insertions in genes and 1,039 insertions close to genes (1,000 bp upstream or 500 bp downstream). Information about the WRRC *Brachypodium* insertional mutant population is available in a searchable website designed to allow researchers to order T-DNA lines with mutations in genes of interest. Protocols for working with *Brachypodium*, information about the T-DNA project, and instructions for ordering T-DNA lines are available at http://brachypodium.pw.usda.gov. We have just completed the first cycle of funding and significantly exceeded our key objectives. We created 1,200 more lines and sequenced the flanking DNA in 1,111 more lines than planned. The two postdocs working on this project have both moved to new positions: one is doing research in China and the other is now working on our related phenomics project. A new postdoc and technician have just been hired to continue the project. The goal of the renewal project is to generate another 30,000 T-DNA lines. We have also established collaborations with eight laborato-

ries from five countries to create an international *Brachypodium* T-DNA collection and, together with our collection, we have plans to create a collection of 65,000 T-DNA lines.

144

Characterization of Novel Cell Wall Mutants in Maize

F.J. Kraemer,[1] T. Thomik,[1] S. Hake,[3] and **M. Pauly**[1,2]

[1]Energy Biosciences Institute, University of California-Berkeley; [2]Department of Plant and Microbial Biology, University of California-Berkeley; and [3]Plant Gene Expression Center, USDA-ARS, Albany, Calif.

In recent years plant cell walls and the polymers that constitute them have received increased attention as a potential highly abundant renewable resource for biofuel production. The feasibility of using plant feedstocks will in part be dependent on the optimization of plant wall composition, as it can directly influence conversion yield. Therefore the understanding of plant cell wall polymer biosynthesis and metabolism and the genes involved therein will be essential. Multiple members of the *Poales* order have been proposed as potential bioenergy feedstocks as they combine multiple desirable traits such as C4 photosynthesis, large biomass yield and fast growth. Examples include switchgrass, *Miscanthus* or sugar cane. Also crop residues like corn stover or wheat straw could be utilized.

We performed a forward genetic screen to identify mutants with alterations in their cell wall monosaccharide composition. Mutaganized lines of *Zea mays* (chemical mutagenesis) were analyzed and multiple lines with altered monosaccharide composition have been identified. A summary of the screen and detailed data on promising maize candidates will be presented and discussed.

145

Identification of Candidate Genes Using Rice Mutants for Biomass Engineering In Switchgrass

B. Broeckling,[1] M. Baroidan,[2] C.E. Jahn,[1] **J.K. McKay**,[1] **J.E. Leach**[1]* (Jan.Leach@colostate.edu), **D.R. Bush**,[1] and **H. Leung**[1]

[1]Program in Plant Molecular Biology, Colorado State University, Fort Collins; and [2]Plant Breeding, Genetics and Biotechnology Division, International Rice Research Institute (IRRI), Metro Manila, Philippines

Project Goals: Our goal is to provide the applied biomass research community and industry with information to allow exploitation of the genes and pathways relevant to biomass accumulation in grasses. The specific objectives for this project are to: Objective 1: Identify genes involved in biomass accumulation. Using genetic populations for rice lines that exhibit extremes in biomass accumulation, we will (a) map QTL using a comprehensive phenotyping/genotyping approach, (b) confirm the QTL location and phenotypic effect and (c) screen a deletion mutant collection to identify large deleted regions corresponding to biomass accumulation. Objective 2: Dissect the QTL using an integrated analysis. First, we will use an expedited approach to generating near isogenic lines containing each QTL. Second, we will integrate genome-wide data to refine the QTL regions including (a) association mapping, (b) expression profiling, (c) mutant analysis, (d) fine scale mapping, and (e) comprehensive sequencing.

Developing a sustainable biofuels program that makes significant contributions to the national energy budget requires unprecedented inputs of biomass for energy conversion. Our specific objectives are to identify genes and pathways in rice (*Oryza sativa*) that increase plant biomass in order to translate this information into cultivar improvement for new energy crops, such as switchgrass. We have screened over 12,000 chemical and irradiation-induced mutants of rice and identified 13 mutants with increased biomass. Our initial effort focused on one diepoxybutane mutant (DEB1) that has reproducibly shown a two-fold increase in biomass under both field and greenhouse conditions. Diepoxybutane is predicted to cause small deletions in the genome and to identify the deletions in DEB1, we performed a comparative genomic hybridization experiment using an *Oryza sativa* whole genome array. We identified 25 deletions in DEB1, ranging in size from 90 to 5,721 bps. Identification of the deletion responsible for the high biomass phenotype is currently in progress. Comparative genomic hybridization, plus the incorporation of new approaches such as short read whole genome sequencing, will be used to identify candidate genes from the remaining 12 mutants. The contribution of candidate genes to biomass improvement will be validated in both rice and switchgrass using a transgenic approach. Plants with perturbations in these candidate genes will be comprehensively phenotyped by transcript and physiological profiling. This systems biology approach will enable identification of key regulatory networks relevant to bioenergy traits for further rational engineering of switchgrass.

