Speaker Abstracts

Joint Meeting 2011

U.S. Department of Energy
Genomic Science Awardee Meeting IX

and

U.S. Department of Agriculture–DOE
Plant Feedstock Genomics for Bioenergy Awardee Meeting

April 10–13, 2011

Hyatt Regency Crystal City in Arlington, Virginia

United States Department of Agriculture

National Institute of Food and Agriculture

U.S. DEPARTMENT OF ENERGY | Office of Science

Both the speaker and poster abstracts for this meeting are available at genomicscience.energy.gov.

Table of Contents

Table of Contents

Table of Contents

Monday, April 11, 9:00 a.m.–9:30 a.m.
Plenary Session: DOE Bioenergy Research Centers

Optimizing Plant Cell Walls for Efficient Biofuel Production

Henrik Vibe Scheller

Feedstocks Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, California; Department of Plant and Microbial Biology, University of California, Berkeley

Producing biofuels from biomass requires decomposition of the polymers, which mainly are composed of polysaccharides and lignin in plant cell walls. These polymers are recalcitrant to degradation, and some degradation products cannot be converted efficiently into fuels or may even be inhibitory. Despite rather detailed information on the structure of cell wall polysaccharides, little is known about their biosynthesis. The key enzymes in this process are glycosyltransferases (GTs); plants need hundreds of GTs and other transferases, hydrolases, and transporters to synthesize the complex polysaccharides present in cell walls. A better understanding of polysaccharide biosynthesis will enable development of crops with improved properties as biofuel feedstocks. The Joint BioEnergy Institute has identified several enzymes involved in xylan synthesis, the major hemicellulose in lignocellulosic biomass. By altering the expression of these enzymes, we have obtained plants with improved saccharification properties and lower content of inhibitory compounds such as acetate and ferulate esters. Hemicelluloses and lignin have important roles in the plant, and their biosynthesis cannot simply be blocked without adverse effects on plant growth. We therefore have developed methods to spatially and temporally fine-tune the deposition of xylan and lignin.

Monday, April 11, 9:30 a.m.–10:00 a.m.
Plenary Session: DOE Bioenergy Research Centers

Gene Targets for Improving Biomass Deconstruction and Fermentation in Switchgrass

Richard A. Dixon

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

Switchgrass is a C_4 perennial forage grass native to most areas of the North American grasslands. It has been proposed as a major perennial bioenergy feedstock in the United States because it is widely adapted, has high biomass production, and is efficient in C_4 photosynthesis and water and nitrogen use. Alamo switchgrass was evaluated, selected, and released in 1978 by the Texas Agricultural Experiment Station as a superior ecotype for conservation performance, forage yield, and seed production, performing well from clays to fine sands.

The DOE BioEnergy Science Center (BESC) is developing switchgrass as a target for genetic modification to improve biomass yields and saccharification efficiency for production of fermentable liquid biofuels. Approaches involve development of extensive genetic and genomic resources for the Alamo variety, high-throughput gene discovery platforms, in-depth biochemical analyses, and genetic transformation strategies. These novel resources will be described, and case studies provided to show how new knowledge generated is being translated into improved switchgrass varieties that will facilitate consolidated bioprocessing for liquid fuels. Specific emphasis is placed on modification of lignin content through the genetic manipulation of biosynthetic pathway genes or transcription factors.

Monday, April 11, 10:30 a.m.–11:00 a.m.
Plenary Session: DOE Bioenergy Research Centers

Understanding and Manipulating Hemicellulose Biosynthesis

Kenneth Keegstra,* David Cavalier, Linda Danhof, Jonathan Davis, Benjamin Fode, Jacob Jensen, Barbara Reca, Yan Wang, and Curtis Wilkerson

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

*Speaker

Hemicellulosic polysaccharides constitute a significant portion of plant biomass and need to be fully utilized to make biofuel production sustainable. Although the hemicellulosic fraction of most common forms of biomass consists mainly of xylan polymers containing pentose monomers, some biomass contains hemicellulosic polysaccharides that are rich in hexoses. An important prelude to manipulating hemicellulose quality and quantity is to understand both the biochemical details of hemicelluose biosynthesis as well as the regulatory events that control hemicellulose synthesis and deposition. Considerable progress has been made in recent years in identifying the genes and proteins responsible for hemicellulose biosynthesis. The cellulose synthase-like (CSL) proteins are involved in making the backbone of many hemicellulosic polysaccharides, while several classes of glycosyltransferases are required for addition of side chain residues. Recent progress will be reviewed by providing an overview of the strategies used to study this problem and the major conclusions that have resulted. Many important issues, however, still need to be elucidated, especially related to the regulation of hemicellulose biosynthesis. Some of these issues and our approaches to resolving them will be presented.

Monday, April 11, 11:00 a.m.–12:00 p.m.
Keynote Presentation

Symbiosis: The Once and Future Sustainability?

Sharon R. Long (SRL@stanford.edu)

Department of Biology, Stanford University

cmgm.stanford.edu/biology/long/

The symbiosis of rhizobia with legumes results in bacterial nitrogen fixation that can provide the plant with reduced N for protein synthesis. The symbiosis takes place in root nodules, which are multitissue organs built through signal exchange, morphogenesis, and molecular differentiation. This symbiosis historically has been recognized as a source of plant nutrition and employed in crop rotation, co-cropping, and other uses.

Through molecular genetic, biochemical, and cell biological approaches, many details about early stages of nodulation have been revealed. The plant and bacteria exchange plant-produced flavonoids and bacterially synthesized nod factors (NF) that act as species-specific molecular signals. Live bacteria or pure NF result in a suite of early (~24 h) plant behaviors: root hair deformation; ion flux and calcium spiking; and mRNA expression of nodule-specific proteins (ENODs).

Forward genetic screens of host plants yielded mutants corresponding to receptors, ion channels, nucleoporin, a calcium/calmodulin-dependent protein kinase, and several transcription factors. Identifying these loci allows detailed study of the proteins and their activities. We used reverse genetics to show a role in symbiosis for flotillins, membrane-associated proteins thought to mediate receptor action in animal cells. We have now shown that LYK3, a plant receptor for bacterial NF signals, shows increased co-localization with a flotillin in root hairs exposed to bacteria compared to untreated root hairs. This may yield insights into the mechanism of infection.

Nodulation is a multistep process, and both plant and bacterial genes are required through the final stages of differentiation and nitrogen fixation. Forward genetic screens in plants resulted in a series of mutants we called defective in nitrogen fixation (dnf). So far, we have cloned two of these proteins; each appears to play a role in signaling to the bacteria and in supporting bacterial viability and gene expression. One of them, DNF1, encodes a signal peptidase required to process plant-encoded proteins that probably act as signals for bacterial differentiation. A second, DNF2, encodes a homology of bacterial phospholipase C.

These molecular details suggest many questions about the evolution of symbiosis. One of the most noteworthy findings about plant genes for NF signal transduction was the discovery that at least three of them are also used for mycorrhizal symbiosis. This implies that rhizobial symbioses may have evolved as duplications and specializations of the ancient and more broadly distributed mycorrhizal associations. Does this suggest pathways to engineer symbiosis in other organisms? It may be possible to construct transgenics that display some early responses to rhizobial signals. In my view, it is less certain that we could transfer to nonlegumes more complex processes, such as infection and C/N root-shoot resource allocation, that are required for effective nourishment of a plant by the rhizobial symbionts. Also, would N-fixing monoculture be a good idea? These questions remain to be discussed.

Monday, April 11, 2:10 p.m.–2:30 p.m.
Breakout Session A: Plant Feedstocks Genomics for Bioenergy

Poplar Biomass Protein-Protein Interactions

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]Department of Horticulture, Virginia Polytechnic Institute and State University, Blacksburg, Va.; [2]Department of Forest Resources and Environmental Conservation, Virginia Tech; and [3]Virginia Bioinformatics Institute at Virginia Tech

*Speaker

xylome.vbi.vt.edu

Proteins are molecular machines 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 (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 (323 DB x 335 AD) 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 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.

Monday, April 11, 2:30 p.m.–2:50 p.m.
Breakout Session A: Plant Feedstocks Genomics for Bioenergy

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

Edward S. Buckler (esb33@cornell.edu),[1*] **Michael D. Casler**,[2] and **Jerome H. Cherney**[3]

Co-authors: Denise Costich,[1] Michael D. Casler,[2] Jerome H. Cherney,[3] Elhan Ersoz,[4] Moria Sheehan,[5]; Fei Lu,[6] Alexandra Lipka,[6] and Edward S. Buckler[1]

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

*Speaker

www.maizegenetics.net

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 5 or more years between cycles of advancement, while genomic selection permits advancement every time a cross can be made, for example, every 1 to 2 years. To enable genomic selection and genome-wide association analysis in switchgrass, we have developed germplasm resources from upland switchgrass, which include 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 strongly favors using the tetraploids. Finally, genotyping-by-sequencing approaches have been applied to switchgrass 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.

Monday, April 11, 2:50 p.m.–3:10 p.m.
Breakout Session A: Plant Feedstocks Genomics for Bioenergy

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; [3]Department of Plant Pathology, University of California-Davis

*Speaker

Cell wall synthesis is subjected to precise temporal and spatial regulation. Multiple components including, for example, cellulose, hemicellulose, lignin, pectin, and proteins have to be synthesized and deposited coordinately in the cell wall. Therefore, it is conceivable that some global regulators must be present 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 are 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 the cell wall.

Monday, April 11, 3:30 p.m.–3:50 p.m.
Breakout Session A: Plant Feedstocks Genomics for Bioenergy

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; [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, U.S. Department of Agriculture's Agricultural Research Service (USDA-ARS), Lincoln, Neb.; and [6]USDA-ARS, Lubbock, Texas.

*Speaker

www.sorghum.genome.ufl.edu

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, developing cultivars and hybrids adapted to a wide range of environments is feasible. Sweet sorghums can reach heights of up to 6 m and 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 (day length, temperature, pests, and diseases) and local conditions (soil quality, water, and nutrient availability), regionally adapted cultivars and hybrids need to be developed. **The goal of this project is to better understand 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, nonsweet 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.

To improve the biomass-to-fuel conversion, we are focusing on *brown midrib* (*bmr*) mutants. The *bmr* mutations change the color and 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, B. S., et al. 2009. "Improved Sugar Conversion and Ethanol Yield for Forage Sorghum (*Sorghum bicolor* (L.) Moench) Lines with Reduced Lignin Contents," BioEnergy Research 2, 153–64.
2. Murray, S.C, et al. 2009. "Sweet Sorghum Diversity and Association Mapping for Brix and Height," *Plant Genome* **2**, 48–62.
3. Murray, S. C., et al. 2008. "Genetic Improvement of Sorghum as a Biofuel Feedstock I: Quantitative Loci for Stem Sugar and Grain Nonstructural Carbohydrates," *Crop Science* **48**, 2165–79.
4. Porter, K.S., et al. 1978. "Phenotype, Fiber Composition, and *in vitro* Dry Matter Disappearance of Chemically Induced *brown midrib* (*bmr*) Mutants of Sorghum," *Crop Science* **18**, 205–9.
5. Saballos, A., et al. 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**, 783–95.
6. Saballos, A., et al. 2008. "Allelic Association, Chemical Characterization and Saccharification Properties of *brown midrib* Mutants of Sorghum (*Sorghum bicolor* (L.) Moench)," *BioEnergy Research* **2**, 193–204.
7. Sattler, S., et al. 2009. "A Nonsense Mutation in a Cinnamyl Alcohol Dehydrogenase Gene is Responsible for the Sorghum *brown midrib6* Phenotype," *Plant Physiology* **150**, 584–95.
8. Vermerris, W. 2011. "Survey of Genomics Approaches to Improve Bioenergy Traits in Maize, Sorghum, and Sugarcane," *Journal of Integrative Plant Biology* **53**, 105–19.
9. Xin, Z., et al. 2008. "Applying Genotyping (TILLING) and Phenotyping Analyses to Elucidate Gene Function in a Chemically Induced Sorghum Mutant Population," *BMC Biology* **8**, 103.

