Sixth Annual DOE Joint Genome Institute User Meeting

Sponsored By

## U.S. Department of Energy Office of Science

March 22-24, 2011 Walnut Creek Marriott Walnut Creek, California

## Contents

Speaker Presentations	1
Poster Presentations	9
Attendees	
Author Index	

## Speaker Presentations

Abstracts alphabetical by speaker

## **Comparative Metagenomics of Gut and Ocean: Identification of Microbial Marker Genes for Complex Environmental Properties**

**Peer Bork** (bork@embl-heidelberg.de)

European Molecular Biology Laboratory, Heidelberg, Germany

Although application of modern sequencing technologies to environmental sequencing<sup>1</sup> enables a wealth of metagenomics data, our understanding of microbial community functioning remains limited, both in terms of internal interactions, but also in the context of environmental properties. Using a metagenomics pipeline, SMASH,<sup>2</sup> we analyzed stool samples from individuals from 6 countries and identified three preferred community compositions, dubbed enterotypes. These are driven by networks of interacting genera and seem to be independent of a number of host properties studied such as nationality, age, gender or body mass index.<sup>3</sup> However, we did find genes or pathways that correlate well with each of the latter properties. Similarly, we also observed adaptation of functional composition of ocean surface communities to various environmental properties related to climate and nutrition. Markers were found even for complex properties such as productivity.<sup>4</sup>

1) Qin et al., Nature. 2010, 464, 59-65.

2) Arumugam, M. et al., Bioinformatics 2010, 26,2977-2978.

3) Arumugam, M. et al. Nature. 2011, in press.

4) Raes, J., Letunic, I., Yamada, T., Jensen, L.L. and Bork, P., Mol.Sys.Biol. 2011, in press.

## Insights from Sequencing the World's Most Diverse Crop: Maize

Edward Buckler (esb33@cornell.edu)

Maize HapMapV2 Project and Maize Diversity Project, Institute for Genomic Diversity, Cornell University, Ithaca, New York

Over the last year, we have combined whole genome sequencing approaches and genotyping-by-sequencing approaches to gain insight into the evolution, domestication, genetic architecture, and now breeding potential of maize. However, there are substantial challenges to working with genomes that have tremendous sequence variability. We found that current pipelines result in two-thirds of the SNP calls coming from paralogues and, hence, we have developed population genetic filters to identify appropriately mapped variation. Through whole genome surveys we have identified 56 million SNPs, and we identified several hundred domestication loci in maize. There were numerous structural variants, which result in with maize lines varying by 30% in genome size, though much of this variation is attributable to a relatively small classes of repetitive elements. We found that 80% of the genome has demonstrable variation in genome size, and through GWAS we found that copy number variants and presence-absence variants were substantially enriched for controlling the phenotypic variation of maize. Shockingly, the ability of the genome to maintain diversity while harboring retroelements is so extreme that sequence comparisons with maize's sister genus found shared genic diversity with maize, despite the

chromosomes having fissioned and most of the retroelements having been replaced. Through the use of genotyping-by-sequencing approaches, we are in the process of genotyping tens of thousands of maize lines for high-resolution genome wide association studies (GWAS) and genomic selection (GS). These studies are beginning to relate this tremendous molecular variation to phenotypic variation and to routes for accelerated breeding.

## How to Eat a Wooden Ship: A Genomic View of a Wood-eating Bacterial Endosymbiosis in the Shipworm *Bankia setacea*

**D.L. Distel**<sup>1</sup> (distel@oglf.org), J. Fung,<sup>1</sup> K. Sharp,<sup>1</sup> B. Henrissat,<sup>2</sup> M. Altamia,<sup>3</sup> E. Lamkin,<sup>4</sup> C. O'Neil,<sup>4</sup> J. Benner,<sup>4</sup> R. Malmstrom,<sup>5</sup> J. Lee,<sup>5</sup> S. Tringe,<sup>5</sup> and T. Woyke<sup>5</sup>

<sup>1</sup>Ocean Genome Legacy, Ipswich, Massachusetts; <sup>2</sup>CNRS & Université de la Méditerranée, Marseille, France; <sup>3</sup>University of the Phillipines, Marine Science Institute, Manilla, Phillippines, <sup>4</sup>Division of Chemical Biology, New England Biolabs, Ipswich, Massachusetts; and <sup>5</sup>DOE Joint Genome Institute, Walnut Creek, California

Shipworms, wormlike wood-boring marine clams of the family *Teredinidae*, are the most prolific consumers of wood in marine environments and are responsible for billions of dollars in damage to wooden ships, piers, and fishing equipment. They are thought to employ a system of wood (lignocellulose) digestion that differs anatomically, functionally, and phylogenetically from all others described to date. The most obvious distinction between the shipworm system and those of termites, ruminants and other previously described lignocellulose-degraders, is that shipworms contain few microbial cells in their digestive tract. Instead, they harbor dense populations of intracellular bacterial endosymbionts within specialized cells of their gills, an organ located far from the gut. It has been proposed that enzymes encoded in and synthesized by the genomes of these gill symbionts are transported to the gut where they contribute to lignocellulose degradation. We have examined the lignocellulolytic system in the shipworm *Bankia setacea* using a combination of highly parallel genomic and metagenomic sequencing and proteomic (2D LC MS/MS) analysis. We provide evidence (1) that the shipworm gill symbiont community, though phylogenetically simple, encodes a rich and diverse variety of lignocellulose-active proteins, including many that are structurally novel, and (2) that a subset of these proteins are selectively transported to the shipworm gut. Thus, the shipworm system appears to be a natural analog of industrial biomass conversion systems that employ separate enzyme synthesis and saccharification. The simplicity of this system, as compared to those of termites or ruminants, makes it highly amenable to genomic and proteomic analyses and genetic manipulation and therefore a potentially informative model for the bioenergy industry.