146

Identification of Genes That Regulate Phosphate Acquisition and Plant Growth During Arbuscular Mycorrhizal Symbiosis in *Brachypodium distachyon*

Jeon J. Hong,[1] S. Karen Gomez,[1] Quan Zhang,[1] Gopala Battu,[2] Liudmila S. Yafremava,[2] Kranthi Varala,[2] Matthew Hudson,[2] and Maria J. Harrison[1]* (mjh78@cornell.edu)

[1]Boyce Thompson Institute for Plant Research, Ithaca, N.Y.; and [2]Department of Crop Science, University of Illinois, Urbana

Project Goals: The goals of this part of our project are to establish *Brachypodium distachyon* as a model to study AM symbiosis in grasses of relevance as bioenergy crops and to use this system to dissect the molecular basis of differences in functionality in different AM symbioses.

Most vascular flowering plants are able to form symbiotic associations with arbuscular mycorrhizal (AM) fungi. The symbiosis develops in the roots and has a profound effect on plant productivity, largely through improvements in plant mineral nutrition (Smith and Read, 2008). All proposed bioenergy crops including legumes, grasses and trees, are capable of forming AM symbioses, and therefore have the potential to benefit from phosphorus and nitrogen acquisition through the symbiosis. This is significant because phosphorous and nitrogen are the two mineral nutrients whose availability most frequently limits plant growth (Vance, 2001; Vance et al., 2003; Tilman et al., 2006)

Brachypodium distachyon is a wild grass species that serves as a model for the temperate grasses, including those proposed as bioenergy crops. There are several detailed ecological studies of AM symbioses of grasses, including studies of the genus *Brachypodium* (van der Heijden et al., 2003; van der Heijden et al., 2006), and these indicate the potential of the AM symbiosis for increasing plant growth in low phosphate soils. However, these studies also illustrate significant differences in plant performance depending on the AM fungal species involved. Variation in plant performance during symbiosis with different AM fungal symbionts is a well documented phenomenon; however, the molecular basis is not understood.

The goals of this part of our project are to establish *Brachypodium distachyon* as a model to study AM symbiosis in grasses of relevance as bioenergy crops and to use this system to dissect the molecular basis of differences in functionality in different AM symbioses. Sequence-based transcript profiling transcript will be used to document transcript profiles in *B. distachyon* AM symbioses that differ with respect to plant performance. From these datasets, transcript profiles and potentially plant processes associated with maximal plant performance will be identified.

Towards these goals, growth systems that enable development of *B. distachyon*-AM symbioses have been established and the interactions of *Brachypodium distachyon* with *Glomus versiforme, Glomus intraradices, Glomus white, Gigaspora gigantea* and *Gigaspora decipiens* have been assessed. These AM fungi all colonize *B. distachyon* roots and establish symbiosis, but their effects on growth of *B. distachyon* and on mineral nutrition, vary significantly.

In *Medicago truncatula* and rice, array-based transcript profiling has been used to document the transcriptional responses to development of AM symbiosis and sets of AM symbiosis-induced genes have been identified (Liu et al., 2003; Güimil et al., 2005; Liu et al., 2007; Gomez et al., 2009). We identified *B. distachyon* orthologs of rice and *M. truncatula* AM-symbiosis induced genes and then examined their expression in *B. distachyon* during symbiosis with different AM fungi. While the expression patterns of the 'AM-specific genes' are conserved in *M. truncatula*, rice and *B. distachyon*, there were some surprising differences in AM-symbiosis induced gene expression patterns in *B. distachyon* relative to rice. For example, expression of a rice peroxidase, OSAM1, is highly induced in rice-AM symbiosis but the predicted ortholog is not induced in *B. distachyon* AM symbiosis. To extend these profiling analyses, Illumina-based transcript profiling of *B. distachyon* during AM symbiosis with three AM fungi is in progress.