Monday, April 11, 3:50 p.m.–4:10 p.m.
Breakout Session A: Plant Feedstocks Genomics for Bioenergy

Identifying Genes and Networks for Increasing Biomass Production in New Energy Grasses by Using Rice as a Model System

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

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

*Speaker

Meeting the energy demands of the future using "new energy" crops as sources for biomass will require significant improvement in agronomic and biomass productivity traits. This targeted improvement, either by traditional breeding or biotechnology-based strategies, requires identification of genetic variation as sources of high biomass traits. Because the genetic resources and genomic tools necessary to enhance feedstock traits of target crops such as switchgrass are limited, we exploited the genetic and genomic tools readily available in rice to establish a pipeline for trait and gene discovery. To date, this rice-focused pipeline has yielded candidate genes, candidate QTL, mutant lines that promise more gene candidates, and a deep understanding of morphological and physiological traits related to biomass. In addition, our data point to simple screening assays relevant to biomass accumulation and cell wall composition. Integration of our data to establish causal links between variation at the sequence level of genes (or genetic regions, in the case of multigenic traits) and phenotype (morphological and physiological signatures of variation in biomass accumulation and composition) in rice is now guiding strategies for validating candidate gene function in switchgrass.

Monday, April 11, 4:10 p.m.–4:30 p.m.
Breakout Session A: Plant Feedstocks Genomics for Bioenergy

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

Kevin L. Childs and **C. Robin Buell***

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

*Speaker

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, variable amounts of genome sequence resources are available, ranging from complete genome sequences (e.g., for *Populus trichocarpa* and *Zea mays*) to transcriptomic datasets in the form of expressed sequence tags (e.g., *Pinus glauca* and *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 and 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 for 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 (bfgr.plantbiology.msu.edu), we have created a comprehensive, uniform, and well-annotated resource for 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, and the predicted gene sequences from fully sequenced plant species. Orthologous gene groups have been identified to allow users to easily identify orthologous and paralogous relationships between their genes of interest and other sequences in the resource and 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 Gene Ontology terms. Additionally, sequences can be searched for SNP and SSR markers.

Monday, April 11, 4:30 p.m.–4:50 p.m.
Breakout Session A: Plant Feedstocks Genomics for Bioenergy

Genetic and Genomic Resources in Prairie Cordgrass

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

South Dakota State University

*Speaker

Prairie cordgrass (*Spartina pectinata*) is a tall (180 to 250 cm) robust rhizomatous perennial grass native to the prairies of North America. It grows well in a wide range of conditions, including wet and dry marginal lands and salty soils. Natural populations of prairie cordgrass can be found as far north as 60˚N, making this species ideal for cultivation in the Great Plains. Prairie cordgrass is a C_4 species with a wide ecological amplitude especially acclimated to low temperatures, allowing early growth in the spring. Breeding efforts in this species have been developed only recently. With the goal of developing 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.

Monday, April 11, 2:10 p.m.–2:40 p.m.
Breakout Session B: Metabolic Pathway Analysis

Metabolic Potential of Cyanobacteria for Advanced Biofuel Production

Himadri Pakrasi

Washington University

Cyanobacteria constitute a vastly diverse group of oxygenic photosynthetic prokaryotes with metabolic versatility that allows them to adapt to a wide range of habitats. These microbes have attracted interest as catalysts for the conversion of solar energy into a variety of valuable products. A quest for the identification of novel pathways has initiated large-scale cyanobacterial genome sequencing projects. Integrated systems-level studies composed of *in silico* modeling; global transcriptomic, proteomic, and metabolomic analyses; and genetic engineering efforts have established the viability of the cyanobacterial platform for producing various advanced fuel molecules. Much exciting work has been done in the past 5 years demonstrating that cyanobacteria can produce desirable industrial platform chemicals like hydrogen, alkanes, fatty acids, butanol/iso-butyraldehyde, and isoprenes. Our recent study demonstrating high rates of aerobic hydrogen production by a native, wild type *Cyanothece* strain by employing appropriate physiological perturbations calls to attention the potential of naturally evolved metabolic capabilities in various cyanobacterial strains. The availability of six additional *Cyanothece* genomes has revealed diverse metabolic capabilities of this group of unicellular diazotrophic cyanobacteria that include various fermentative and anaerobic pathways, unique nitrogen and phosphorous metabolism pathways, as well as butanol and alkane production pathways. Finally, for such synthetic biology experiments in cyanobacterial hosts, efforts are under way to develop controllable promoter systems that will allow high levels of product formation in engineered strains without compromising growth and while simultaneously avoiding product toxicity issues.

Monday, April 11, 2:40 p.m.–3:10 p.m.
Breakout Session B: Metabolic Pathway Analysis

Design, Evolution, and Ensemble Modeling of High Flux Microbial Synthesis Networks

James C. Liao

University of California, Los Angeles–DOE Institute of Genomics and Proteomics; DOE BioEnergy Science Center

Higher alcohols possess several characteristics that make them attractive as biofuels and chemical feedstock. We have designed a metabolic network to synthesize these molecules in high yield and further developed strategies to produce them from various carbon sources including cellulose, CO_2, and proteins. These metabolic engineering successes largely come from rational designs based on biochemical knowledge and insight.

For high-flux production, a quantitative method is needed to fine-tune the metabolic network. Current approaches address this problem either at the stoichiometric level or require detailed enzyme kinetics. We developed a high-throughput modeling strategy called "ensemble modeling," which allows the use of typical end-point data (e.g., product titers) to construct ODE-based dynamic models and circumvents the need for detailed kinetic parameters. This strategy screens all allowable kinetic models based on product titers of gene knockout strains and enzyme-overexpression modifications. The sets of models converge rather quickly to describe similar phenotypes and thus predict results of further genetic modifications. Because these models are screened based on typical metabolic engineering data, they do not require specialized experiments for validation and can learn from the data.

Monday, April 11, 3:30 p.m.–4:00 p.m.
Breakout Session B: Metabolic Pathway Analysis

Using Computations to Facilitate Metabolic Reconstructions and Guide Strain Optimization

Costas D. Maranas

Department of Chemical Engineering, Pennsylvania State University

In this presentation, we will introduce the genome-scale model *Zea mays* iRS1563 for maize that contains 1,563 genes and 1,825 metabolites involved in 1,985 reactions from primary and secondary maize metabolism. For approximately 42% of the reactions, we found direct literature evidence for the participation of the reaction in maize. As many as 674 metabolites and 893 reactions present in *Z. mays* iRS1563 are not accounted for in maize C4GEM. All reactions are elementally and charged balanced, linked with GPR associations, and localized in six different compartments (i.e., cytoplasm, mitochondrion, plastid, peroxisome, vacuole, and extracellular). *Z. mays* iRS1563 corresponds to the largest and most complete effort to date at cataloguing metabolism for a plant species. In addition, we will describe progress in the reconstruction of metabolic models for cyanobacteria.

In the second part of the presentation, we will highlight milestones achieved from an integrated computational and experimental study (using OptForce) aimed at improving the availability of malonyl-CoA in *Escherichia coli*. By successively implementing the hierarchy of OptForce suggestions, we have successfully constructed a recombinant strain of *E. coli* that exhibits improved malonyl-CoA levels. Finally, we will overview progress in the construction of the MetRxn web resource that contains standardized metabolite and reaction entries from 8 databases and more than 30 genome-scale metabolic models. All reactions in the database are charge and elementally balanced, allowing for their direct use in model construction, curation, and simulation. As of February 2011, this dataset consists of 34,984 distinct metabolites participating in 67,944 reactions.

Monday, April 11, 4:00 p.m.–4:30 p.m.
Breakout Session B: Metabolic Pathway Analysis

Exploring Microbiomes for New Enzyme Activities and Re-Engineering Their Function

Andrzej Joachimiak

Midwest Center for Structural Genomics and Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, Ill.

Human microbiomes harbor a large number of microbial species, many of which are uncharacterized or unculturable but have a significant impact on human physiology and disease. The gut microbiome performs a diverse set of functions supplementing those performed by the host. It is believed to significantly enhance the metabolism of complex carbohydrates and other compounds and contributes a considerable fraction of host energy intake. The analysis of gut microbes with completed genomes reveals an abundance of carbohydrate-metabolizing enzymes and resulting metabolic pathways. Moreover, the frequencies of the glycoside hydrolase families differ significantly in human gut microbiomes as compared to a set of completed, nonhuman-related microbial genomes.

We have expressed and purified more than 300 glycoside hydrolase family members from human microbiome bacteria. A number of structures from normal and enteric microflora have been determined, leading to a better understanding of the metabolic pathways they enable. For example, the α-glucosidase, a member of the GH31 family, shows substrate preference for α(1–6) over α(1–4) glycosidic linkages and produces glucose from isomaltose as well as maltose. The preference can be switched by a point mutation at its active site, as predicted from structure. The re-engineered enzyme prefers α(1–4) over α(1–6) glycosidic linkage and is more active. Our results suggest a widespread adaptation to a variety of polysaccharides by intestinal microorganisms. Knowledge of carbohydrate metabolic pathways in these organisms may be exploited for new functions relevant to bioenergy applications.

Advanced structural studies, such as obtaining structures of protein complexes or more precisely identifying substrates and products of enzymatic reactions, are expected to contribute to the understanding of metabolic pathways and guide their engineering.

This work was supported by National Institutes of Health Grant GM074942, GM094585 and by the U.S. Department of Energy's Office of Biological and Environmental Research under contract DE-AC02-06CH11357.

Monday, April 11, 4:30 p.m.–5:00 p.m.
Breakout Session B: Metabolic Pathway Analysis

ELSI Pilot: Assessing and Mitigating the Risks of Large-Scale Metabolic Engineering

Nathan J. Hillson

Lawrence Berkeley National Laboratory

The Advanced Biofuels Process Development Unit (ABPDU)—funded by the Department of Energy's Office of Energy Efficiency and Renewable Energy—will come online in spring 2011 and house two 300-liter microbial fermentation tanks. In the extremely unlikely event that ABPDU's post-fermentation microbicidal protocol (e.g., base treatment and neutralization) should catastrophically fail, broth that harbors viable genetically engineered microorganisms could be purged directly to downstream waste-water treatment processes. This ethical, legal, and social implications (ELSI) pilot study seeks to quantitate the risks associated with such a scenario by measuring the viability of the engineered microbes (and, perhaps more importantly, their embedded genes) in mock sewage reactors that mimic the conditions and microbial communities found in real-world waste-water treatment plants. Furthermore, investigation of how differential genetic backgrounds (e.g., gene deletions) impact survival and gene transmission to sewage sludge communities will guide subsequent forward-engineering efforts to further reduce risk. This pilot study establishes methodologies (leveraging only recently available technologies) for assessing and mitigating the risks of future large-scale metabolic engineering microbial projects, including those extending beyond the bioreactor.