## The Turn-on of LCLS: The X-Ray Free-Electron Laser at SLAC

Persis Drell (persis@slac.stanford.edu)

SLAC National Accelerator Laboratory, Menlo Park, California

## The Gulf Oil Spill – Ecogenomics and Ecoresilience

#### Terry Hazen (TCHazen@lbl.gov)

Lawrence Berkeley National Laboratory, Berkeley, California, and Joint BioEnergy Institute, Walnut Creek, California

The explosion on April 20, 2010 at the BP-leased Deepwater Horizon drilling rig in the Gulf of Mexico off the coast of Louisiana, resulted in oil and gas rising to the surface and the oil coming ashore in many parts of the Gulf, it also resulted in the dispersment of an immense oil plume 4,000 feet below the surface of the water. Despite spanning more than 600 feet in the water column and extending more than 10 miles from the wellhead, the dispersed oil plume was gone within weeks after the wellhead was capped – degraded and diluted to undetectable levels. Furthermore, this degradation took place without significant oxygen depletion. Ecogenomics enabled discovery of new and unclassified species of oil-eating bacteria that apparently lives in the deep Gulf where oil seeps are common. Using 16s microarrays, functional gene arrays, clone libraries, lipid analysis and a variety of hydrocarbon and micronutrient analyses we were able to characterize the oil degraders. In collaboration with JGI we obtained sequence data. These samples include the deepwater oil plume and uncontaminated water collected at the same depth, as well as contaminated beach samples collected at several time points. Metagenomic sequence data was obtained for the deep-water samples using the Illumina platform. In addition, single cells were sorted and sequenced for the some of the most dominant bacteria that were represented in the oil plume; namely uncultivated representatives of Colwellia and Oceanospirillum. In addition, we performed laboratory microcosm experiments using uncontaminated water collected from The Gulf at the depth of the oil plume to which we added oil and COREXIT. These samples were characterized by 454 pyrotag. The beach samples were also sequenced by 454 pyrotag sequencing and monitored for hydrocarbon degradation. The results provide information about the key players and processes involved in degradation of oil, with and without COREXIT, in different impacted environments in The Gulf of Mexico. We are also extending these studies to explore dozens of deep sediment samples that were also collected after the oil spill around the wellhead. This data suggests that a great potential for intrinsic bioremediation of oil plumes exists in the deep-sea and other environs in the Gulf of Mexico.

### Genomic Analysis of Speciation and Adaptation in Aquilegia

Scott Hodges (hodges@lifesci.ucsb.edu)

University of California, Santa Barbara, California

## The Physiological Genomics of Panicum: Exploring Switchgrass Responses to Climate Change

**Tom Juenger** (tjuenger@austin.utexas.edu)

University of Texas at Austin, Austin, Texas

# Spatially and Temporally Resolved Studies of the Human Microbiome

**Rob Knight** (rob@spot.colorado.edu)

Department of Chemistry & Biochemistry, University of Colorado at Boulder, Boulder, Colorado

We are all >99.9% identical in terms of our human genomes, yet the assemblages of microbial species that outnumber us 10 to 1 in our own bodies can be 90% different. For personalized medicine, it therefore makes sense to look for the sources of individual variation where the variation actually is. In this seminar, I discuss the tools we use to study bacterial communities with next- gen sequencing and some of the remarkable discoveries about how different sites within the same body, and the communities that inhabit different people, actually are. For example, our skin microbial communities are so distinct that people can be traced to the objects they touch. I also describe some of the implications of our studies of obesity in different mouse models for human disease.

#### **Relating Host Genetic Variation to the Microbiome**

Ruth Ley (rel222@cornell.edu)

Cornell University, Ithaca, New York

## **Metatranscriptomics of Marine Microbial Communities**

Mary Ann Moran (mmoran@uga.edu)

University of Georgia, Athens, Georgia

## How to Lose Half Your Genome in 10 million Years and Live to Tell the Tale: Comparative Genomics of *Arabidopsis lyrata* and *A. thaliana*

Magnus Nordborg (magnus.nordborg@gmi.oeaw.ac.at)

Gregor Mendel Institute, Austrian Academy of Sciences, Vienna, Austria

I will present our analysis of the 207 Mb genome sequence of the outcrosser *Arabidopsis lyrata*, which diverged from the self-fertilizing species *A. thaliana* about 10 million years ago. It is generally assumed that the much smaller *A. thaliana* genome, which is only 125 Mb, constitutes the derived state for the family. Apparent genome reduction in this genus can be partially attributed to the loss of DNA from large-scale rearrangements, but the main cause lies in the hundreds of thousands of small deletions found throughout the genome. These occurred primarily in non-coding DNA and transposons, but protein-coding multi-gene families are smaller in *A. thaliana* as well. Analysis of deletions and insertions still segregating in *A. thaliana* indicates that the process of DNA loss is ongoing, suggesting pervasive selection for a smaller genome.

# **Evolutionary Genomic Analyses of Insect Society: Eat, Drink, and Be Scary**

Gene Robinson (generobi@illinois.edu)

University of Illinois at Urbana-Champaign, Urbana, Illinois

## **Remote Detection of Marine Microbes in Coastal Waters and the Deep Sea Using the Environmental Sample Processor (ESP)**

Christopher Scholin (scholin@mbari.org)

Monterey Bay Aquarium Research Institute, Moss Landing, California

The advent of modern molecular biological techniques has revolutionized our understanding of the diversity, function and community structure of marine microorganisms. Increasingly, tools and techniques derived from the biomedical diagnostics and research industries are used in parallel with sensors that characterize the physical, chemical and optical properties of ocean waters. This juxtaposition has fueled the notion of developing an "ecogenomic sensor," a device that could be used to apply molecular analytical techniques in situ as one part of a larger integrated ocean/earth observing network. The Environmental Sample Processor (ESP) was built with this idea in mind – an instrument to help define both the technological and operational elements that underlie the ecogenomic sensor concept. The ESP currently employs low-density DNA probe and protein arrays to assess in near real-time the presence and abundance of specific organisms, their genes and/or metabolites. In addition, a 2-channel real-time PCR module supports deployment of a variety of user-defined master mixes, primer/probe combinations and control templates. The ESP can also be used to preserve samples for a variety of laboratory tests once the instrument is recovered, including metatranscriptomic analyses of natural microbial populations. The ESP is battery operated (12V) and has been deployed on a variety of platforms including coastal moorings, a coastal pier, an open ocean drifter, a research ship, and a 4000m-rated "elevator" compliant with use on deep sea cable observatories. It supports two-way communications for transmitting results of analyses as well as for downloading new instructions so that its mode of operation can be altered remotely. This presentation will highlight the architecture of the ESP and the analytical methods used onboard the instrument, results of recent shallow and deep water field deployments, and an outline of plans for further ESP development. With the expansion of coastal and global ocean observatories, opportunities abound for developing and fielding ecogenomic sensors on a variety of fixed and mobile ocean observing platforms. For the first time, ocean observing systems that allow investigators to carry out interactive experiments and test hypotheses remotely *in situ* from a molecular biological perspective are within reach.