References

1. Gomez, S.K., Javot, H., Deewatthanawong, P., Torrez-Jerez, I., Tang, Y., Blancaflor, E.B., Udvardi, M.K., and Harrison M, J. (2009). *Medicago truncatula* and *Glomus intraradices* gene expression in cortical cells harboring arbuscules in the arbuscular mycorrhizal symbiosis. *BMC Plant Biology* 9, 1-19.
2. Güimil, S., Chang, H.-S., Zhu, T., Sesma, A., Osbourn, A., Roux, C., Ioannidis, V., Oakeley, E.J., Docquier, M., Descombes, P., Briggs, S.P., and Paszkowski, U. (2005). Comparative transcriptomics of rice reveals an ancient pattern of response to microbial colonization. *Proceedings of the National Academy of Sciences of the United States of America* 102, 8066-8070.
3. Liu, J., Maldonado-Mendoza, I.E., Lopez-Meyer, M., Cheung, F., Town, C.D., and Harrison M, J. (2007). The arbuscular mycorrhizal symbiosis is accompanied by local and systemic alterations in gene expression and an increase in disease resistance in the shoots. *The Plant Journal* 50, 529-544.
4. Liu, J., Blaylock, L., Endre, G., Cho, J., Town, C.D., VandenBosch, K., and Harrison, M.J. (2003). Transcript profiling coupled with spatial expression analyses reveals genes involved in distinct developmental stages of the arbuscular mycorrhizal symbiosis. *Plant Cell* 15, 2106-2123.
5. Smith, S.E., and Read, D.J. (2008). *Mycorrhizal Symbiosis.* (San Diego, CA: Academic Press, Inc.).
6. Tilman, D., Hill, J., and Lehman, C. (2006). Carbon-negative biofuels from low-input high-diversity grassland biomass. *Science* 314, 1598-1600.
7. van der Heijden, M.G.A., Wiemken, A., and Sanders, I.R. (2003). Different arbuscular mycorrhizal fungi alter coexistence and resource distribution between co-occurring plant. *New Phytol.* 157, 569-578.
8. van der Heijden, M.G.A., Streitwolf-Engel, R., Riedl, R., Siegrist, S., Neudecker, A., Ineichen, K., Boller, T., Wiemken, A., and Sanders, I.R. (2006). The mycorrhizal contribution to plant productivity, plant nutrition and soil structure in experimental grassland. *New Phytol.* 172, 739-752.
9. Vance, C. (2001). Symbiotic Nitrogen Fixation and Phosphorus Acquisition. Plant Nutrition in a World of Declining Renewable Resources. *Plant Physiol.* 127, 390-397.
10. Vance, C.P., Uhde-Stone, C., and Allan, D.L. (2003). Phosphorus acquisition and use: critical adaptations by plants for securing a nonrenewable resource. *New Phytol.* 157, 423-447.

147

Sucrose Transporter Genes of *Populus*: An Investigation of Their Importance as Regulators of Biomass and Carbon Partitioning in Trees

Scott A. Harding* (sharding@uga.edu), **C.J. Tsai**, Christopher J. Frost, Duo Peng, Liangjiao Xue, and Kate H.C. Tay

Warnell School of Forestry and Natural Resources, and Department of Genetics, University of Georgia

http://aspendb.uga.edu

Project Goals: see below

Sucrose is the long-distance transport form of carbon in most temperate zone plants. Its export from source organs and subsequent trafficking among various sink organs, including sites of lignocellulosic biomass accrual, depend on the activity of trans-membrane sucrose transporters (SUT).

Project Goals: We seek to learn how SUT proteins function, individually and in cooperation with one another, to facilitate:

1. **the export of sucrose from source leaves.**
2. **the retention, trafficking and utilization of sucrose for lignocellulose production in wood tissues.**

Gene profiling and network analysis as well as metabolite profiling will be used to investigate SUT function with regard to water use, constitutive defense and biomass growth/composition in *Populus*.

Basic Approach:
Transgenic manipulation of the relative expression of three *Populus SUT* genes in a tissue-specific manner will be used in combination with drought and defoliation/stem girdling treatments. The overarching plan is to vary the ratio of SUT-mediated supply to SUT-mediated trafficking and sink retrieval processes within the plant .

Rationale:
The use of wood as a lignocellulosic feedstock for bioenergy is expected to be accompanied by expanded cultivation of woody species in agriculturally marginal sites. This means that both the yield and composition of harvestable feedstocks will be subjected to environmental constraints that affect the biosynthesis and subsequent utilization of photoassimilate (sucrose). The importance of SUT as an index or marker of tree productivity remains under-investigated. One benefit of proposed research will be to obtain a thorough understanding of SUT function in the context of other more examined gene markers (*e.g.*, sucrose, starch and lignin biosynthetic pathway enzymes). We also expect to learn about the degree to which the distinctive chemical milieu of *Populus* may factor into SUT protein function. Leaves and stems of *Populus* and *Salix* are rich in defensive metabolites, the salicin-derived phenolic glycosides (PGs) that do not occur in other agriculturally important species. Salicin is a substrate and potential inhibitor of SUT proteins. Environmental constraints brought on by limiting nutritional and water status expected to be characteristic of potential bioenergy plantations are therefore likely to perturb PG homeostasis, and possibly sucrose utilization in *Populus*.

Progress:
Phylogenetic analysis of the amino acid sequences classified the *Populus* SUT family into the three major groups characteristic of other dicots. Group-1 *PtaSUT3* gene transcripts were localized to leaf vascular traces and stem developing xylem; Group-4 *PtaSUT 4* to leaf spongy mesophylls, stem cambium, developing xylem and phloem; Group-2 *PtaSUT5/6* to all leaf cells, stem developing xylem and phloem fibers. Based on these data, PtaSUT4 is a much more important regulator of sucrose transport in *Populus* than it is in herbaceous annuals. Subcellular localization has been carried out to confirm that in contrast to other SUT proteins, PtaSUT4 is tonoplast-localized. Additional GFP-SUT fusion constructs have been assembled to investigate the subcellular localization of the other SUT proteins. SUT4-RNAi transgenic plants demonstrated a shift of biomass allocation from stem to leaf in both nitrogen (N)-replete and N-limited plants. In those plants, sucrose exhibited a complex pattern of hyper-accumulation in exporting leaves and vascular tissues of the stem, with a slight decrease in the shoot tip and sink leaves. RNAi silencing of SUT4 reduced water uptake from root tissues during drought simulation. Significant RNAi effects on secondary metabolite accumulation and on the transcript levels of carbohydrate-related genes were observed in exporting source leaves.