Monday, April 11, 2:00 p.m.–2:20 p.m.
Breakout Session C: Databases and Functional Annotation

Improving the Consistency and Completeness of Microbial Genome Annotations in an Integrated Database Context

Victor M. Markowitz

DOE Joint Genome Institute; Computational Research Division, Lawrence Berkeley National Laboratory

Comparative analysis of genomic and metagenomic data is critical for their biological interpretation. The success of comparative analysis depends on the efficiency of genome sequence data integration. It also depends on the consistency and completeness of the integrated data, which in turn are determined by the quality of annotations and the level of detail in metabolic pathway characterization and phenotype prediction.

The Integrated Microbial Genomes (IMG) family of data management and analysis systems enables scientists to study microbial genomes (IMG: img.jgi.doe.gov) and microbial community aggregate metagenomes (IMG/M: img.jgi.doe.gov/m) in the integrated context of a rapidly expanding universe of genomic and metagenomic datasets. The quality of genome annotations in systems such as IMG depends largely on the original dataset that providers and various functional resources employed in the functional characterization of genomes, with errors often carried over into public resources and difficult to correct.

We will present the mechanisms and tools provided by the IMG systems for improving the consistency and completeness of microbial genome annotations and the re-annotation of all microbial genomes in the context of IMG.

Monday, April 11, 2:20 p.m.–2:40 p.m.
Breakout Session C: Databases and Functional Annotation

Functional Characterization by Mapping Protein-Ligand Interactions

Danielle Corgliano, Michael Endres, Elizabeth Landorf, Sarah Zerbs, and **Frank R. Collart***

Biosciences Division, Argonne National Laboratory, Lemont, Ill.

*Speaker

Accurate functional annotation is essential to define the role of key microbial and plant proteins that mediate the response to ecosystem characteristics. We are improving experimental approaches for functional annotation by developing label-free methods that can be globally applied at the genome scale. These *in vitro* methods map protein-ligand interactions as an effective approach for functional interrogation of unknown proteins. Or, in the cases of proteins with inferred or known binding ligands, the approach can validate or extend the current biochemical and biophysical characterization. One approach uses a fluorescence-based thermal shift (FTS) assay for high-throughput screening of protein-ligand interactions. A second complementary approach applies bio-layer interferometry as a screen for protein-ligand interactions in a microwell plate–compatible format. The methods have been used to derive ligand profiles for ABC transporter-binding proteins and two-component sensory systems and to identify effector molecules and DNA-binding regions for bacterial transcriptional regulators. The sensory and transport profiles provide insight for the molecular connections between the cell and the environment. This information can be integrated with functional characterization of the transcription factors, whose activity is mediated by the environmental ligands or their metabolic derivatives. A coupling of the regulatory ligands with the DNA-binding regions of the transcription factors allows the association of metabolic pathways with the regulatory network. This improved functional characterization of protein and DNA elements for these essential biological processes will facilitate the functional design of biological systems for DOE mission needs.

This contribution originates from the Environment Sensing and Response Scientific Focus Area (SFA) program at Argonne National Laboratory. This research was supported by the U.S. Department of Energy's Office of Biological and Environmental Research as part of BER's Genomic Science program.

Monday, April 11, 2:40 p.m.–3:00 p.m.
Breakout Session C: Databases and Functional Annotation

The Use of Protein Localization Prediction in Functional Annotation: Obstacles and Opportunities

Margie Romine

Pacific Northwest National Laboratory

Prediction of protein function from sequence data alone is largely based on using (1) homology-based methods (e.g., BLAST) to identify defined domains or experimentally characterized orthologs and (2) genome-context analysis to infer function based on recurring connections to other genes and regulatory features. Although many tools have been developed for predicting the subcellular localization of proteins, they are underutilized in functional prediction, and current methods used do not sufficiently capitalize on knowledge gathered during the annotation process. In this presentation, I will provide examples of how these predictions are useful in annotation, describe the basis by which they are derived, and suggest a generalized workflow that uses different types of predictors and evidence types to provide a more informed genome-wide prediction of protein localization. I will conclude with pointing out key sources of error in predictions and what might be done to improve them.

Monday, April 11, 3:00 p.m.–3:20 p.m.
Breakout Session C: Databases and Functional Annotation

Functional Annotation of Putative Enzymes in Methanogens

John Orban

University of Maryland

Methane-producing organisms provide an efficient and cost-effective biofuel that is self-harvesting and can be distributed readily using infrastructure already in place. As with other genomes, however, accurate functional annotation in methanogens lags significantly behind a large body of sequence data, representing a sizable gap in our understanding of biology in these organisms. We are using the methanogenic archaeon *Methanosarcina acetivorans* (MA) as a model system for developing experimental tools to rapidly and reliably annotate and validate function. The target genes are putative enzymes in MA with detectable *in vivo* expression.

Our experimental approach uses a combination of methods for rapid function assignment. Nuclear magnetic resonance (NMR) spectroscopy is used to screen for putative substrates, products, or their structural analogs. Where possible, we have followed up on function assignments by checking to see if the MA gene can complement the corresponding *E. coli* knockout. We have used this approach to both validate and correct functional assignments in MA target genes, as will be illustrated with examples. Additional discussion will focus on the insights into functional annotation of "hypotheticals" that are being obtained by integrating mass spectrometry–based metabolite profiles of gene knockouts with NMR-based approaches.

Monday, April 11, 3:40 p.m.–3:50 p.m.
Breakout Session C: Databases and Functional Annotation

GO Consortium Announcement: MENGO–The Microbial ENergy Processes Gene Ontology Workshop

João Carlos Setubal

Department of Computer Science, Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University

The MENGO project is a community-oriented, multi-institutional collaborative effort that aims to develop new gene ontology terms to describe microbial processes related to bioenergy. The Gene Ontology (GO) Consortium was formed in 1998 to create universal descriptors that can be used to describe functionally similar gene products and their attributes across all organisms. MENGO, an interest group within the GO Consortium, depends on the bioenergy community to help generate the desired terms. This presentation will provide an overview of the project and explain how the bioenergy community can participate in MENGO. A MENGO workshop will be held April 13–14 directly following this year's Genomic Science contractor-grantee meeting (same venue).

Monday, April 11, 3:50 p.m.–4:10 p.m.
Breakout Session C: Databases and Functional Annotation

Version 3 of the MG-RAST Server: A Metagenomics Portal for a Fully Democratized Sequencing World

Folker Meyer

Argonne National Laboratory

metagenomics.anl.gov

MG-RAST is the most widely used metagenomic annotation and comparison engine. Its latest version (v.3) offers a variety of novel features, comprehensive GSC metadata support, and enhanced analysis capabilities (including comparisons to COGs, EGGnogs, and KEGG). Complementing this resource are comprehensive user interface–driven tools for sample comparison that allow extraction of subsets of sequence data based on user criteria.

As sequencer output grows, standards on data analysis become more important, MG-RAST v.3 uses the M5 open standard tool set to perform quality control on input sequence sets developed in cooperation with the DOE Joint Genome Institute.

Monday, April 11, 4:10 p.m.–4:30 p.m.
Breakout Session C: Databases and Functional Annotation

The Open Data Framework for Access and Analysis of Microbiome Data

Victor Felix,[1] Narayan Desai,[2] Greg Caporaso,[3] Samuel Angiuoli,[1] Rob Knight,[3] Folker Meyer,[2] Anup Mahurkar,[1] and **Owen White**[1*]

[1]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore; [2]Argonne National Laboratory, Argonne, Ill.; [3]Department of Chemistry and Biochemistry, University of Colorado, Boulder.

*Speaker

We are developing an Open Data Framework (ODF) to provide microbiome datasets and analysis resources to the scientific community. ODF consists of a database, a data exchange format, and an associated RESTful application programming interface (API) that supports data retrieval and submissions from the community. The initial system will contain microbiome data from ~700 human microbiome samples from the NIH-supported Human Microbiome Project (HMP), environmental samples derived from oceans and soil, as well as samples derived from model organisms. All samples in ODF will have associated metadata that meet Genomic Standards Consortium specifications. In addition to raw DNA sequence from HMP samples, ODF will incorporate large-scale analysis, including genome assemblies, alignment to reference genomes, gene predictions, functional annotations, and metabolic reconstructions recently generated by the HMP consortium. Unlike conventional public archives such as GenBank, EMBL, or DDBJ that utilize centralized databases, file systems, and web servers to deliver data to users, ODF will use cloud computing and cross-site replication technologies to support a distributed data access and storage model. This infrastructure is designed to be scalable to accommodate rapidly growing metagenomic datasets derived from next-generation sequencing and promote data sharing. ODF will support access through an API and pre-defined virtual machines, including Qiime and CloVR. The API is designed for use in both analytical pipelines and online web resources. ODF also takes advantage of virtual machines to access and analyze data. In previous efforts, we have shown that users are able to launch ~1000 virtual machines on cloud-based systems, giving unprecedented access to remote, dynamically scalable computational power to the average user. This will enable complex analytical and visualization tools such as MG-RAST, Galaxy, and QIIME to query, analyze, and display microbiome data, including phylogenetic and metabolic reconstructions, and allow comparison of the metabolism and annotations of one or more metagenomes and genomes. ODF is intended to break the paradigm of conventional web-based genome resources (e.g., the UCSC Genome Browser, FlyBase, WormBase) typically used by model organism databases that are confined to monolithic web servers with centralized data.

Tuesday, April 12, 8:30 a.m.–9:00 a.m.
Plenary Session: User Facilities

DOE Joint Genome Institute Update

Eddy Rubin
DOE Joint Genome Institute

The DOE Joint Genome Institute (JGI) is transitioning from being a user facility that provides primarily large-scale-production DNA sequencing to a "genomic foundry" where a variety of state-of-the-art genomic capabilities are made available to users. Presently, these capabilities include cell sorting, single-cell genomics, transcriptomics, DNA synthesis, genomic data analysis, and access to high-performance computing, as well as large-scale DNA sequencing. These capabilities will enable users to address complex energy and environmental issues involving plants, animals, and microbes in their natural habitats. The massive discovery of biomass-degrading genes and genomes of hard-to-culture microbes in cow rumen will be described as an example of the various capabilities that the DOE JGI is making available to users to solve complex problems. In the rumen study, microbial cell sorting, DNA synthesis, single-cell sequencing, terabase-scale sequence analysis, and high-performance computing were all exploited in the discovery of functional genes and genomes from a complex microbial community.

Tuesday, April 12, 9:00 a.m.–9:30 a.m.
Plenary Session: User Facilities

The Environmental Molecular Sciences Laboratory (EMSL): High-Impact Science and State-of-the-Art Capabilities at a National Scientific User Facility

Allison Campbell
Environmental Molecular Sciences Laboratory

EMSL provides unique suites of capabilities that combine experiments with high-performance computing to advance biology, geochemistry, subsurface science, and interfacial chemistry. EMSL recently elevated its capabilities with 31 new instruments funded with $60 million from the American Recovery and Reinvestment Act. New instruments include a multispectral confocal microscope for biofilms, next-generation DNA sequencers, and the only helium ion microscope available in the United States. These systems will speed development of new energy sources, guide environmental cleanup, and provide users with data much faster and with greater detail than current instruments. They will enable advancements in biofuel research and the harnessing of microbes for reduction of harmful and toxic materials in the environment. EMSL specializes in conducting integrated high-throughput functional genomics and proteomics studies, as well as electron and fluorescence microscopy at high spatial and temporal resolutions. These capabilities provide the foundation for attaining a molecular-level understanding of both cellular networks and microbial-community dynamics, thus enabling the pursuit of systems biology. Capabilities available to EMSL users include high-performance computing and state-of-the-art mass spectrometry, nuclear magnetic resonance, multi-modal microscopy and imaging, metabolite analysis, flow cytometry, and cell isolation and systems analysis—all located under one roof. EMSL collaborates directly with users and other user facilities in support of bioenergy research centers, such as recent studies of the symbiotic relations of leafcutter ants, fungi, and bacteria for the University of Wisconsin–Madison and the DOE Great Lakes Bioenergy Research Center. Construction has begun on a new Quiet Wing—to be completed in 2012—that will provide users with new instruments requiring extremely low electromagnetic field and vibrational interference as well as high-temperature stability. EMSL also is building a data management system (myEMSL) that will enable searching and integration of the data generated across all of its instruments, thus supporting the formation of new multidisciplinary research teams. The myEMSL system also will enable entirely new approaches for performing data-driven research and help accelerate the convergence of biology with the physical and mathematical sciences.