#### **Evolutionary and Comparative Genomics**

Stephen Schuster (scs@bx.psu.edu)

Penn State University, University Park, Pennsylvania

## Designing Biological Systems for Sustainability and Programmed Environmental Interface

Pamela A. Silver (Pamela\_Silver@hms.harvard.edu)

Department of Systems Biology Harvard Medical School and Wyss Institute of Biologically Inspired Engineering, Harvard University, Boston, Massachusetts

Biology presents us with an array of design principles. From studies of both simple and more complex systems, we understand some of the fundamentals of how Nature works. We are interested in using the foundations of biology to engineer cells in a logical way to perform certain functions. By necessity, the predictable engineering of biology requires knowledge of quantitative behavior of individual cells and communities and the ability to construct reliable models. By building and analyzing synthetic systems, we learn more about the fundamentals of biological design as well as engineer useful living devices with myriad applications. For example, we are interested in building cells that can perform specific tasks, such as counting mitotic divisions and remembering past events. Moreover, we design cells with predictable biological properties that serve as potential therapeutics, cell-- based sensors, factories for generating useful commodities including bio-- fuels and improved centers for carbon fixation. In doing so, we have made new findings about how cells interact with and impact on their environment.

## Low Temperature Regulatory Networks Controlling Cold Acclimation in *Arabidopsis*

Michael F. Thomashow (thomash6@msu.edu)

MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, Michigan

Plants from temperate regions increase in freezing tolerance in response to low nonfreezing temperatures, a phenomenon known as cold acclimation. In Arabidopsis, cold acclimation is associated with a major reconfiguring of the transcriptome involving the induction and repression of hundreds of genes. Our goal is to determine the regulatory logic used to bring about these changes in gene expression and to identify the regulons that condition acclimation to low temperature. One important pathway is the CBF cold response pathway which involves action of three transcription factors—CBF1, CBF2 and CBF3—that control expression of more than 100 cold-regulated genes that contribute to freezing tolerance. The analysis of transgenic plants constitutively overexpressing wildtype and dominant negative versions of CBF2 reveal a network of CBF-dependent, CBFcoregulated and CBF-independent cold response pathways and provide evidence for novel pathways that contribute to freezing tolerance. Recent results also establish a role for CAMTA (calmodulin-binding transcription activator) family transcription factors in CBF regulation and cold acclimation and suggest a possible molecular mechanism for the long sought link between low-temperature calcium signaling and cold-regulated gene expression. Finally, we have identified components of the circadian clock that have roles in both circadian regulation and cold induction of the CBF cold response pathway and have found that these clock components are required for plants to attain maximum freezing tolerance at both non-acclimating and cold acclimating temperatures. These results suggest that the integration of cold signaling pathways with the circadian clock may have been an important evolutionary step contributing to plant adaptation to cold environments. The research reported was supported by grants to MFT from the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, U.S. Department

of Energy (DE-FG02-91ER20021); the National Science Foundation Plant Genome Program (DBI 0110124); and the Michigan Agricultural Experiment Station.

## **Tackling Metagenomics' Largest Challenge: The Great Prairie Project**

**Jim Tiedje<sup>1</sup>**\* (tiedjej@msu.edu), Adina Howe,<sup>1</sup> Regina Lamendella,<sup>2</sup> Susannah Tringe,<sup>2</sup> Rachel Mackelprang,<sup>2</sup> C. Titus Brown,<sup>1</sup> Tijana Glavina del Rio,<sup>2</sup> Stephanie Malfatti,<sup>2</sup> Patrick Chain,<sup>3</sup> Nikos Kyrpides,<sup>2</sup> and Janet Jansson<sup>2</sup>

<sup>1</sup>Michigan State University, East Lansing, Michigan; <sup>2</sup>DOE Joint Genome Institute, Walnut Creek, California; and <sup>3</sup>JGI-Los Alamos National Laboratory, Los Alamos, New Mexico

Soil represents the greatest challenge for metagenomics due to the high microbial diversity and complexity of the soil environment. Therefore, JGI has selected soil metagenomics as a Grand Challenge for development and optimization of a metagenomics sequencing, assembly, and annotation pipeline. As a pilot study we are focusing on the Great Prairie of the U.S. which represents the largest expanse of the world's most fertile soils. Samples were collected from paired native prairie and long-term agriculture sites in Wisconsin, Iowa, and Kansas. Two additional sites were sampled in Wisconsin: recently established switchgrass and restored prairie sites that had previously been in continuous corn. We evaluate the effects of stable (native prairie) versus regularly disturbed (cropped) land use on (gene) diversity and selection as well as the effect of restored of grasslands (switchgrass and prairie) on the same parameters. Seven soil cores were collected from each of the eight sampling sites along a 10 m spatial transect, for a total of 56 soil cores. DNA was extracted from each of the cores and the 16S rRNA genes were sequenced using the 454 platform. The resulting pyrotag data clearly demonstrated that soil management altered the composition of the soil microbial communities showing a separation of the native prairie soil microbial communities from the managed soil communities in each state. The Wisconsin switchgrass and restored prairie soil communities clustered more closely to the cultivated corn communities, thus reflecting their previous management history. Metagenome sequencing was also performed on central (reference) cores from each site using the 454 and Illumina (GAII and HiSeq) platforms, which yielded more than a terabase of sequence data. Between 80 - 200 Gigabases of sequence data were obtained per site. The sequencing per se proved to be less challenging than assembly of the sequence data. Several different approaches were used to assemble the data. We focused primarily on assembly of the two Iowa samples, because the Iowa corn samples had two dominant species (*Frauteria* and *Saprospora*) that we hypothesized would simplify the assembly. For one approach to assembly, we use a novel pre-filtering approach to partition the dataset into reads that are likely to assembled together. Using this method, the assembly of Iowa corn Illumina sequence (176 Gb) resulted in 148,053 contigs (>1000 bp). Based on k-mer abundances, we estimate 2-6x maximum coverage for sequencing efforts of Iowa corn soil. Our assembly approach for partitioning large datasets works well, scales to commodity hardware, and has a freely available implementation. Overall, our efforts demonstrate and evaluate the use of next generation sequencing of soils for understanding the biological basis and ecosystem services of its microbial community.