148

Identifying Genes Controlling Feruloylation in Grass Cell Walls

Marcia Maria de Oliveira Buanafina[1]* (mmb26@psu.edu), and Mandeep Sharma[1]

[1]The Pennsylvania State University, Biology Department, University Park

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

Cellulosic biomass—particularly from grasses—is expected to supply a large and renewable source of biomass for the production of renewable biofuels. However, utilization of grass cell walls for biofuel production is impeded greatly by ferulic acid residues which are ester linked to arabinoxylans (AX) and have the ability to form ferulate dimers functioning in cell wall cross-linking. Such cross-linking is among

the main factors inhibiting the release of fermentable carbohydrates from grasses for biofuel production.

We have shown previously that the expression of ferulic acid esterase (FAEA) in different grass species resulted in a substantial reduction in cell-wall-esterified ferulates and diferulates and increased cell wall hydrolysis. Controlling the level of total feruloylation should have a direct impact on the level of cross-linking and thereby on cell wall degradation.

Currently, the genes underlying AX feruloylation have not been identified and the isolation of such genes could be of great importance in manipulating ferulates accretion to the wall. Mutation of the feruloyl transferase gene(s) should lead to less ferulates secreted to the cell wall and reduced ferulate cross-linking.

We have developed EMS mutagenized populations of model grass species *Brachypodium distachyon* with EMS to be used as a resource for identification of the genes involved in feruloylation, synthesis of the xylan backbone and new genes that are responsible for the assembly of lignin into the cell wall. We have used spectrophotometric, microscopy, HPLC and HPIEC screening techniques to select for new genetic variation in *Brachypodium*.

EMS populations were developed from over 28,000 mutagenized seeds generating over 5,000 M2 families. A total of 12,793 plants have been screened and 1,233 have been selected. Here we report on the potential mutants with *–altered levels of cell wall ferulates, lignification and cell wall AX* – that have been selected.

We also report on the considerable variation on the level of cell wall ferulates, AX, ligninification and cellulase mediated release of sugars, among different *Brachypodium* accessions to be selected to generate our mapping population.

149

Genetic Dissection of Bioenergy Traits in Sorghum

Wilfred Vermerris[1,*] (wev@ufl.edu), Ana Saballos,[1] Terry Felderhoff,[2] Sharon Mitchell,[3] **William Rooney**,[2] **Seth Murray**,[2] **Stephen Kresovich**,[3,4] **Jeffrey F. Pedersen**,[5] **Scott Sattler**,[5] and Zhanguo Xin[6]

[1]Genetics Institute and Agronomy Department, University of Florida, Gainesville; [2]Dept. of Soil and Crop Sciences, Texas A&M University, College Station, Tex.; [3]Institute for Genomic Diversity, Cornell University, Ithaca, N.Y.; [4]Office of Research, University of South Carolina, Columbia; [5]Grain, Forage, and Bioenergy Research, USDA-ARS, Lincoln, Nebr.; and [6]USDA-ARS, Lubbock, Tex.

http://www.sorghum.genome.ufl.edu

Project Goals: The goals of this project are to gain better understanding of the genetic basis of both sugar accumulation and cell wall biosynthesis in (sweet) sorghum, in order to facilitate development of improved germplasm. This is accomplished by identifying the gene(s) underlying a quantitative trait locus (QTL) for stem sugar concentration, by identification of QTL for juice volume, by characterizing novel brown midrib mutants, and by cloning Brown midrib genes.

Sorghum is an attractive biomass crop for ethanol production because of its low water and fertilizer requirements, tolerance to heat and drought, and high biomass yield. Because of the species' great genetic diversity (Murray et al. 2009), and the fact that sorghum is a diploid, seed-propagated crop, development of cultivars and hybrids adapted to a wide range of environments is feasible. Sweet sorghums are sorghums that can reach heights of up to 6 m and that accumulate soluble sugars in their stems. After squeezing the stalks, these sugars can be fermented directly and conveniently to ethanol or other biofuels. The crushed stems (bagasse) can then be processed as lignocellulosic biomass. Sweet sorghum thus represents an ideal bridge between sugar-based and cellulosic fuels, and, given the rapid establishment of sweet sorghum, this species is expected to be of particular value in extending the processing window of sugarcane-based biorefineries (Vermerris, 2011).

In order to expand the area where sweet sorghum can be produced, both in terms of geographic location (daylength, temperature, pests and diseases) and local conditions (soil quality, water and nutrient availability), regionally adapted cultivars and hybrids need to be developed. **The goals of this project are to gain better understanding of the genetic basis of both sugar accumulation and cell wall biosynthesis, in order to facilitate development of improved germplasm.**

Quantitative trait loci (QTL) associated with sugar concentration of the juice were identified in a recombinant inbred line population derived from the sweet sorghum 'Rio' and the grain sorghum BTx623 (Murray et al. 2008). A major QTL for sugar concentration is located on chromosome 3. We are employing high-throughput transcriptome profiling using the Solexa next-generation sequencing platform to identify the gene(s) underlying this QTL. This approach relies on comparing gene expression profiles of heterogeneous inbred families that are genetically highly similar except for the region containing the QTL. RNA was extracted from several different tissues and developmental stages and expression data are in the process of being analyzed. In addition, we have mapped QTL for juice volume, using a population of recombinant inbred lines derived from the dry-stem, non-sweet grain sorghum BTx3197 and the sweet sorghum 'Rio'. Novel germplasm with an overall higher sugar yield can be developed by combining QTL (and ultimately loci) controlling juice volume and juice concentration.