Tuesday, April 12, 9:30 a.m.–10:00 a.m.
Plenary Session: User Facilities

The DOE Systems Biology Knowledgebase Update

Susan Gregurick

DOE Office of Biological and Environmental Research

The DOE Systems Biology Knowledgebase (Kbase) for energy and the environment is a cyberinfrastructure to integrate, search, and visualize in an open environment experimental data, associated information (metadata), corresponding models, and analysis tools. The vision and justification for Kbase were defined in a May 2008 workshop, and an implementation plan was developed based on a series of community workshops held in 2010. The Kbase implementation plan articulates scientific objectives in microbial, plant, and metacommunity research; outlines a plan for implementing these objectives; and describes architectural details of the system. The implementation plan and scientific objectives will be discussed.

Tuesday, April 12, 10:30 a.m.–11:00 a.m.
Plenary Session: Plant Feedstocks Genomics for Bioenergy

Toward an Understanding of Arbuscular Mycorrhizal Symbiosis for Growth of Biomass Feedstocks in Low-Input Systems

Maria J. Harrison (mjh78@cornell.edu)

Boyce Thompson Institute for Plant Research, Tower Road, Ithaca, N.Y.

Most vascular flowering plants, including all proposed biofeedstock species, are able to develop symbiotic associations with arbuscular mycorrhizal (AM) fungi. AM symbiosis develops in roots where the fungus colonizes cortical cells to gain access to carbon from the plant. In addition to growth within the root cortex, the fungal symbiont establishes an extensive network of hyphae in the surrounding soil, and this extra-radical mycelium delivers mineral nutrients to the root, in particular phosphorus (P) as phosphate. AM symbiosis profoundly affects plant productivity, largely through improvements in plant mineral nutrition. This is significant because phosphate availability limits crop production (Vance 2001). Moreover, rock phosphate reserves, the raw material for superphosphate fertilizers, are being depleted, and fertilizer costs are predicted to rise (Runge-Metzger 1995; Vaccari 2009). The proposed use of marginally fertile land for producing biofeedstocks will further increase the need for improved and sustainable P fertilization strategies (Hill et al. 2006). Crops such as alfalfa require high P fertilization, and even low-input prairie grass systems are predicted to become P limited (Hill et al. 2006; Tilman, Hill, and Lehman 2006). Consequently, there are strong economic and ecological incentives for increasing the efficiency of P-fertilizer use. Modifying agricultural practices to incorporate partnerships such as AM symbiosis is one part of the solution to enable economic and environmentally sustainable feedstock production.

The objectives of our work are to identify and characterize plant genes involved in AM symbiosis regulation and functioning in *Medicago truncatula* and *Brachypodium distachyon*. *M. truncatula* is a legume and well-established model for analysis of AM symbiosis. In this system, we are focusing on identifying *M. truncatula* transcription factors that regulate development of the arbuscule–cortical cell interface, which is where mineral nutrients are transferred from the fungus to the plant cell. Additionally, we have begun to identify a network of transcription factors that regulate the *M. truncatula* phosphate transporters responsible for acquiring phosphate from the fungus. Although it is a useful model and relevant to other dicotyledonous biofeedstock species, *M. truncatula* may not be the optimal system to understand AM symbiosis in biofeedstock grass species. Consequently, we have initiated analyses of AM symbiosis in a grass model, *B. distachyon*. Here, we are using sequence-based transcript profiling approaches to dissect the molecular basis of functionality differences in different AM symbioses. Progress with these projects will be discussed.

References

1. Vance, C. 2001. "Symbiotic Nitrogen Fixation and Phosphorus Acquisition. Plant Nutrition in a World of Declining Renewable Resources," *Plant Physiology* **127**(2), 390–97.
2. Runge-Metzger, A. 1995. "Closing the Cycle: Obstacles to Efficient P Management for Improved Global Food Security, 27–42. In *Phosphorus in the Global Environment: Transfers, Cycles and Management*. Ed. H. Tiessen, John Wiley and Sons, New York.
3. Vaccari, D. A. 2009. "Phosphorus: A Looming Crisis," *Scientific American* **300**(6), 54–59.
4. Hill, J., et al. 2006. "Environmental, Economic, and Energetic Costs and Benefits of Biodiesel and Ethanol Biofuels," *Proceedings of the National Academy of Sciences* USA **103**(30), 11,206–10.
5. Tilman, D., J. Hill, and C. Lehman. 2006. "Carbon-Negative Biofuels from Low-Input High-Diversity Grassland Biomass," *Science* **314**(5805), 1598–1600.

Tuesday, April 12, 11:00 a.m.–11:30 a.m.
Plenary Session: Plant Feedstocks Genomics for Bioenergy

Setaria as a Model for Grass Genetics and Biofuel Feedstock Genomics

Jeff Bennetzen

Department of Genetics, University of Georgia, Athens

The grass family Poaceae has been a major target for genetic research since the dawn of the 20th century, partly because of the ease of forward genetics in such species as maize and barley but also because of the importance of many grasses as crops for human food, feed, and forage. Some grasses like switchgrass and *Miscanthus* have recently become candidates for biofuel feedstocks, thereby adding one more incentive to the demand for basic research on the Poaceae family. Several grass species have been proposed as models for basic research, including rice, sorghum, and *Brachypodium*. Among these, maize stands out for its ease of outcrossing and history of genetic, morphological, biochemical, and physiological characterization. Yet maize research is limited by the plant's large size, long generation time, and relatively inefficient and expensive transformation.

Recently, the cereal species *Setaria italica* (foxtail millet) and its wild progenitor (green foxtail, or *Setaria viridis*) have been proposed as an exceptionally powerful model grass system. The DOE Joint Genome Institute and its collaborators have fully sequenced the ~500 Mb *S. italica* genome, and the high-quality assembly is now being confirmed with an ~1000-marker genetic map. About 400,000 expressed sequence tags also have been sequenced, along with low-pass sequences of eight *S. viridis* accessions. These two *Setaria* species are diploid C_4 grasses that last shared a common ancestor with switchgrass ~13 million years ago. Hence, the switchgrass genome sequence likely can be assembled onto a *Setaria* backbone. Some *S. viridis* accessions can produce five generations per year and can mature in small pots at a height of ~20 cm. Genetic diversity is high within *S. viridis*, and a collection of the most diverse materials is being assembled at the University of Georgia. As one of the world's most broadly distributed plant species, *S. viridis* is expected to show an unusual breadth of specialized adaptations. Recent work indicates that *Agrobacterium*-mediated transformation is as efficient on *S. viridis* as it is with rice. Outcrossing is a challenge, but we have generated five populations from crosses of different *S. viridis* accessions to the sequenced foxtail millet cultivar, Yugu1. The progeny of these crosses are advancing rapidly toward recombinant inbred populations that will be available to all interested researchers, and variously mutagenized populations for community use also are planned. Hence, *Setaria* provides an intriguing opportunity to serve as the optimal model grass, especially for studies of C_4 genetics or for its application to closely related crops like switchgrass, pearl millet, and napier grass.

Tuesday, April 12, 11:30 a.m.–12:00 p.m.
Plenary Session: Plant Feedstocks Genomics for Bioenergy

Integration of Switchgrass into the Grass Consensus Genome

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]Forage Improvement Division, Samuel Roberts Noble Foundation, Ardmore, Okla.; [3]Department of Public Health Sciences, Pennsylvania State University, Hershey

*Speaker

Switchgrass (*Panicum virgatum* L.), a warm-season C_4 perennial grass native to North America, is well suited as a feedstock for renewable bioenergy. Applying advances in genome science will enable development of improved cultivars through molecular breeding faster than would otherwise be possible. Several approaches have been taken by our group to provide an improved understanding of the polyploid switchgrass genome and its relationship to other grass systems.

Switchgrass linkage maps were constructed and are now being applied to identify QTL for important agronomic traits. This was accomplished

using a full-sib mapping population consisting of 238 individuals that was genotyped with EST-SSR, genomic SSR, and EST-STS markers. Single-dose markers were identified and analyzed following the two-way pseudotestcross strategy. The framework map length and total number of framework and accessory markers were 1,376 cM and 563 in the female and 1,645 cM and 543 in the male map. Each map consisted of 18 linkage groups arranged into nine homeologous pairs. Ratios of coupling to repulsion-phase linkages agreed with a high degree of preferential pairing. Comparative analysis of each homology group to the sorghum genome identified syntenic relationships and collinear tracts.

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 ranged from 129 to 268 cm, with a 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 from 28.2% to 48.8%, with a mean of 37.6%. QTLs associated with the traits of interest will be detected through genotyping with the same markers used for map construction.

The allopolyploid or autopolyploid origins of the subgenomes in switchgrass are unclear. Ancestral diploid progenitors have not been identified. We have used cytogenetics as one means to establish further evidence for either case. From among the progeny of the mapping population, two dihaploid (2n = 18) individuals were identified using flow cytometry and mitotic chromosome spreads. These individuals have facilitated karyotyping of the base 9 switchgrass chromosomes and appear to have lost one copy of each subgenome, as at meiosis 18 univalents were observed. First, we obtained relative lengths, centromere locations, and arm ratios. The condensation patterns of prometaphase chromosomes have been quantitatively analyzed and fluorescence in situ hybridization has allowed assignment of 5S rDNA, 45S rDNA, and CentC genetic probes, creating a foundation for future genome analyses based on chromosome structure and identity. Integration of the cytogenetic studies with the genetic map is our current objective and will involve assignment of genetically mapped BAC probes. This work has demonstrated that the karyotypes of the subgenomes are very similar.

Tuesday, April 12, 2:10 p.m.–2:35 p.m.
Breakout Session D: Microbes and the Global Carbon Cycle

The Impacts of Global Warming on the Carbon Cycle of Arctic Permafrost

T. C. Onstott

Department of Geosciences, Princeton University

This presentation considers the impacts of elevated temperature on carbon processing by permafrost microbial communities using a combination of manipulative microcosm experiments on permafrost cores and correlated observations at a well-characterized Arctic field site. The project uses a combination of "omics" techniques, stable isotope probing, and measurements of biogeochemical fluxes of CO_2, CH_4, and organic intermediates. These data will be integrated into a biogeochemical reaction/transport model that identifies specific pools of soil carbon most vulnerable to degradation and predicts CO_2 and CH_4 production as permafrost thaws.

Tuesday, April 12, 2:35 p.m.–3:00 p.m.
Breakout Session D: Microbes and the Global Carbon Cycle

Carbon Dynamics in Rhizosphere Soil

Mary Firestone

Department of Environmental Science, Policy, and Management, University of California, Berkeley

This presentation will focus on the role of soil microbial communities in degradation and transformation of carbon released from the roots of grasses and, in particular, the impact of elevated CO_2 concentrations on these processes. This project uses soil microcosms containing seedlings of *Avena barbata* (a grass common to the western United States) and a combination of stable isotope probing and metatranscriptomic sequencing to identify metabolic pathways of soil microbes responsible for carbon processing. Fluorescent in situ hybridization probing and NanoSIMS analysis allow direct identification of active community members and visualization of spatial associations between microbes and plant roots.