## **Resequencing in Populus: Towards Genome Wide Association Genetic**

Jerry Tuskan (gtk@ornl.gov)

Oak Ridge National Laboratory, Oak Ridge, Tennessee

# Hardware and Software Trends in Computational Systems for Biology

Kathy Yelick (KAYelick@lbl.gov)

National Energy Research Scientific Computing Center (NERSC), Lawrence Berkeley National Laboratory, Berkeley, California

## **Poster Presentations**

Posters alphabetical by first author. \*Presenting author

## Improving the Genome Annotation of Inky Cap Mushroom *Coprinopsis cinerea*

**C.H. Au**,<sup>1</sup> C.K. Cheng,<sup>1</sup> S.K. Wilke,<sup>2</sup> C. Burns,<sup>3</sup> M.E. Zolan,<sup>3</sup> P.J. Pukkila,<sup>2</sup> and H.S. Kwan<sup>1</sup>\* (hoishankwan@cuhk.edu.hk)

<sup>1</sup>The Chinese University of Hong Kong, Hong Kong, China; <sup>2</sup>The University of North Carolina at Chapel Hill, North Carolina; and <sup>3</sup>Indiana University, Indiana

The genome sequence of the model mushroom Coprinopsis cinerea recently published is an important resource in fungal genomics. Gene models currently available are mainly derived from various computer prediction algorithms with help of EST sequences and manual curation. Accuracy of the gene models was evaluated by examining 5' SAGE and microarray datasets we generated and the RNA-Seq dataset from Zemach et al. (Science 328:916-919). Gene model annotations of transcription start sites, 5' and 3' untranslated regions and exons are added or modified. Gene expression levels of different developmental stages are also incorporated. The new genome annotation datasets will be accessible to the research community through a GBrowse-based website, with crossreferences to other genome databases with C. cinerea data. This resource will be useful in C. cinerea functional genomics and comparative genomics of related fungal species. Further investigation of the datasets will also be presented.

## Metagenomic Insights into Lignocellulose Degradation in Leafcutter Ant Fungus Gardens

**Frank O. Aylward**<sup>1,2</sup>\* (faylward@wisc.edu), Kristin E. Burnum,<sup>3</sup> Jarrod J. Scott,<sup>1,2,4</sup> Garret Suen,<sup>1,2</sup> Susannah G. Tringe,<sup>5</sup> Sandra M. Adams,<sup>1,2</sup> Kerrie W. Berry,<sup>5</sup> Carrie D. Nicora,<sup>3</sup> Samuel O. Purvine,<sup>3</sup> Gabriel J. Starrett,<sup>1,2</sup> Lynne A. Goodwin,<sup>5</sup> Richard D. Smith,<sup>3</sup> Mary S. Lipton,<sup>3</sup> and Cameron R. Currie<sup>1,2</sup>

<sup>1</sup>Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin; <sup>2</sup>DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, Wisconsin; <sup>3</sup>Pacific Northwest National Laboratory, Richland, Washington; <sup>4</sup>Smithsonian Tropical Research Institute, Balboa, Ancon, Panama; and <sup>5</sup>DOE Joint Genome Institute, Walnut Creek, California

Leaf-cutter ants are dominant herbivores in Neotropical ecosystems that forage on massive quantities of fresh foliar biomass to cultivate specialized fungus gardens for food. The gardens of these ants are thought to represent external digestive systems that convert recalcitrant plant biomass into forms usable by the ants. To investigate the diversity of microorganisms present in these ecosystems, we conducted metagenomic analyses on fungus gardens collected from two species of leaf-cutter ants in Panama. We show that a diversity of bacteria co-habit fungus gardens with the fungal cultivar, and that the genera *Enterobacter, Klebsiella, Pantoea, Escherichia*, and *Citrobacter* appear to be numerically dominant. Analysis of the lignocellulolytic potential of bacteria.in these ecosystems identified numerous oligosaccharide-degrading enzymes but only few enzymes predicted to target more recalcitrant plant polymers, suggesting that the fungal cultivar may be responsible for the initial breakdown of lignocellulose. Diverse biosynthetic potentials were recovered from the bacterial portion of the metagenomes, suggesting that bacteria

may be responsible for enriching the plant forage of the ants with amino acids, proteins, and B-vitamins. This work provides insight into the process through which herbivores gain access to nutrients in plant biomass by associating with microbial communities.

### **Biopig, a Set of Cloud Computing Apps for Next-Gen Sequence Analysis**

Karan Bhatia\* (kbhatia@lbl.gov) and Zhong Wang

Joint Genomics Institute, Research and Development Group, Lawrence Berkeley National Laboratory, Berkeley, California

Next Generation sequencing is producing ever larger data sizes with a growth rate outpacing Moore's Law. The data deluge has made many of the current sequence analysis tools obsolete because they do not scale with data. Here we present BioPig, a collection of cloud computing tools to scale data analysis and management. Pig is a flexible data scripting language that uses Apache's Hadoop data structure and map reduce framework to process very large data files in parallel and combine the results. BioPig extends Pig with capability with sequence analysis. We will show the performance of BioPig on a variety of bioinformatics tasks, including screening sequence contaminants, Illumina QA/QC, and gene discovery from metagenome data sets using the Rumen metagenome as an example.