In order to improve the biomass-to-fuel conversion, we are focusing on *brown midrib* (*bmr*) mutants. The *bmr* mutations change the color and the chemical composition of the vascular tissue. Four independent loci were identified by Saballos et al. (2008) in a collection of mutants first described by

Porter et al. (1978). Additional *bmr* mutants were identified in the TILLING population of Xin et al. (2008). Several *bmr* mutants from both populations have been shown to result in enhanced yields of fermentable sugars following enzymatic saccharification of sorghum biomass, even after thermochemical pretreatment (Saballos et al. 2008; Dien et al., 2009; Pedersen et al.; *in preparation*). As part of this project we have also cloned the *Bmr6* and *Bmr2* genes. The *Bmr6* gene encodes the monolignol biosynthetic gene cinnamyl alcohol dehydrogenase (CAD) (Saballos et al. 2009; Sattler et al. 2009). The *Bmr2* gene also encodes a cell wall biosynthetic enzyme (Saballos et al.; *in preparation*). Knowing the identity of the *Bmr* genes and the nature of the mutations in these genes has enabled the development of allele-specific markers that will allow more efficient use of these mutations in commercial breeding programs.

Funding from the Office of Science (BER), U.S. Department of Energy grant DE-FG02-07ER64458 is gratefully acknowledged.

References

1. Dien BS, Sarath G, Pedersen JF, Sattler SE, Chen H, Funnell-Harris DL, Nichols NN, and MA Cotta (2009) Improved sugar conversion and ethanol yield for forage sorghum (*Sorghum bicolor* (L.) Moench) lines with reduced lignin contents. Bioenerg. Res. 2: 153-164
2. Murray SC, Rooney WL, Hamblin MT, Mitchell SE and S Kresovich. 2009. Sweet sorghum diversity and association mapping for brix and height. Plant Genome 2:48-62.
3. Murray SC, Sharma A, Roone WL, Klein PE, Mullet JE, Mitchell SE and S Kresovich. 2008. Genetic improvement of sorghum as a biofuel feedstock I: quantitative loci for stem sugar and grain nonstructural carbohydrates. Crop Sci. 48:2165-2179.
4. Porter KS, Axtell JD, Lechtenberg VL, and VF Colenbrander (1978) Phenotype, fiber composition, and in vitro dry matter disappearance of chemically induced *brown midrib* (*bmr*) mutants of sorghum. Crop Sci. 18: 205-209
5. Saballos A, Ejeta G, Sanchez E, Kang CH, and W Vermerris. 2009. A genomewide analysis of the cinnamyl alcohol dehydrogenase family in sorghum [*Sorghum bicolor* (L.) Moench] identifies *SbCAD2* as the *Brown midrib6* gene. Genetics 181: 783-795.
6. Saballos A, Vermerris W, Rivera L, and G Ejeta. 2008. Allelic association, chemical characterization and saccharification properties of *brown midrib* mutants of sorghum (*Sorghum bicolor* (L.) Moench). BioEnerg. Res. 2: 193-204
7. Sattler S, Saathoff AJ, Haas EJ, Palmer NA, Funnell-Harris DL, Sarath G, and JF Pedersen. 2009 A nonsense mutation in a *cinnamyl alcohol dehydrogenase* gene is responsible for the sorghum *brown midrib 6* phenotype. Plant Physiol. 150:584-595.
8. Vermerris W (2011) Survey of genomics approaches to improve bioenergy traits in maize, sorghum and sugarcane. J. Integr. Plant Biol. 53: 105-116
9. Xin Z., Wang ML., Barkley NA, Burow G, Franks C, Pederson G, and J Burke. 2008 Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. BMC Biol. 8: 103

150

Linkage and QTL Mapping of Switchgrass (*Panicum virgatum* L.) Using EST-Microsatellites and Their Use for Comparison Within the Poaceae

Christian Tobias[1]* (christian.tobias@ars.usda.gov), Miki Okada,[1] Hugh Young,[1] Christina Lanzatella-Craig,[1] Brindha Narasimhamoorthy,[2] Malay Saha,[2] Guifen Li,[2] Rongling Wu,[3] and Joe Bouton[2]

[1]USDA-ARS Western Regional Research Center, Albany, Calif.; [2]Samuel Roberts Noble Foundation, Inc., Forage Improvement Division, Ardmore, Okla.; and [3]Department of Public Health Sciences, Pennsylvania State University, Hershey

Project Goals: 1. Develop a comprehensive molecular marker system for switchgrass. 2. Develop a saturated microsatellite linkage map of switchgrass. 3. Identification of markers associated with biomass yield, digestibility, maturity, and other traits and identification of associated markers.