Tuesday, April 12, 3:00 p.m.–3:25 p.m.
Breakout Session D: Microbes and the Global Carbon Cycle

Systems Biology of Methanosarcinales

Bill Metcalf

School of Molecular and Cellular Biology, University of Illinois

This presentation will focus on new systems biology approaches to study methanogenic microbes capable of converting the organic acid acetate to methane (the archaeal class *Methanosarcina*). Since this process is a major source of methane production from global wetlands, understanding the function and regulation of the underlying metabolic processes would allow better prediction of potential responses to climate change variables. The work involves sequencing the genomes of 40 representative species, performing comparative genomic analysis to identify patterns in metabolic and regulatory circuits, assembling transcriptional networks, and developing an integrative metabolic network model of this class of organism.

Tuesday, April 12, 3:45 p.m.–4:10 p.m.
Breakout Session D: Microbes and the Global Carbon Cycle

Atmospheric N Deposition Slows Litter Decay, Alters Fungal Community Composition, and Suppresses Ligninolytic Gene Expression in a Northern Hardwood Forest

Don Zak

Department of Ecology and Evolutionary Biology, University of Michigan

This presentation focuses on understanding the impact of anthropogenic nitrogen deposition on microbially mediated carbon cycling processes in soil, in particular the observation that adding nitrogen results in decreased plant litter decay rates and increased carbon storage in soils in some forest ecosystems. The research being discussed leverages a pre-existing, long-term field study of nitrogen levels in forest soils, uses a combination of community

transcriptomics to assess gene expression by prokaryotic and fungal community members, and directly measures soil carbon cycling processes. It also will link environmental factors affecting the regulation of microbial genes to biogeochemical processes in an attempt to understand specific mechanisms underlying the accumulation of soil carbon in response to nitrogen addition.

Tuesday, April 12, 4:10 p.m.–4:35 p.m.
Breakout Session D: Microbes and the Global Carbon Cycle

Impact of "Cheating" in Microbial Community Carbon Cycle Processes

Chris Blackwood

Department of Biological Sciences, Kent State University

This presentation considers the role of microbial "cheaters" in the degradation of plant matter in soils. The majority of plant biomass is composed of long, complex carbon polymers (e.g., lignocellulose) that are too large for microbes to directly import. Instead, many soil microbes invest energy in producing extracellular enzymes that break down plant matter outside the cell and then they consume the resulting carbohydrate fragments. However, rather than producing their own extracellular enzymes, some soil microbes "cheat" by taking up carbohydrates liberated by those of other organisms. This project uses a combination of molecular approaches and stable isotope probing to identify cheating microbes, investigate the regulation of this metabolic strategy, and examine the impact of cheating on carbon cycling in soil environments.

Tuesday, April 12, 4:35 p.m.–5:00 p.m.
Breakout Session D: Microbes and the Global Carbon Cycle

Omics Techniques for Functional Analysis of Microbial Carbon Cycling

Bob Hettich

Oak Ridge National Laboratory

New developments in analytical chemistry coupled to next-generation omics approaches for molecular microbiology are providing sophisticated new techniques for examining functional processes of microbial communities. Using genomic and metagenomic databases as reference frameworks, these tools allow characterization of proteins and other biomolecules from intact microbial consortia and increasingly complex communities, often directly from environmental samples. This presentation will discuss new advances in this area and consider current technical challenges.

Tuesday, April 12, 2:05 p.m.–2:30 p.m.
Breakout Session E: Genome Structure and Epigenetics

Fidelity and Dynamics of DNA Methylation in Plants

Qin Yao,[1] Changjun Liu,[1] John Shanklin,[1] Chuan He,[2] and **John Dunn**[1*] (jdunn@bnl.gov)

[1]Department of Biology, Brookhaven National Laboratory, Upton, N.Y.; [2]Department of Chemistry and Institute for Biophysical Dynamics, University of Chicago

*Speaker

Epigenetics is defined as the study of heritable changes to genome structure and function that do not change the nucleotide sequence of DNA. Methylation of cytosine to form 5-mC in genomic DNA is an important epigenetic marker that plays a critical role in regulation of gene expression, chromatin structure, and genome stability. In all organisms, cytosine methylation is a postreplicative process. Newly synthesized DNA strands are unmethylated, thus creating hemimethylated duplexes at replication forks. In both mammals and plants, most methylation occurs at the DNA dinucleotide CpG, where both cytosines in the complementary strands of adjacent base pairs are methylated. In mammals, the UHRF1 protein recognizes these hemimethylated CpG sites via its SET-and RING-associated (SRA) domain. Structural studies have shown that the 5-mC residue in hemimethylated DNA bound to UHRF1 is flipped outside the DNA helix into a specific 5-mC–binding pocket within the SRA domain. This causes DNA looping and allows the N-terminal region of the SRA domain to interact with the DNA's major and minor grooves. The residue requiring modification is then flipped out of the helix and presented to the DNA methyl transferase DNMT1 for addition of the methyl group.

In *Arabidopsis*, VIM1 encodes an SRA domain methylcytosine-binding protein that probably functions similarly to UHRF1 in playing a major role in maintaining DNA methylation patterns following DNA replication. To gain further insight into how VIM1 functions, we have cloned, expressed, and purified VIM1 and are using electrophoretic mobility-shift and fluorescence anisotropy titration assays to study its interaction with model duplex DNAs containing cytosine or 5-mC in one or both strands.

A fine-scale mapping tool, Bisulfite Patch PCR (Varley and Mitra 2010), is also being used to discover the potential epigenetic regulation underlying the spatial and temporal expression of 26 genes or transcription factors involved in lignin biosynthesis in *Arabidopsis*. This new approach allows us to process ~100 genes from multiple samples at the same time.

Recent studies of genomic DNA from the human brain, neurons, and mouse embryonic stem cells have demonstrated that these DNAs contain a sixth base, oxidized 5-mC or 5- hydroxymethylcytosine (5-HmC). Current thinking is that 5-HmC does not merely mimic 5-mC groups but likely plays an independent role in yet-unknown epigenetic regulation of various biological processes. We are using bacteriophage T4 β-glucosyltransferase to transfer a glucose moiety containing an azide group onto the hydroxyl group of 5-HmC. The azide group can then be chemically modified with biotin for detection, affinity enrichment, and subsequent sequencing of 5-HmC–containing DNA fragments to reveal the genomic locations of 5-HmC in the DNA (Song et al. 2011). Dot blot detection of biotinylated glucose-HmC shows about 0.06% 5-Hmc in plant leaf genomic DNA. Efforts are under way to map these sites. We are also determining if 5-HmC affects VIM1 binding and if 5-HmC residues can be removed by the *Arabidopsis* DNA glycosylase/lyase, Repressor of Silencing (ROS1), thereby allowing the 5-HmC generated by oxidation of 5-mC to be replaced by C, resulting in active demethylation.

References

1. Varley, K. E., and Mitra, R. D. 2010. "Bisulfite Patch PCR Enables Multiplexed Sequencing of Promoter Methylation across Cancer Samples," *Genome Research* **20**, 1279–87.
2. Song, C. X., et al. 2011. "Selective Chemical Labeling Reveals the Genome-Wide Distribution of 5-Hydroxymethylcytosine," *Nature Biotechnology* **29**, 68–72.

Tuesday, April 12, 2:30 p.m.–2:55 p.m.
Breakout Session E: Genome Structure and Epigenetics

Genomic Analysis of miRNAs and Target RNAs of *Brachypodium*

Pamela J. Green,[1*] Dong-Hoon Jeong,[1] Skye A. Schmidt,[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] and Blake C. Meyers[1]

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

*Speaker

MicroRNAs (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. Although 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. More than 64 million reads were obtained from 12 small RNA libraries, resulting in an average of over 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 miRNAs 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, identifying miRNA and target RNA pairs in *Brachypodium* will help to better understand how small RNAs contribute to the regulation of genes and genomes. Future experiments to expand this analysis to potential bioenergy crops will be discussed.

This work was funded by grant # DE-FG02-07ER64450 of the Plant Feedstock Genomics for Bioenergy program of the U.S. Department of Energy.

Tuesday, April 12, 2:55 p.m.–3:20 p.m.
Breakout Session E: Genome Structure and Epigenetics

DNA Methylation Profiling Using Illumina Sequencers

Matteo Pellegrini

Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles

The development of efficient and cost-effective sequencing strategies during the last decade has revolutionized genome-scale biology. We will present methods for the reconstruction of genome-wide DNA methylation profiles. By including RNA-seq data from the same samples, we are also able to study the relationship between allele-specific methylation and transcription. Finally, by sequencing the methylome of parental strains and their crosses, we are able to observe allele-specific inheritance of methylation and measure the degree to which this is conserved between these.

Tuesday, April 12, 3:20 p.m.–3:45 p.m.
Breakout Session E: Genome Structure and Epigenetics

Maintenance of H3 Lysine 9 Methylation in *Arabidopsis*

Zhicheng Dong and **Judith Bender***

Department of Molecular Biology, Cell Biology, and Biochemistry, Brown University

*Speaker

In many eukaryotes, including mammals and plants, histone H3 lysine 9 methylation (H3meK9) and cytosine methylation (5meC) are targeted to invasive sequences such as transposons, viruses, and transgenes to transcriptionally silence these sequences, thereby preventing deleterious consequences such as the expression of aberrant gene products and transposon mobilization. Accurate maintenance of H3meK9 and 5meC on invasive sequences but not on host genes is required for normal cellular function. In *Arabidopsis*, H3meK9 mediated by SUVH histone methyltransferases (MTases) is associated with 5meC mediated by the CMT3 cytosine MTase. In order to understand targeting of H3meK9, we developed a novel strategy to track hotspots of

SUVH activity. Our strategy uses expression of SUVH active site mutants predicted to catalyze tri-methylation of H3 K9 (H3me3K9), rather than the mono- or di-methylation catalyzed by wildtype SUVH enzymes. We focused on the SUVH6 MTase, which acts preferentially at a transcribed inverted repeat locus, and found that the SUVH6 active site mutant drives H3me3K9 specifically at this preferred target. In addition, the novel H3me3K9 modification results in hyper-5meC at the preferred target through enhanced CMT3 activity. Because both H3me3K9 and hyper-5meC are detectable above the levels of silencing modifications present in the genome of a wildtype plant, the SUVH active site mutant strategy provides a means to track MTase activity genome-wide with high-throughput approaches.

Tuesday, April 12, 4:00 p.m.–4:25 p.m.
Breakout Session E: Genome Structure and Epigenetics

Epigenome Atlas and Comparative Epigenome Analysis

Aleksandar Milosavljevic (amilosav@bcm.edu)
Molecular and Human Genetics Department, Baylor College of Medicine, Houston, Texas

New sequencing technologies are enabling comprehensive mapping of DNA methylation, histone modifications, chromatin structure, and chromatin-associated RNA in humans and other species. These mapping projects are revealing epigenomic variation at all levels of biological complexity, from species to cells. To establish a basis for studying epigenomic variation in humans, the National Institutes of Health Roadmap Epigenomics Initiative (www.roadmapepigenomics.org) aims to produce reference epigenome maps from a representative collection of hundreds of human cell and tissue types and compile them into the Human Epigenome Atlas (www.epigenomeatlas.org). We review bioinformatic frameworks for constructing the Human Epigenome Atlas and for conducting comparative epigenome analysis.