## *Nitrosomonas* sp. Is79: An Ammonia Oxidizer Adapted to Low Ammonium Concentrations

**Annette Bollmann**<sup>1</sup>\* (bollmaa@muohio.edu), Jeannette M. Norton,<sup>2</sup> Daniel J. Arp,<sup>3</sup> Martin G. Klotz,<sup>4</sup> Hendrikus J. Laanbroek,<sup>5</sup> Lisa Y. Stein,<sup>6</sup> Yuichi Suwa,<sup>7</sup> Susanne Nielsen,<sup>8</sup> and Lynne A. Goodwin<sup>9</sup>

<sup>1</sup>Miami University, Department of Microbiology, Oxford, Ohio; <sup>2</sup>Utah State University, Department of Biology, Logan, Utah; <sup>3</sup>Oregon State University, Department of Botany and Plant Pathology, Corvallis, Oregon; <sup>4</sup>University of Louisville, Departments of Biology and Microbiology and Immunology, Louisville, Kentucky; <sup>5</sup>Netherlands Institute for Ecology, Department of Microbial Ecology, Wageningen, The Netherlands; <sup>6</sup>University of Alberta, Department of Biological Sciences, Edmonton, Alberta, Canada; <sup>7</sup>Chuo University, Tokyo, Japan; <sup>8</sup>University of Aarhus, Department of Microbial Ecology, Aarhus, Denmark; and <sup>9</sup>DOE Joint Genome Institute, Los Alamos National Laboratory, Los Alamos, New Mexico

Ammonia-oxidizing bacteria (AOB) carry out the first step of nitrification - the oxidation of ammonium to nitrite. AOB are regularly exposed to low substrate concentrations in the environment. Therefore chemostats were used to enrich AOB at low ammonium concentrations from freshwater sediments in the Netherlands. These enrichments resulted in cultures containing members of *Nitrosomonas oligotropha* cluster (Bollmann and Laanbroek, 2001). Subsequent experiments were conducted with the enrichment culture G5-7, which contained a relative of *Nitrosomonas oligotropha*. Competition experiments showed that the enrichment culture G5-7 was able to grow at lower ammonium concentrations than *Nitrosomonas europaea* while *N. europaea* exhibited faster recovery after starvation than G5-7, indicating that the two AOB occupy different niches (Bollmann et al., 2002). *Nitrosomonas* sp. Is79 was isolated from the enrichment culture G5-7 using serial dilution. Growth experiments showed that heterotrophic bacteria and nitrite oxidizers have a positive influence on the growth of *Nitrosomonas* sp. Is79. The K<sub>m</sub> value of *Nitrosomonas* sp. Is79 is comparable to the K<sub>m</sub> values of other AOB from the *Nitrosomonas oligotropha* cluster and lower than the K<sub>m</sub> values of many other AOB's. All

these experiments show the adaptation of *Nitrosomonas* sp. Is79 to low ammonium concentrations. The whole genome has been sequenced recently and will be used to evaluate the basis of the adaptations to low ammonium at genomic and proteomic levels.

Bollmann, A. and H. J. Laanbroek (2001) Continuous culture enrichments of ammoniaoxidizing bacteria at low ammonium concentrations. FEMS Microbial Ecology 37: 211-221.

Bollmann, A.; M.J. Baer-Gilissen, and H.J. Laanbroek (2002) Growth at low ammonium concentrations and starvation response as potential factors involved in niche differentiation among ammonia-oxidizing bacteria. Applied and Environmental Microbiology 68: 4751-4757.

## The WRRC *Brachypodium distachyon* T-DNA Insertional Mutant Population

**Jennifer Bragg**<sup>1</sup>\* (Jennifer.Bragg@ars.usda.gov), Jiajie Wu,<sup>1,2</sup> Yong Gu,<sup>1</sup> Gerard Lazo,<sup>1</sup> Olin Anderson,<sup>1</sup> and John Vogel<sup>1</sup>

<sup>1</sup>USDA-ARS, Western Regional Research Center, Albany, California, and <sup>2</sup>Shandong Agriculture University, Taian, Shandong, China

The model grass *Brachypodium distachyon* (Brachypodium) is an ideal system for studying the basic biology underlying traits of temperate cereals and forage grasses as well those that control the utility of grasses currently being developed as energy crops. Using our *Agrobacterium tumefaciens*-mediated high-efficiency transformation method (average efficiency 44%), we have have generated >8,700 Brachypodium T<sub>0</sub> lines. We have assigned 8092 flanking sequence tags (FSTs) in 4402 insertional mutant lines to specific locations in the Brachypodium genome. More than 1300 unique genes have been tagged in this population. A profile describing the distribution of the FSTs within the Brachypodium genome will be presented. The success of this project has resulted in plans to generate an additional 30,000 lines for this collection. Information about the WRRC Brachypodium insertional mutant population is available in a searchable website designed to provide researchers with a means to order T-DNA lines with mutations in genes of interest. Protocols for working with Brachypodium, information about the T-DNA project, and instructions for ordering T-DNA lines are available at http://brachypodium.pw.usda.gov.

## Near Real-time Analysis of Streaming DNA Sequence Data using Message Oriented Middleware

**Thomas Brettin**\* (brettintx@ornl.gov), Craig Cunic, Ray Easterday, Daniel Quest, and Robert Cottingham

Oak Ridge National Laboratory, Oak Ridge, Tennessee

Achieving near real-time analysis of DNA sequence reads is required to keep pace with advancing sequencing technologies and will depend on changes to the underlying computational methodology. Genomic sequencing throughput per unit cost continues to exceed Moore's Law and all indications are that this rate of advance will continue into the future. The BioSITES (Biological Signature Identification and Threat Evaluation System) system was proposed in 2009 to leverage recent technological and scientific advances to improve detection of microbes in the growing number of sequence data sets. As part of

that system, an architecture and implementation have been developed. BioSITES is a novel project at Oak Ridge National Laboratory (ORNL) that can provide near real-time integration of DNA and non-DNA sequence data. BioSITES currently integrates DNA sequences with spatial and temporal data using advanced data models. The architecture is based on the integration of systems biology knowledge repositories (catalogs), sequence processing algorithms (sensors) that subscribe to specific streams of sequence data, expert systems for integrating information from multiple sources (controllers), and Web 2.0 technologies for reporting threats to analysts (advisories). Current run-time characteristics of the architecture on 1088 cores demonstrates the scalability of the system as well as limitations of standard configurations of message oriented middleware.