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

To determine marker/QTL linkage phase with certainty during mapping, a four grandparent three-generation population using parents and grandparents from both upland and lowland ecotypes has been created consisting of 188 individuals. These were evaluated over two seasons at two locations in Oklahoma following an R-256 honeycomb design. Continuous variations for biomass and related traits were observed in the population. Plant height varied between 129 – 268 cm with the mean of 201 cm. The most productive plant produced 5.23 kg of fresh biomass at the second year harvest. Striking variations were observed for both Ca and K content. Dry matter digestibility ranged between 28.2 - 48.8% with the mean of 37.6%. QTLs associated with the traits of interest will be detected through genotyping with EST-SSR, STS, and genomic SSR markers.

As part of the larger goals of the funded project we used a full-sib mapping population derived from one pair of grandparents in a pseudo-backcross mapping strategy. This

resulted in linkage maps from each individual grandparent that were found to be highly collinear. The male and female framework map lengths and number of both framework and accessory markers were 1376 cM and 563 in the female map and 1645 cM and 542 in the male map, with 97% of the genome estimated to be within 10 cM of a mapped marker in both maps. Consistent with previous cytological data and existing initial linkage studies, there was nearly exclusive preferential pairing. The map resulting from the pollen donor was more affected by transmission ratio distortion than the female parent. Extensive collinearity with sorghum was apparent with the differences in base chromosome number appearing to result from the fusion of sorghum chromosomes 8 and 9. Sub-genome comparisons within the nine switchgrass homology groups as well as interspecific comparisons are now possible using CMAP (www.gramene.org) and all raw mapping data has been published as supplemental data.

A point that remains unclear is the origin of polyploidy in switchgrass. Autopolyploids originate through whole-genome duplication events while allopolyploids arise through hybridization of distinctly different species. Though we have thus far seen evidence for disomic inheritance this does not provide conclusive evidence supporting either mode of origin. To provide further support for one or the other modes of origin, we have performed fluorescence *in-situ* hybridization (FISH) using centromere-specific repeats, and 45S rDNA sequences in dihaploid individuals. These individuals appear to have lost one copy of each subgenome and at meiosis 18 univalents were observed. The plants are functionally sterile, though partially fertile tillers have arisen that have undergone at least partial chromosome doubling. These dihaploids, represent simplified genetic systems for karyotyping and will provide a useful reference genome allowing direct comparisons of subgenome differentiation. Our attempts to unambiguously discriminate individual chromosomes in these dihaploids will be presented.

151

Molecular Mechanisms of Carbon Partitioning by *cpg13* in Poplar

Matias Kirst* (mkirst@ufl.edu) and **Gary F. Peter**

School of Forest Resources and Conservation, University of Florida Genetics Institute, Gainesville

Project Goals: We have identified the *cpg13* gene (carbon partitioning and growth in LG13) as a key regulator of carbon partitioning to lignin and cellulose, and whole plant biomass productivity in a segregating *Populus* hybrid population. We hypothesize that *cpg13* regulates the metabolic competition for carbon, affecting growth, cellulose biosynthesis and lignification. To verify this hypothesis and begin addressing the molecular function of *cpg13*, our goals are to: a. Characterize *cpg13* function in the regulation of gene expression, metabolites, and cell wall chemistry and structure, and b. Determine the spatial and temporal expression and subcellular localization of *cpg13* to advance our understanding of its molecular role.

Lignin is one of the primary obstacles for efficient bioconversion of wood cellulosics to renewable fuels, and its content is directly and inversely proportional to fermentable sugar yield. In woody species, lignin content is often negatively correlated with biomass productivity. This relationship has been observed in segregating populations of *Populus* and *Eucalyptus*, and in breeding populations and a natural mutant of loblolly pine. In a previous project funded by the Department of Energy ("Genomic Mechanisms of Carbon Allocation and Partitioning in Poplar" – ER64114-1026645-0011741) we identified a major gene involved in the regulation of carbon partitioning to lignin and cellulose, and whole plant biomass productivity. The gene, referred hereafter as *cpg13* (carbon partitioning and growth locus in chromosome XIII), was identified by applying a genetical genomic approach that combines fine-mapping of quantitative trait loci and transcriptome analysis of a *Populus deltoides* and *Populus trichocarpa* hybrid population. However, the functional role of *cpg13* is unknown.

The primary goal of this project is to dissect the molecular mechanism by which *cpg13* regulates carbon partitioning and growth. To uncover the mechanism of carbon regulation we are: (1) characterizing *cpg13* function in the regulation of gene expression, metabolites, and cell wall chemistry and structure, and (2) determining the spatial and temporal expression and subcellular localization of *cpg13* to advance our understanding of its molecular role.