Tuesday, April 12, 4:25 p.m.–4:50 p.m.
Breakout Session E: Genome Structure and Epigenetics

Linking Genetic, Epigenomic, and Transcriptional Variation in 1,001 *Arabidopsis* Genomes

Joseph R. Ecker
Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, Calif.

Traditionally, phenotype is defined by a combination of genetic and environmental interactions. Missing from this equation is an understanding of the impact that epigenetic variation has on phenotypic output. Consequently, we are sequencing the genomes, DNA methylomes, and transcriptomes of >150 *Arabidopsis thaliana* accessions to better understand the interplay and impact of genetic, epigenetic, and transcriptional variation on phenotype. We have chosen to use this model plant species for our studies of natural epigenetic variation because accessions have been collected from diverse geographical regions; the species is naturally inbred, resulting in a nearly isogenic genome; and methods for whole-methylome analysis are now well established. The combination of methylome and transcriptome datasets is allowing us to clearly define genomic regions that display characteristics of epialleles, genes that are silenced by RNA-directed DNA methylation. We have identified over 1,000 epialleles from ~150 accessions that exhibit a decrease in transcription in addition to the presence of DNA methylation characteristic of gene silencing. These epialleles primarily fall into two major classes: (1) transposons/repeats and (2) canonical genes. Epialleles that occur within transposons/repeats occur frequently within the population, typically >50%, whereas epialleles in canonical genes usually occur at a low frequency in the population, typically ~<5%. Our current efforts are focused toward understanding the impact these epialleles have on phenotypic variation. Combining the rich history of genetics research in the *Arabidopsis* reference accession Col-0 and our epialleles is allowing us to address some easily testable hypotheses. These studies will enable us to characterize the genetic nature of epialleles and, for the first time, understand how many of the identified epialleles cannot be explained by a change in genotype, thus paving the way for future studies of epiallelic variation in any organism.

Tuesday, April 12, 4:50 p.m.–5:15 p.m.
Breakout Session E: Genome Structure and Epigenetics

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, Oregon State University (OSU), Corvallis, Ore.; [2]Molecular and Cellular Biology program, OSU; [3]Department of Biochemistry and Biophysics, OSU; [4]Department of Botany and Plant Pathology, OSU; [5]Deptartment of Molecular, Cell, and Developmental Biology, University of California, Los Angeles

*Speaker

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 11 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 1-Kb 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 to 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: poplar-dev.cgrb.oregonstate.edu/cgi-bin/gbrowse/poplar_v2/.

Tuesday, April 12, 2:00 p.m.–2:30 p.m.
Breakout Session F: Computational Biology—
New Mathematical Methods to Model and Understand Biological Processes

Computational Identifications of Bacterial RNA Regulatory Motifs Using a Gibbs Sampling Algorithm

Lauren Alpert, Donglai Wei, and Charles (Chip) E. Lawrence*

Brown University

*Speaker

Regulatory RNAs in bacteria play an essential role in the modulation of gene expression. For example, riboswitches regulate translation, CRISPR arrays provide resistance to bacteriophage infection, and sRNAs alter gene expression in various ways. In each of these cases, the secondary structure of the RNA regulatory element is key to their function. Over the past three decades, substantial progress has been made in predicting RNA secondary structures (Mathews 2006). Here, our focus is on identifying the secondary structure common to a set of RNA sequences believed to share common structural features associated with post-transcriptional regulation. Our goal for this research is the development of statistical inference procedures and algorithms that extend motif-finding algorithms (which have had considerable success in identifying DNA regulatory motifs and the cognate binding sites) to RNA motif finding. Our approach involves a Bayesian statistical procedure known as the Gibbs sampler that we have developed and used for DNA motif finding over a number of years (Lawrence et al. 1993; Newberg et al. 2007; Conlan et.al 2005). The novel and challenging aspect of this problem is to predict RNA secondary structure, itself a difficult problem, while simultaneously performing all the other components of motif finding. We have recently developed a novel Gibbs sampling algorithm for this purpose and shown that it improves positive predictive values over extant technologies by more than 15% over a wide range of sensitivities. Here, we will describe this algorithm and its application to bacterial regulatory RNAs. Preliminary applications to lysine, RFN, THI, and S-box riboswitch families from the Rfam database (Griffiths-Jones et.al. 2005; rfam.janelia.org) show areas under the PPV-sensitivity ROC curves of 93%, 67%, 89%, and 88%, respectively. This indicates that, in three of these families, the algorithm identifies nearly all of the base pairs in the reference structures while making a small number of incorrect predictions. For this meeting, we will report on applications to other bacterial families now under study and on our progress in identifying RNA motifs and sites in full-length UTRs.

References

1. Conlan, S., C. Lawrence, and L. A. McCue. 2005. "*Rhodopseudomonas palustris* Regulons Detected by Cross-Species Analysis of Alphaproteobacterial Genomes," *Applied and Environmental Microbiology* **71**(11), 7442–52.
2. Griffiths-Jones, S., et al. 2005. "Rfam: Annotating Non-Coding RNAs in Complete Genomes," *Nucleic Acids Research* **33**, D121–41.
3. Lawrence, C. E., et al. 1993. "Detecting Subtle Sequence Signals: A Gibbs Sampling Strategy for Multiple Alignment," *Science* **262**(5131), 208–14.
4. Matthews, D. H. 2006. "Revolutions in RNA Secondary Structure Prediction," *Journal of Molecular Biology* **359**(3), 526–32.
5. Newberg, L. A., et al. 2007. "A Phylogenetic Gibbs Sampler that Yields Centroid Solutions for cis-Regulatory Site Prediction," *Bioinformatics* **23**(14), 1718–27.

Tuesday, April 12, 2:30 p.m.–3:00 p.m.
Breakout Session F: Computational Biology—
New Mathematical Methods to Model and Understand Biological Processes

Comparative Genomics Approaches for Reconstruction of Transcriptional Regulatory Networks in Bacteria

Dmitry Rodionov[1*] (rodionov@burnham.org) and **Pavel Novichkov**[2] (PSNovichkov@lbl.gov)

[1]Sanford-Burnham Medical Research Institute, La Jolla, Calif.; [2]Lawrence Berkeley National Laboratory, Berkeley, Calif.

*Speaker

Genome-scale annotation of gene regulatory features and reconstruction of transcriptional regulatory networks in a variety of diverse microbes are

critical tasks in modern genomics and systems biology. Constituting important challenges, these tasks are prerequisites for understanding molecular mechanisms of transcriptional regulation in prokaryotes; identifying regulatory circuits; and interconnecting them with each other and with various metabolic, signaling, and other cellular pathways. A growing number of complete prokaryotic genomes allows us to extensively use comparative genomic approaches to infer cis-acting regulatory elements (e.g., transcription factor binding sites and riboswitches) in regulatory networks of numerous bacteria groups. Two major components of this analysis are propagation of previously known regulons from model organisms to other species and *ab initio* prediction of novel regulons, as implemented in the integrative web-server tool RegPredict (regpredict.lbl.gov).

We developed and used an integrative comparative genomics approach to infer transcriptional regulatory networks in ~100 microbial genomes from 10 distinct taxonomic groups of bacteria (each including 6 to 16 genomes): *Shewanella*, Desulfovibrionales, Enterobacteriales, cyanobacteria, Thermotogales, Bacillales, *Streptococcus*, *Staphylococcus*, *Ralstonia*, and *Corynebacteria*. A limited input of established regulon members was provided by publications on particular transcription factors in individual species (*Escherichia coli*, *Bacillus subtilis*, and *Staphylococcus aureus*). Reconstructed regulatory networks for the key pathways involved in central metabolism, production of energy and biomass, metal homeostasis, stress response, and virulence include over 450 regulators, >22,000 of their DNA-binding sites, and 25 families of RNA regulatory elements (e.g., riboswitches). The obtained reference set of microbial regulons is captured in the RegPrecise database within the taxonomic group–specific collections (regprecise.lbl.gov). Many novel regulons first predicted and reconstructed by the comparative genomics techniques were validated by targeted *in vivo* and *in vitro* experiments. The network of regulatory interactions obtained by genomic analysis provides a framework for interpreting gene expression data in model species.

Tuesday, April 12, 3:00 p.m.–3:30 p.m.
Breakout Session F: Computational Biology—
New Mathematical Methods to Model and Understand Biological Processes

Satisfying Flux Balance and Mass-Action Kinetics in a Network of Biochemical Reactions

Michael Saunders

Department of Management Science and Engineering, Stanford University

Computational modeling of biochemical reaction networks will become increasingly dependent on large-scale numerical optimization as network dimensionality continues to rise. Linear optimization, the mainstay for flux balance analysis, is trivial for well-scaled metabolic models but nontrivial for integrated metabolic and macromolecular synthetic networks because of the size of the stoichiometric matrix and the wide temporal scales involved. We describe the first such reconstruction for *Escherichia coli*, which accounts for ~43% of all its genes, and a similar reconstruction for *Thermotoga maritime*, as well as numerical challenges associated with flux balance analysis.

Beyond linear optimization, we have established that a sequence of parametric convex optimizations can be used to model the inherently nonlinear kinetic relationship between reaction flux and metabolite concentration. Convex optimization preserves many of the attractive numerical features of linear optimization algorithms. For example, given a model of a biochemical network, one can establish *a priori* whether a solution will exist. The nonexistence of a solution is an indication of a malformed model arising from, for example, incorrect stoichiometry or reaction directionality.

With a sequence of convex optimization problems, establishing the existence of a fixed point for such a sequence is equally important for indicating when a model is malformed. We summarize recent results establishing necessary conditions for existence of a flux and concentration at such a fixed point. The approach also leads to a numerically tractable algorithm for simultaneous satisfaction of flux balance

and mass-action kinetics for a well-scaled stoichiometric model, subject to certain conditions.

The overall objective of this work is to provide the systems biology community with a new wave of computationally efficient algorithms for physicochemically realistic modeling of genome-scale biochemical networks. We highlight open-source algorithms already released.

Tuesday, April 12, 3:45 p.m.–4:15 p.m.
Breakout Session F: Computational Biology—
New Mathematical Methods to Model and Understand Biological Processes

Proteomics of Environmental Samples Using Genome-Based Optimizations and Spectral Library Developments

William Cannon

Pacific Northwest National Laboratory

We present a method for identifying peptides and proteins from CID MS/MS spectra that increases the number of spectra identified by 60% to 147%. The method is a hybrid spectral library/database search, and the large increase in identifications results from several factors. First, we employ a probabilistic method for incorporating intensities into the identification process that sufficiently allows for variability in intensities between spectra. This enables the use of model spectra that have a wide range of fidelity to the experimentally observed spectrum. Second, the nature of the statistical distributions of scores obtained using the different model spectra allows us to analyze the error rates using a single method regardless of whether the model spectrum was obtained from a spectral library or a database search. We applied this approach to proteomics analysis of *Synechococcus sp.* PCC 7002, a cyanobacterium that is a model organism for studies of photosynthetic carbon fixation and biofuels development. The increased specificity and sensitivity of this approach allowed us to identify many more peptides involved in processes important for photoautotrophic growth. In addition, we discuss the identification of proteins from unknown (environmental) samples from MS/MS spectra using a novel identification strategy. The method provides an effective way to control the false discovery rate for environmental samples and offers an alternative to *de novo* peptide sequencing.