## A Tale of Two Corals: A Large-scale Transcriptomics Synthesis to the Study of Coral-algal Symbiosis Establishment and Disruption (Bleaching) in *Montastraea faveolata* and *Acropora palmata*

**E. Buschiazzo**<sup>1</sup>\* (ebuschiazzo@ucmerced.edu), S. Sunagawa,<sup>2</sup> M. K. DeSalvo,<sup>3</sup> C.R. Voolstra,<sup>4</sup> M. Aranda,<sup>4</sup> T. Bayer,<sup>4</sup> and M. Medina<sup>1</sup>

<sup>1</sup>School of Natural Sciences, University of California, Merced, California; <sup>2</sup>European Molecular Biology Laboratory, Meyerhofstraße, Heidelberg, Germany; <sup>3</sup>Department of Anesthesia and Perioperative Care, University of California, San Francisco, California; and <sup>4</sup>King Abdullah University for Science and Technology, Thuwal, Saudi Arabia

Corals are the building blocks of one of the most biodiverse and dynamic ecosystems in the world. The intricate relationship that evolved between the coral polyp and its endosymbiotic microalgae (genus Symbiodinium) is key to many of the biological processes that allow corals to thrive in their natural environment. In particular, corals offer protection and access to light, while microalgae provide readily available and energy-rich products of photosynthesis. In times of stress (e.g. high temperature and high UV radiation), this relationship may be disrupted; the host expels most or all of its colored endosymbionts, and the bright white coral skeleton is then revealed through the now transparent tissue (i.e. appearing to be freshly "bleached"). The coral may recover or die, depending on whether or not conditions returned to "normal" and the associated Symbiodinium spp. are present in the environment and successfully re-integrate into the host tissue. Recently, three independent studies based on medium-scale cDNA microarrays have sought to follow gene expression during the early establishment and the disruption (thermostress-related bleaching) of the coral-algal symbiosis in two threatened Caribbean corals, Montastraea faveolata and Acropora palmata. An additional study has also investigated the transcriptomics of post-bleaching recovery in *M. faveolata*. While these reports uncovered at least some of the genes, pathways and processes that may be crucial to the specific establishment of the coral-algal symbiosis, its disruption and its recovery, they were limited by the size of the cDNA microarrays (1,314 and 2,055 cDNAs for M. *faveolata* and A. *palmata*, respectively). Here, we report the development of two large large-scale microarrays (11,216 and 14,400 cDNAs, respectively), which were used to hybridize cDNA from the same samples used in those previous experiments, and we synthesize our results in a single transcriptomics study of the evolution of coral-algal symbiosis from establishment through to disruption (A. palmata and M. faveolata) and recovery (*M. faveolata* only).

# Engineering of Stress Tolerance by Introducing *GT78* Gene Sequence from *Selaginella moellendorffii* into Higher Plants

**Ulla Christensen**<sup>1</sup>\* (uchristensen@lbl.gov), Peter Benke,<sup>1</sup> Agnieszka Zygadlo Nielsen,<sup>2</sup> Bodil Joergensen,<sup>2</sup> Jesper Harholt,<sup>2</sup> Patrick Canlas,<sup>3</sup> Pamela C. Ronald,<sup>3</sup> Peter Ulvskov,<sup>2</sup> and Henrik Vibe Scheller<sup>1</sup>

<sup>1</sup>Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California; <sup>2</sup>Department of Plant Biology and Biotechnology, University of Copenhagen, Faculty of Life Sciences, Copenhagen, Denmark; and <sup>3</sup>Department of Plant Biology and Pathology, University of California, Davis, California

A comparative genomic study of the recent JGI sequenced *Selaginella moellendorffii* (a spikemoss) and *Physcomitrella patens* (a moss) revealed the presence of sequences encoding members of the Glycosyltransferase family 78 (GT78). GT78 genes are not found among higher plants. However, the GT78 proteins have been found in the themophilic bacterium *Rhodothermus marinus* and the red algae *Griffithsia japonica*. The *S. moellendorfii GT78* gene encodes a mannosylglycerate synthase, which catalyzes the synthesis of 2-O- $\alpha$ -D-mannosyl-D-glycerate (MG) from GDP-mannose and glycerate. MG is a widespread compatible solute among thermophilic and hyperthermophilic bacteria and archaea. Like other compatible solutes, MG is a low molecular mass organic compound that accumulates in response to environmental stress such as salinity, drought and elevated temperature. The intracellular concentration of these types of molecules is high but they are compatible with the metabolism of the cell and provide cell stability and thermo protectance.

The presence of MG synthase in a terrestrial plant suggested that this stress tolerance strategy could function in higher plants and increase tolerance to environmental stress. To test this idea, the *GT78* gene from *S. moellendorfii* was codon optimized for expression in higher plants and introduced into *Arabidopsis thaliana*, *Oryza sativa* and *Brachypodium distachyon*. Transgenic lines with inducible and constitutive expression of GT78 have been verified and the MG has been detected using LC/MS metabolic profiling.

# Illumina GA IIx & HiSeq 2000 Production Sequencing and QC Analysis Pipelines at the DOE Joint Genome Institute

**Chris Daum**\* (cgdaum@lbl.gov), James Han,<sup>1</sup> Matt Zane,<sup>1</sup> Angela Tarver,<sup>2</sup> Alex Copeland,<sup>2</sup> Mingkun Li,<sup>2</sup> JGI Production Illumina Sequencing & Rolling QC Teams, and Susan Lucas<sup>1</sup>

<sup>1</sup>Lawrence Livermore National Laboratory, Livermore, California, and <sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, California, and DOE Joint Genome Institute, Walnut Creek, California

The U.S. Department of Energy (DOE) Joint Genome Institute's (JGI) Production Sequencing group is committed to the generation of high-quality genomic DNA sequence to support the mission areas of renewable energy generation, global carbon management, and environmental characterization and clean-up.

Within the JGI's Production Sequencing group, a robust Illumina Genome Analyzer and HiSeq pipeline has been established. Optimization of these sequencer pipelines has been ongoing with the aim of continual process improvement of the laboratory workflow, reducing operational costs and project cycle times to increase sample throughput, and

improving the overall quality of the sequence generated. A sequence QC analysis pipeline has been implemented to automatically generate read and assembly level quality metrics.

The foremost of these optimization projects, along with sequencing and operational strategies, throughput numbers, and sequencing quality results will be presented.

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH112.