To achieve these goals we have generated 20 transgenic lines, over-expressing (OE) and down-regulating the expression of *cpg13*. For down-regulating *cpg13* we developed two RNAi constructs, that target (1) the short N-terminus and (2) the C-terminus, which contains a domain of unknown function (DUF). No transgenic lines could be recovered for the second construct, suggesting that it interferes with expression of other genes that contain the DUF domain. Three to five biological replicates were grown in a greenhouse for all viable transgenic lines during 2010. Instead of proceeding to immediately characterize the transcriptome, metabolome, and wood growth and properties, as originally planned, we chose to narrow the selection to 5 OE and RNAi expressing lines. These lines are now being cloned to obtain 20+ biological replicates, for precise estimates of plant biomass growth properties.

In addition to the analysis of plants over-expressing and down-regulating the expression of *cpg13* we are taking two complementary approaches to determine what cell types express *cpg13* and its subcellular localization. In the first approach, green fluorescent protein (GFP) has been fused in frame at its carboxy terminus to *cpg13* and its native promoter, and the construct has been transformed into poplar. The transgenic plants are being used to analyze expression in different organs, cell types and ontogenetic stages. Overall, initial observations indicate the expression of *cpg13* in lignifying vascular tissues (Figure 1). In the second approach, we are developing antibodies that specifically recognize *cpg13* and use them to confirm protein abundance,

subcellular localization by transmission electron microscopy and biochemical fractionation studies. *Cpg13* has been expressed and efforts are currently under way to purify it.

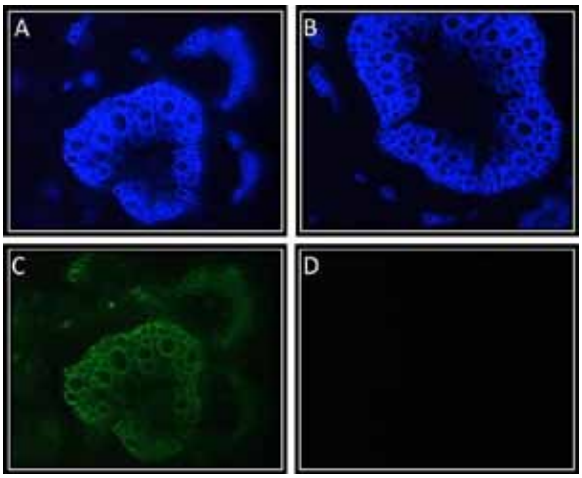


Figure 1. Lignin localization and expression of cpg*13* in poplar petioles (30 micrometer) . Lignin UV-Auto-Fluorescence in (A) cpg13::cpg13:GFP (B) wild-type. GFP fluorescence in (C) cpg13::cpg13:GFP (D) wild-type (negative control). Magnification at 20× and exposure time of 3.54s.

152

The Role of Small RNA in Biomass Deposition and Perenniality in Andropogoneae Feedstocks

Matthew Hudson* (mhudson@illinois.edu)

University of Illinois

Project Goals: We have already shown that a large number of new small RNA species are expressed in *Miscanthus*. We will investigate the extent to which these species target genes known to be involved in biomass production, and how much they contribute to development and variability of biomass. In particular, we will identify small RNA elements that regulate deposition of the cell wall of cells in the plant stem, a critical process in biomass production. We will also investigate the involvement of small RNA in processes that underlie traits critical to regional biomass production, for example flowering and maturity time and overwintering.

The small RNA (sRNA) transcriptome is a complex network that regulates gene expression, antiviral responses and chromatin remodeling, and controls the self-spread of mobile elements. We employed Illumina deep sequencing of sRNAs, mRNAs, and sRNA-mediated cleavage products of mRNA (the acmRNA or degradome) to probe the role of sRNA in growth, development and biomass deposition in *Miscanthus X giganteus* (Mxg). This perennial allo-triploid grass species has significant potential as a bioenergy crop. Using the genomes of *Sorghum*, maize, and rice as references, we predicted 120 new miRNA families from Mxg that are also conserved in other grasses, and identified 61 Mxg orthologs of known miRNA from other species. For 261 mRNAs computationally predicted to be targeted by miRNAs, miRNA-directed cleavage was observed experimentally. This result validates targets of many the newly discovered miRNAs and shows some mRNAs are cleaved by multiple miRNAs. The majority of miRNAs were differentially expressed to a statistically significant degree among the sampled tissues or organs. In several cases strong negative correlation was observed between miRNA and target mRNA expression patterns, likely indicating a predominant role for the miRNA in tissue-specific mRNA expression. The majority of new Mxg miRNAs were present at highest levels in the inflorescence and rhizome tissues that contribute to a perennial growth habit. Similarly, many miRNAs identified in stem tissues are associated with the regulation of cell wall biology, targeting members of the expansin, β-tubulin, and callose synthase gene families in a tissue-specific manner. Our results suggest the existence of several previously-undescribed miRNAs that control expression of genes likely contributing to perenniality and biomass deposition in grass species.