Tuesday, April 12, 4:15 p.m.–4:45 p.m.
Breakout Session F: Computational Biology—
New Mathematical Methods to Model and Understand Biological Processes

Ultrascale Computational Modeling of Phenotype-Specific Metabolic Processes in Microbial Communities

Nagiza F. Samatova[1,2*] (samatovan@ornl.gov), **Chongle Pan**[2] (pcl@ornl.gov), **Robert (Bob) L. Hettich**[2] (hettichrl@ornl.gov), and **Jill Banfield**[3] (jbanfield@berkeley.edu)

[1]North Carolina State University, Raleigh; [2]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]University of California, Berkeley

*Speaker

Many microbial communities in natural environments exhibit phenotypes of interest to DOE, including oxidization of pyrite ore that leads to acid mine drainage, breakdown of the lignocellulosic barrier of biomass, and biodegradation of various environmental contaminants. Addressing bioremediation and bioenergy problems will require understanding how interacting biochemical pathways in these communities lead to specific phenotypic traits (e.g., nitrogen and carbon fixation, resistance to heavy metals, and tolerance to pH perturbations). This problem cannot be solved by experiments alone. Computational modeling methods are needed that will reveal phenotype-related "signals" and their combinatorial interplay by comparing potentially hundreds of microorganisms with millions of genes organized into thousands of metabolic pathways that are uncertainly defined. These methods are being applied to the acid mine drainage (AMD) community to help answer long-standing questions regarding the role of fine-scale variation in adaptation to dynamic environmental conditions and community composition.

We performed and published a quantitative proteomics comparison of field AMD biofilm to laboratory AMD biofilm. To enable laboratory studies of growth, production, and ecology of AMD microbial communities, a culturing system was designed to reproduce natural biofilms, including organisms

recalcitrant to cultivation. A comprehensive metabolic labeling-based quantitative proteomic analysis was utilized to verify that natural and laboratory communities were comparable at the functional level. Results confirmed that the composition and core metabolic activities of laboratory-grown communities were similar to a natural community, including the presence of active, low-abundance bacteria and archaea that have not yet been isolated. However, laboratory growth rates were slow compared to natural communities, and this correlated with increased abundance of stress response proteins for the dominant bacteria in laboratory communities. Modification of cultivation conditions reduced the abundance of stress response proteins and increased laboratory community growth rates. This was the first application of a metabolic labeling–based quantitative proteomic analysis at the community level and resulted in a model microbial community system ideal for testing physiological and ecological hypotheses.

We also performed and published a quantitative proteomics study of pH perturbation to functionally characterize laboratory-cultivated acidophilic communities sustained in pH 1.45 or 0.85 conditions. The distributions of all proteins identified for individual organisms indicated biases for either high or low pH and suggest pH-specific niche partitioning for low-abundance bacteria and archaea. Although the proteome of the dominant bacterium, *Leptospirillum* group II, was largely unaffected by pH treatments, analysis of functional categories indicated proteins involved in amino acid and nucleotide metabolism as well as cell membrane and envelope biogenesis were generally more abundant at high pH. Results indicate solution pH may play an important role in shaping community membership and biofilm structure. Proteomic analysis of communities also revealed differences in the number of phage proteins detected across biological replicates. Stochastic spatial heterogeneity of viral outbreaks may also play a role in shaping community structure. Quantitative proteomic comparisons showed distinct differences in community composition and metabolic function of individual organisms during different pH treatments and confirmed the importance of specific geochemical parameters that fine-tune acidophilic microbial community structure and function at the species and strain level.

In support of these proteomics studies, we developed, published, and released a *de novo* sequencing algorithm, **Vonode**, to exploit the potential of high-resolution MS/MS data by using a unique tag scoring function and a novel type of spectrum graphs. When compared to an established *de novo* sequence algorithm, PepNovo v2.0, the Vonode algorithm inferred sequence tags for 11,422 (vs. 2,573) spectra at an average length of 5.5 (vs. 6.0) residues with 84% (vs. 65%) accuracy of inferred consensus sequence tags.

We also developed and released a module for **ProRata**, a data analysis algorithm for quantitative proteomics, to address two critical needs: (1) combining multiple replicates and assessing the reproducibility of measurements to obtain reliable quantification information and (2) comparing two unlabeled field samples of interest to a labeled reference sample grown in the laboratory because we cannot label metabolic labeling to a field sample.

We developed three graph-theoretical and statistical methods for *in silico* prediction of cellular subsystems related to the expression of a target phenotype:

- The Network Instance-Based Biased Subgraph Search (**NIBBS**) is capable of comparing hundreds of genome-scale metabolic networks to identify ***metabolic subsystems*** that are statistically biased toward phenotype-expressing organisms.
- The ***α,β*-motifs** approach allows for identification of ***functional modules*** that, in addition to metabolic subsystems, could include their regulators, sensors, transporters, and even uncharacterized proteins predicted to be related to the target phenotype.
- The Dense ENriched Subgraph Enumeration (**DENSE**) algorithm allows for incorporating partial ***prior*** knowledge about the proteins involved in a phenotype-related process and enriches that knowledge with newly identified sets of functionally associated proteins present in individual phenotype expressing organisms.

From the results obtained, for example, we were able to predict cellular subsystems that are likely related to various phenotypes such as acid tolerance, biohydrogen production, and aerobic and anaerobic respiration. Also, the genome-scale comparative network analysis enabled us to predict pathway crosstalks and to perform a systematic study on various mechanisms underlying crosstalks.

This research is supported by both the Office of Biological and Environmental Research and by the Office of Advanced Scientific Computing Research of the U.S. Department of Energy.

Wednesday, April 13, 8:30 a.m.–9:00 a.m.
Plenary Session: Innovative Tools for Genomic Sciences

Radiotracer Imaging and Genomic Science

William W. Moses

Lawrence Berkeley National Laboratory

Radiotracer imaging has been used for over half a century in nuclear medicine. PET (positron emission tomography) and SPECT (single photon emission computed tomography) cameras routinely make three-dimensional images of radiotracer distributions in patients, with 5 to 10 mm FWHM (full width at half maximum) spatial resolution and 0.05% to 1% detection efficiency (for radionuclides in the camera's field of view). Since radioactivity distribution depends on the details of the radiopharmaceutical used (as opposed to the radioisotope), many different disease states and biochemical conditions can be probed by using different radiopharmaceuticals.

In the past decade, these techniques have applied to preclinical studies that image small animals, especially mice. Imaging instruments have been modified to accommodate the finer spatial resolution and smaller field of view dictated by the smaller "patients," and spatial resolutions as fine as 0.5 to 1 mm FWHM have been achieved. There are many reasons for applying these radiotracer imaging techniques to small animals. They are truly "tracer" techniques because nanomolar concentrations can be detected and measured and the measurement does not perturb the system. This enables repeat studies to be conducted on the same animal so that it can act as its own control, allowing accurate measurement of how a disease or biochemical state evolves. The vast number of radiopharmaceuticals available also allows many biological questions to be addressed.

These same techniques can be adapted to genomic science applications in plant or microbial systems. Research on such systems needs to map biological processes on many different size and time scales—from local cellular trafficking to global biochemical pathways and from seconds and minutes to days and weeks. Although optical imaging and autoradiography are the imaging tools most commonly used in these applications, they generally can be applied only to small objects and have limited depth of field. The penetrating nature of gamma rays makes radionuclide imaging well-suited to plant, microbial, and environmental research. As biofuel development progresses, researchers will need to quantitatively and noninvasively measure increasingly larger model systems, and radionuclide techniques may be able to play a significant role. In environmental remediation, important processes occur on the micron, meter, and hundred meter size scales. While appropriate imaging tools are available at the micron and hundred meter scales, few are available at the meter scale, which is well-suited to radionuclide imaging. Therefore, an opportunity exists to modify radiotracer imaging devices and radiotracer probes for biofuel optimization and environmental remediation research.

This presentation will explore how these techniques can be adapted to suit the unique requirements of biofuel development and environmental remediation. By using radiotracer isotopes with longer half-lives than those used for imaging patients or small animals, processes with time scales as long as months can be explored. Radioligands that probe specific plant and microbial processes can be developed and used to quantitatively measure flow and metabolic activity with high sensitivity. By changing camera geometry, size scales as small as tens of microns or as large as meters can be imaged. In short, the same aspects of radiotracer imaging that make it attractive for both clinical and small animal applications are likely to be attractive to the genomic science community.

Wednesday, April 13, 9:00 a.m.–9:30 a.m.
Plenary Session: Innovative Tools for Genomic Sciences

Metabolomics and Chemical Imaging for Probing Cellular Heterogeneity

Jonathan V. Sweedler,[1*] **Bruce Fouke**,[1] and **Paul W. Bohn**[2]

[1]Institute for Genomic Biology, University of Illinois at Urbana-Champaign; [2]Department of Chemical and Biomolecular Engineering and Department of Chemistry and Biochemistry, University of Notre Dame, Ind.

*Speaker
www.scs.illinois.edu/sweedler/
mcb.illinois.edu/faculty/profile/fouke/
www.nd.edu/~bohngrp/

Given the rapid advances in measurement sciences over the past decades, what measurement challenges remain? Although multiple techniques can characterize a surface and provide either chemical or spatial information, they tend not to provide such information at single-cell resolution, and oftentimes there is a trade-off between spatial resolution and chemical information. Whether studying plant growth, microbial communities, or even a mammalian organ, approaches that provide multidimensional and contextual information are essential. Highlighted in this discussion is a range of approaches based on mass spectrometry (MS) that provide unmatched chemical information on the cellular metabolome. For example, both direct matrix-assisted laser desorption/ionization time-of-flight MS and capillary-electrophoresis electrospray ionization time-of-flight MS provide metabolome coverage at the level of individual cells. This allows us to investigate metabolic heterogeneity among selected cells within a cell population, providing unique information on cell function and fate.

While traditional MS-based approaches provide little spatial information, several approaches combine the chemically information-rich data obtained from MS with spatial information. The addition of imaging to MS allows rare cells to be characterized, provides information on the functional context of detected analytes, and allows us to probe the interface between cells and their environment. MS-based imaging can use several mass spectrometry platforms, including secondary ion mass spectrometry and laser desorption/ionization mass spectrometry. Using these approaches, we have characterized large numbers of metabolites and cell-to-cell signaling molecules in several well-defined cellular networks.

To enhance chemical and spatial information, Raman microspectroscopy and MS imaging are combined to provide unmatched figures of merit for spatially resolved imaging of lignocellulosic materials. The application of such heterocorrelation spectroscopies to visualizing the structural and chemical changes during the processing of lignocellulosic materials demonstrates the advantages of multimodal chemical imaging approaches. The potential for other multidimensional information-rich characterization approaches will be discussed.

Wednesday, April 13, 9:30 a.m.–10:00 a.m.
Plenary Session: Innovative Tools for Genomic Sciences

Phenotype Sequencing: Identifying the Genes that Cause a Phenotype Directly from Pooled Sequencing of Independent Mutants

Chris Lee

Center for Computational Biology, University of California, Los Angeles

Random mutagenesis and phenotype screening provide a powerful method for dissecting microbial functions, but their results can be laborious to analyze experimentally. Each mutant strain may contain 50 to 100 random mutations, necessitating extensive functional experiments to determine which one causes the selected phenotype. To solve this problem, we propose a "phenotype sequencing" approach in which genes causing the phenotype can be identified directly from sequencing of multiple independent mutants. We developed a new computational analysis method showing that (1) causal genes can be identified with high probability from even a modest number of mutant genomes and (2) costs can be cut many-fold compared with a conventional genome sequencing approach via an optimized strategy of library-pooling (multiple strains per library) and tag-pooling

(multiple tagged libraries per sequencing lane). We have performed extensive validation experiments on a set of *Escherichia coli* mutants with increased isobutanol biofuel tolerance. We generated a range of sequencing experiments varying from 3 to 32 mutant strains, with pooling on 1 to 3 sequencing lanes. Our statistical analysis of these data (4,099 mutations from 32 mutant genomes) successfully identified 3 genes (acrB, marC, acrA) that have been independently validated as causing this experimental phenotype. It must be emphasized that our approach reduces mutant sequencing costs enormously. Whereas a conventional genome sequencing experiment would have cost $7,200 in reagents alone, our phenotype sequencing design yielded the same information value for only $1,200. In fact, our smallest experiments reliably identified acrB and marC at a cost of only $110 to $340.