IM release number: LLNL-POST-468754

#### A Proposal for a New Finishing Standard

**Karen Davenport**<sup>1</sup>\* (kwdavenport@lanl.gov), Hajnalka Daligault,<sup>1</sup> Lynne Goodwin,<sup>1</sup> Linda Meincke,<sup>1</sup> Olga Chertkov,<sup>1</sup> Tanja Woyke,<sup>2</sup> Cliff Han,<sup>1</sup> and Chris Detter<sup>1</sup>

<sup>1</sup>Los Alamos National Laboratory, Los Alamos, New Mexico, and <sup>2</sup>DOE Joint Genome Institute, Walnut Creek, California

As massively parallelized sequencing technologies enable draft genome sequencing at minimal cost, the cost for closing gaps in the draft genome becomes disproportionally high. Ribosomal DNA sequences are important for identification and classification of organisms and transposon sequences are important for genome evolution study, but these are not generally useful in most biological studies of a bacterial species. The current finishing standard requires each copy of the same transposon or ribosomal DNA region to be finished separately. The cost for finishing repeat regions represents a significant portion of the total finishing cost. We are exploring ways to reduce finishing costs while still providing the genome sequences with quality high enough for determining gene functions in microorganisms. One possibility could be to insert a consensus sequence for rDNA and transposon regions to eliminate the costs of finishing these regions, while repetitive regions with other functional genes will be finished with current finishing methods. The final product of the modified finishing process would be a contiguous sequence for each replicon with unique regions finished to a high quality standard (< 1 errors / 10 kb) while repetitive rDNA regions and transposon regions would be finished to a degree that will assure the correctness of the chromosomal structure. According to our analysis, this would significantly reduce finishing costs while enabling metabolic reconstruction, fulfilling the key need of the user community.

#### Catabolism of Plant Biomass by Streptomyces Bacteria

Jennifer R. Davis\* (jennifer\_reeve@brown.edu) and Jason K. Sello

Brown University, Providence, Rhode Island

The search for a renewable energy source to act as an alternative to fossil fuel is of global importance. The use of plant biomass as a source of low-value carbon that can subsequently be used to produce high-value biofuels has shown great potential. Most attention is focused on the conversion of the cellulose component of plant biomass; however, this process is impeded by the presence of lignin, a complex aromatic polymer found in the cell walls of plants. The means to effectively and efficiently degrade lignin would enhance the ability to harness the energy stored in plant biomass into a usable fuel. Several ligninolytic species of *Streptomyces* bacteria have been identified, including *S. viridosporus, S. badius,* and *S. setonii.* Although it is known that members of the

Streptomyces genus are able to depolymerize lignin and catabolize its products, little is known about the underlying genetics and biochemistry. In collaboration with the Joint Genome Institute (JGI), the genomes of S. viridosporus and S. setonii are currently being sequenced in an effort to identify the enzymatic machinery for lignin depolymerization. In parallel, we are also characterizing the major catabolic pathway for lignin-derived aromatic compounds in microorganisms, the β-ketoadipate pathway. Through this pathway, the aromatic compounds (*i.e.*, protocatechuate and catechol) are converted to acetyl coenzyme A and succinyl coenzyme A. We have found that transcription of genes encoding enzymes of the protocatechuate branch of the  $\beta$ -ketoadipate pathway are induced by protocatechuate in *Streptomyces coelicolor*. Disruption of a gene, *pcaV*, encoding a MarR family transcription factor resulted in constitutive transcription of the genes, indicating that it acts as a transcriptional repressor. We hypothesize that the PcaV transcription factor regulates transcription of the catabolic genes by binding to their promoter in a protocatechuate-dependent fashion. Currently, biochemical and structural studies of PcaV are underway to elucidate the mechanism by which its DNA binding is affected by protocatechuate. Our combined genomic, biochemical and structural analyses will provide critical insights into bacterial consumption of the lignin component of plant biomass.

#### Feedstock-adapted Anaerobic Consortia Derived from Tropical Forest Soils

**Kristen M. DeAngelis**<sup>1,2</sup>\* (kdeangelis@lbl.gov), Julian Fortney,<sup>1</sup> Sharon Borglin,<sup>1</sup> Whendee Silver,<sup>1,3</sup> and Terry C. Hazen<sup>1,2</sup>

<sup>1</sup>Ecology Department, Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; <sup>2</sup>Microbial Communities Division, Joint BioEnergy Institute, Emeryville, California; and <sup>3</sup>Department of Ecosystem Science, Policy and Management, University of California, Berkeley, California

Tropical soils in Puerto Rican rain forests are capable of deconstructing biofuel plant materials to basic components, and frequent episodes of anoxic conditions make it likely that these decomposing consortia are primarily bacteria, not fungi as are usually observed in temperate systems. We cultivated feedstock-adapted anaerobic consortia (FACs) derived from Puerto Rico forest soils and added the terminal electron acceptors nitrate, sulfate, or iron to examine the effect on switchgrass deconstruction. Soils from two forest types were used as inoculum; short cloud forest (SCF) soils are perennially soaked and anaerobic, while Bisley Ridge soils (BisR) are more iron-rich and experience fluctuating redox from completely oxic to completely anoxic over a range of hours to days. Soil communities were anaerobically passaged through a succession of transfers in minimal media with switchgrass as the sole carbon source. Metagenomic analysis was performed on third-transfer FACs derived from BisR inoculum, from switchgrass only and ironamended microcosms, revealing that the iron-amended FAC contained 324 distinct taxa compared to 81 taxa in the unamended. The soil-only FACs from BisR and SCF were used as inoculum for nitrate-, iron-, sulfate-, switchgrass- and double-switchgrass-amended FACs, which were passaged once more with the additional TEAs and then characterized. Based on methane and carbon dioxide production rates, nitrate and iron caused the highest C mineralization in BisR-FACs, while switchgrass alone had the highest C mineralization in SCF-FACs. Specific enzyme activity rates were higher overall in SCF-FACs compared to BisR-FACs, perhaps a reflection of the need for periodic oxygen availability in BisR soils that is absent in SCF soils. Microbial community profiling was performed using PLFA and pyrotag sequencing of the small subunit ribosomal RNA gene, revealing

Actinobacteria and Gammaproteobacteria as dominant organisms. The diversity of anaerobic degraders found in these soils reiterates the importance of anaerobic decomposition in these environments and highlights the potential for discovery. Functional and phylogenetic screening will indicate target samples for future metagenomics, with the goal of discovering the enzymes responsible for switchgrass decomposition in the anaerobic FACs.