153

Development of Genomic Tools to Improve Prairie Cordgrass (*Spartina pectinata*), a Highly Productive Bioenergy Feedstock Crop

Jose L. Gonzalez, Kristene Gedye, Vance Owens, and Arvid Boe

South Dakota State University

Project Goals: Development of SSR markers in prairie cordgrass and initiate linkage map development

Prairie cordgrass (*Spartina pectinata*) is a tall (180-250 cm) robust rhizomatous perennial grass native to the prairies of North America, grows well in a wide range of conditions, including wet and dry marginal lands, as well as salty soils. Natural populations of PCG can be found as far north as 60°N, making this species ideal for cultivation in the Great Plains of North America. Prairie cordgrass is a C4 species with a wide ecological amplitude especially acclimated to low temperatures allowing early growth in the spring. The breeding efforts in this species have been only recently developed. With the goal of developing a molecular breeding tools we have identified over 1,000 SSR loci derived from both genomic and EST sequences.

EST sequencing has yielded the first transcriptome for this species consisting of 26,302 contigs derived from 454 reads. Selected SSR markers have been used to develop an initial linkage map using a population of 94 individuals. Other genetic and genomic resources are being developed including a clonal germplasm collection, BAC library, additional ESTs sequencing and tissue culture and genetic transformation protocols.

154

Transcriptome Profiling in *Brachypodium distachyon*

Todd Mockler * (tmockler@cgrb.oregonstate.edu)

Oregon State University

Project Goals: The project goals were to design a whole genome DNA microarray for the model grass *Brachypodium distachyon* and to use the array platform for expression profiling in *Brachypodium* under a variety of environmental conditions and in various tissues and developmental stages.

Brachypodium distachyon is the premier experimental model grass platform and is related to candidate feedstock crops for bioethanol production. Based on the recent DOE-JGI *Brachypodium* Bd21 genome sequence and annotation we designed a whole genome DNA microarray platform. The quality of this array platform is unprecedented due to the exceptional quality of the *Brachypodium* genome assembly and annotation and the stringent probe selection criteria employed in the design. We worked with members of the international community and the bioinformatics/design team at Affymetrix at all stages in the development of the array. We have used the *Brachypodium* arrays to interrogate the transcriptomes of plants grown in a variety of environmental conditions including diurnal and circadian light/temperature conditions and a variety of light qualities. We have also examined the transciptional responses of *Brachypodium* seedlings subjected to various abiotic stresses including heat, cold, salt, and high intensity light. We are also generating a gene expression atlas representing various organs and developmental stages. The results of these efforts including all microarray datasets are available through the BrachyBase.org community genome and annotation database.

155

The Biofuel Feedstock Genomics Resource—A Comparative Gene and Transcript Annotation Database for Lignocellulosic Feedstock Species

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

Plant Biology Department, Michigan State University, East Lansing, Mich.

Project Goals: Complete genome sequences are only available for two lignocellulose biofuel species: poplar and maize. Limited amounts of transcript sequence are available for many biofuel species, but those resources are also incomplete. Researchers would benefit from comparative analyses between their genes-of-interest in their species-of-interest with sequence resources from many other species. Such a resource is not available for biofuel species. The goals of this project are to provide a comparative resource where gene and transcript sequences from model species and biofuel species are thoroughly and consistently annotated and where those sequences are related to each other in numerous ways. The success of the project depends on the ability of a researcher to move easily between sequences from multiple species that are related to each other based on functional annotation and/or sequence similarity.

Candidate plant species to be used for lignocellulosic ethanol production include a large number of species within the Poaceae, Pinaceae, and Salicaceae families. For these biofuel feedstock species, there are variable amounts of genome sequence resources available, ranging from complete genome sequences (e.g. *Populus trichocarpa*, *Zea mays*) to transcriptome data sets in the form of Expressed Sequence Tags (e.g. *Pinus glauca*, *Panicum virgatum*). Some species with importance to biofuel feedstock research have negligible sequence resources. While obtaining genome or transcript sequence is the initial step in a genomics-based approach to biological research, the more challenging step in genomics is the process of understanding gene function and how genes and their products confer the underlying processes/traits in plant biology. This challenge is mostly attributable to two issues: a large percentage of genes within genomes have no known function and experimental approaches to determining gene function on a per gene basis are fiscally prohibitive. One method to improve our understanding of gene function is through comparative approaches in which sequence similarity is used to cross-annotate orthologs and paralogs thereby leveraging all available functional annotation data to improve the annotation of genes in species with limited annotation data. For the current release of the Biofuel Feedstock Genomics Resource (http://bfgr.plantbiology.msu.edu/), we have created a comprehensive, uniform, well annotated resource for data-mining genomic data for biofuel feedstock species. To augment comparative analyses, the predicted genes from seven sequenced plant genomes and the predicted transcriptomes from 44 species, including biofuel feedstock target species and their phylogenetic relatives, are annotated within the database. All sequences in the resource have been aligned to Uniref proteins, InterPro domains, KEGG Orthologs, as well as the predicted gene sequences from fully sequence plant species. Orthologous gene groups have been identified in order to allow users to easily identify orthologous and paralogous relationships between their genes of interest and other sequences in the resource and in order to aid in cross-referencing to sequences from other species. Numerous search functions are provided to allow users to find sequences based on sequence alignments, functional annotation, sequence identifiers, homology to KEGG orthologs, InterPro domain names and GO terms. Additionally, sequences can be searched for SNP and SSR markers.