Wednesday, April 13, 10:30 a.m.–10:50 a.m.
Plenary Session: Early Career Program

A Systems Biology, Whole-Genome Association Analysis of the Molecular Regulation of Biomass Growth and Composition in *Populus deltoides*

Matias Kirst

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

Poplars are the principal short-rotation woody crop species for providing clean, renewable, and sustainable fuels in North America because of their fast, perennial growth habit and wide natural distribution in a broad range of environments. Although poplars provide the benefits of an ideal bioenergy crop, with few exceptions, the genes regulating productivity and biomass composition are largely unknown despite their critical relevance for efficient conversion of biomass to biofuels. This gap is the main barrier for efficient molecular breeding and selection of superior poplar germplasm and, consequently, the extensive adoption of this woody crop as a renewable bioenergy source. Association genetics has become the primary approach for identifying genes that regulate complex traits in human genetics, agriculture, and forestry because this strategy captures information at high resolution on a broad range of alleles that control phenotypic variation. Poplars are particularly suited to unveil the molecular basis of biomass productivity and composition using association genetics because of minimal domestication, large open-pollinated native populations with limited genetic structure, and high levels of genetic and phenotypic variation. However, most association genetic studies in plants have been hampered by limited gene and polymorphism coverage because of incomplete knowledge of the genetic variants and the low multiplexing capacity of genotyping platforms available to plant species. Consequently, only a fraction of the genetic diversity impacting phenotypic variation has been uncovered for most traits analyzed to date.

This project's main goal is to apply an association genetics approach to unveil the molecular basis of biomass productivity and composition. To comprehensively capture the genetic variants that regulate traits of value for bioenergy production, we are combining sequence-capture and high-throughput sequencing to genotype coding and regulatory sequences in the whole-genome of *Populus deltoides*. To achieve this goal we have (1) optimized sequence-capture for unbiased, high-throughput, and low-cost recovery of target coding and regulatory sequences in *P. deltoides*. A set of over 220,000 probes that efficiently capture exon and 500 bp of putative regulatory sequences of 24,000 genes has been developed so far. Next, we are (2) genotyping a *P. deltoides* unstructured population for association mapping. Oligonucleotides optimized for recovery of target coding and regulatory sequences are being used for sequence capture in 500 individuals of an association population. Captured fragments will be resequenced and polymorphisms genotyped for association analysis. (3) Upon completion of genotyping, we will identify significant SNP-trait associations with biomass growth and carbon partitioning to define genes and alleles that regulate trait variation. Alternative alleles detected in polymorphic sites will be tested individually and in a combined model for marker-trait association to identify the genes that regulate biomass growth and partitioning of carbon into lignocellulosics.

Wednesday, April 13, 10:50 a.m.–11:10 a.m.
Plenary Session: Early Career Program

Insights into the Ecophysiology of Diverse Archaeal-Bacterial Symbioses Mediating the Anaerobic Oxidation of Methane

Victoria Orphan (vorphan@gps.caltech.edu),[1*] Anne Dekas,[1] Abigail Green,[1] Rachel Poretsky,[1] Shawn McGlynn,[1] Jeremy Sandler,[1] Shulei Sun,[2] Mark Ellisman,[2] Tom Deerinck,[2] Robert Hettich,[3] Chongle Pan,[3] and Karuna Chourey[3]

[1] Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena; [2]University of California San Diego, La Jolla; [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.

*Speaker

The overarching project objective is to develop culture-independent methods for studying syntrophic associations between environmental microorganisms, focusing study on the anaerobic oxidation of methane.

Symbiotic methane-consuming associations between uncultured archaea (ANME groups 1–3) and sulfate-reducing bacteria (*Desulfosarcina / Desulfococcus* and members of *Desulfobulbaceae*) are central to carbon and sulfur cycling in hydrate-associated and methane seep environments. The ecophysiology, specificity, and dynamics of these symbiotic associations are poorly understood. Applying a combination of targeted metagenomics using the selective immunofluorescence capture technique Magneto-FISH (Pernthaler et al. 2008), environmental proteomics, and fluorescence *in situ* hybridization coupled with secondary ion mass spectrometry (FISH-nanoSIMS; Dekas et al. 2009), we are expanding our understanding of the range of physiological capabilities and unique species-specific traits of these diverse ANME and sulfate-reducing bacterial consortia. Genomic DNAs were recovered from Magneto-FISH experiments targeting consortia associated with the methanotrophic ANME-2c subgroup, as well as *Desulfosarcina / Desulfococcus* and *Desulfobulbaceae*-associated sulfate-reducing bacteria in samples from methane seep sediments originating from the Eel River Basin in California and Hydrate Ridge in Oregon. Comparative metagenomics of the independent Magneto-FISH pyrosequenced datasets was used to characterize the diversity and specificity among different methanotrophic consortia and to examine their metabolic potential (i.e., nutrient acquisition; utilization of carbon, nitrogen, and sulfur; and electron transfer). Preliminary environmental proteomic analysis of methane seep sediment samples support the *in situ* expression of proteins affiliated with the canonical seven-step methanogenic pathway (likely associated with methanotrophic ANME archaea and methanogens) as well as proteins involved in dissimilatory sulfur metabolism by sulfate-reducing deltaproteobacteria. In addition, the recovery of protein fragments assigned to sulfide-oxidizing gamma- and epsilonproteobacteria suggest the occurrence of an active sulfur cycle within the seep sediments. Using these metagenomic and proteomic data as a guide, sediment microcosm incubation experiments amended with stable isotope–labeled substrates were established and analyzed by FISH-nanoSIMS to test hypotheses regarding the metabolic potential of different ANME/SRB partnerships. Results from experiments using nitrogen 15–labeled substrates (e.g., NH_4^+, NO_3^-, N_2) will be highlighted in this talk, focusing on the observed variation in resource utilization and niche differentiation between the methanotrophic ANME-2 in symbiotic association with members of *Desulfosarcina / Desulfococcus* or *Desulfobulbaceae*. Their potential involvement in benthic nitrogen fixation and nitrate utilization also will be discussed.

References

1. Dekas, A. E., R. S. Poretsky, and V. J. Orphan. 2009. "Deep-Sea Archaea Fix and Share Nitrogen in Methane-Consuming Microbial Consortia," *Science* **326**, 422–26.
2. Pernthaler, A., et al. 2008. "Diverse Syntrophic Partnerships from Deep-Sea Methane Vents Revealed by Direct Cell Capture and Metagenomics," *Proceedings of the National Academy of Sciences USA*, **105**, 7052–57.

Wednesday, April 13, 11:10 a.m.–11:30 a.m.
Plenary Session: Early Career Program

Applying the Biology of Brown Rot Fungi to Consolidated Bioprocessing

Jonathan S. Schilling (schilling@umn.edu)
University of Minnesota

Hypothesis: During brown rot, oxidative pretreatments occur ahead of enzymatic saccharification, spatially, and the fungus partitions these reactions using gradients in pH, lignin reactivity, and plant cell wall porosity. These can be recreated without the fungus present for faster bioconversion, integrating otherwise incompatible steps.

Consolidated bioprocessing (CBP) of lignocellulose combines enzymatic sugar release (saccharification) with fermentation, but pretreatments remain separate and costly. In nature, lignocellulose-degrading brown rot fungi consolidate pretreatment and saccharification, likely using spatial gradients to partition these incompatible reactions. To characterize this relevant biological system, my objectives are to (1) physically sample wood degraded by the brown rot fungus *Postia placenta*, spatially map the coincident locations of pretreatment and saccharification reactions, and correlate with pH and lignin chemistry; (2) to image pH and porosity at the fungus-plant interface and layer this data with images showing cellulase ingress; and (3) map, along the active hyphal front, the co-occurring expression of iron reductases associated with pretreatment and of cellulase involved in saccharification. These are spatially focused goals. Therefore, my respective approaches involve either small-scale, spatially resolved characterization (Objective 1) or appropriately resolved microscopy (Objectives 2 and 3). Small-scale physical sample analysis includes traditional wet chemical characterization, coupled both with spin-trap adduct recovery of hydroxyl radicals produced by the fungus and with C13-labeled tetramethyl ammonium hydroxide thermochemolysis for specific brown rot lignin modifications. For microscopy, I am using fluorescence lifetime imaging (FLIM) with confocal detection for pH measurements, cryo-transmission electron microscopy (TEM) with electron tomography for porosity measures, a complementary scanning transmission X-ray approach for porosity, and traditional TEM with immunolabeling to track cellulase ingress. I am also planning to co-localize chitinous fungal biomass, imaged using a traditional WGA-FITC dye, with fluorescence in situ hybridization to measure mRNA transcribed from iron reductase and endoglucanase DNA sequences, made possible by the recent DOE JGI-funded annotation of the *P. placenta* sequence. Overall, this research will help resolve how brown rot fungi consolidate oxidative pretreatments with enzyme-based saccharification, so that we might better understand and exploit natural synergies between bioconversion steps currently approached as separate, distinct steps.

Wednesday, April 13, 11:30 a.m.–11:50 a.m.
Plenary Session: Early Career Program

Novel Proteomic Approaches for Quantitative Profiling of Important Protein Posttranslational Modifications

Wei-Jun Qian[1*]

Co-authors: Jong-Seo Kim,[1] Lu Wang,[1] Dian Su,[1] Uma K. Aryal,[1] Mary S. Lipton,[1] Richard D. Smith,[1] Ziyu Dai,[2] and Scott E. Baker[2]

[1]Biological Sciences Division, Pacific Northwest National Laboratory (PNNL), Richland, Wash.; [2]Energy and Efficiency Division, PNNL

*Speaker

The overall objective of this early career research project is to develop spatial and temporal proteomic technologies that enable quantitative measurements of protein posttranslational modifications (PTMs) in subcellular compartments to gain an understanding of posttranslational and spatial regulation of cellular machineries on environmental eukaryotic systems. In particular, our efforts are focused on three broad and important classes of PTMs—glycosylation, proteolytic processing, and reversible thiol oxidation—that are ubiquitous in both prokaryotic and eukaryotic cells and play crucial roles in cellular regulation and signaling. In this presentation, we highlight progress on several mass spectrometry–based approaches for quantitative characterization of PTMs in these three classes.

The approaches include site-specific identification of N-linked protein glycosylation sites by integrating hydrazide and lectin chemistries with nano-scale LC and tandem mass spectrometry; identification of proteolytic products or proteolytic sites by selective enrichment of protein N-terminal peptides (i.e., N-terminome); and the characterization of reversible thiol oxidation by selective reduction and subsequent specific enrichment of thiol-containing peptides. We will discuss the concepts behind these novel approaches and present preliminary data and insights gained from studies of model systems, including the filamentous fungus *Aspergillus Niger,* one of the most important microorganisms for biomass degradation and biofuel production.