### The FUNG-GROWTH Database: Linking Growth to Genome

**Ronald P. de Vries**<sup>1</sup>\* (r.devries@cbs.knaw.nl), A. Wiebenga,<sup>1</sup> Vincent Robert,<sup>1</sup> Pedro M. Coutinho,<sup>2</sup> and Bernard Henrissat<sup>2</sup>

<sup>1</sup>CBS-KNAW Fungal Biodiversity Centre, Utrecht, The Netherlands, and <sup>2</sup>AFMB, Marseille, France

Fungal genome sequences demonstrate the potential to utilize a variety of different carbon sources. Natural carbon sources for many fungi are based on plant biomass and often consist of polymeric compounds, such as polysaccharides. They cannot be taken up by the fungal cell and are extracellularly degraded by a complex mixture of enzymes. Plant polysaccharide degrading enzymes have been studied for decades due to their applications in food and feed, paper and pulp, beverages, detergents, textile and biofuels. These enzymes have been classified based on amino acid sequence modules (www. cazy.org).

Based on the hypothesis that fungal genomes have evolved to suit their ecological niche, we have performed a comparative study using >60 fungal species. In this study we have compared growth profiles on 36 different carbon sources (consisting of mono-, oligo- and polysaccharides, lignin, protein and crude plant biomass) to the CAZy annotation of the genomes to identify correlations between growth and genomic potential.

Highlights of these comparisons will be presented as well as the importance of growth profiling for both fundamental and applied fungal research. The data from our study is accessible through a public database that will also be presented in this presentation.

## Characterization of CAZY-like Enzymes from HT Sequencing of Microbial Communities

**Michael J. Dougherty**<sup>1</sup>\* (mjdougherty@lbl.gov), Patrik D'haeseleer,<sup>2</sup> Blake A. Simmons,<sup>2</sup> Paul Adams,<sup>1</sup> and Masood Hadi<sup>1</sup>

<sup>1</sup>Technology Division and <sup>2</sup>Deconstruction Division, Joint BioEnergy Institute, Emeryville, California

Advances in sequencing technology have made metagenomic analysis of complex environmental samples feasible, and these analyses have generated a vast amount of diverse sequence information, unlocking potential new biocatalysts useful for biotechnology applications. However, moving from gene sequences to functional characterization is still a major bottleneck due to the effort required to identify appropriate expression hosts and/or conditions for protein purification and characterization.

The development of processes for the conversion of biomass into fuels and chemicals is currently a major research goal. The conversion of lignocellulosic biomass into fermentable sugars via chemical pretreatment and enzymatic saccharification is currently one of the most expensive steps in processes for biofuel production. Identifying and/or engineering glycoside hydrolases (GHs) and other carbohydrate-active enzymes with improved enzymatic properties is a major research challenge in this effort. We have performed a metagenomic analysis of a switchgrass-adapted compost microbial community and identified genes in this community that putatively encode enzymes with diverse activities, including endoxylanase,  $\beta$ -xylosidase, and  $\alpha$ -arabinofuranosidase.

In order to validate the metagenomic approach for finding new biocatalysts for biomass deconstruction these ORFs have been cloned, expressed, and assayed for various hemicellulase activities which can then be prioritized for diverse biofuel process conditions. The most promising candidate genes have been characterized in more detail focusing on properties important for the process of biomass hydrolysis, such as thermostability, pH dependence, and ionic liquid tolerance. These gene products include GH43 family bifunctional  $\beta$ -xylosidase/ $\alpha$ -arabinofuranosidases and endoxylanases from the GH10 and GH11 families. Some of these enzymes may be the starting points for further protein engineering towards the goal of developing enzyme cocktails for various process conditions.

## Metagenomic Gene Annotation by a Homology Independent Approach

**Changbin Du**<sup>1,2</sup>\*(cdu@lbl.gov), Jeff Froula,<sup>1,2</sup> Tao Zhang,<sup>1,2</sup> Annette Salmeen,<sup>1,2</sup> Matthias Hess,<sup>3</sup> Cheryl A. Kerfeld,<sup>1,2,4</sup> and Zhong Wang<sup>1,2</sup>

<sup>1</sup>Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California; <sup>2</sup>DOE Joint Genome Institute, Walnut Creek, California; <sup>3</sup>Washington State University, Department of Molecular Biosciences, Richland, Washington; and <sup>4</sup>Department of Plant and Microbial Biology, University of California, Berkeley, California

Fully understanding of the genetic potential of a microbial community requires functional annotation of all the genes it encodes. The recently developed deep metagenome sequencing approach has enabled rapid identification of millions of genes from a complex microbial community without cultivation. Current homology-based gene annotation fails to detect distantly-related or structural homologs. Furthermore, homology searches with millions of genes are very computational intensive.

To overcome these limitations, we developed rhModeller, a homology independent software pipeline to efficiently annotate genes from metagenomic sequencing projects. Using cellulases and carbonic anhydrases as two independent test cases, we demonstrated that rhModeller is much faster than HMMER but with comparable accuracy, at 94.5% and 99.9% accuracy, respectively. More importantly, rhModeller has the ability to detect novel proteins that do not share significant homology to any known protein families.

As  $\sim$ 50% of the 2 million genes derived from the cow rumen metagenome failed to be annotated based on sequence homology, we tested whether rhModeller could be used to annotate these genes. Preliminary results suggest that rhModeller is robust in the presence of missense and frameshift mutations, two common errors in metagenomic genes. Applying the pipeline to the cow rumen genes identified 4,990 novel cellulases candidates and 8,196 novel carbonic anhydrase candidates.

In summary, we expect rhModeller to dramatically increase the speed and quality of metagnomic gene annotation.

## Genomics and Systematics of Wood Decay by the White Rot Fungus *Phlebia radiata*

**Jaana Ekojärvi**<sup>1</sup>\* (jaana.ekojarvi@helsinki.fi), Miia Mäkelä,<sup>1</sup> Ilona Oksanen,<sup>1</sup> Pia K. Laine,<sup>2</sup> Lars Paulin,<sup>2</sup> Petri Auvinen,<sup>2</sup> and Taina Lundell<sup>1</sup>

<sup>1</sup>Department of Food and Environmental Sciences, Division of Microbiology, Fungal Biotechnology Laboratory and <sup>2</sup>Institute of Biotechnology, DNA Sequencing and Genomics Laboratory, University of Helsinki, Helsinki, Finland

*Phlebia radiata Fr.* is a saprobic filamentous fungal species belonging to the family Corticiaceae in the class Agaricomycetes (Basidiomycota). It is a common Eurasian species able to cause a white rot type of decay both in dead hardwood (angiosperms) and softwood (conifers). *P. radiata* is able to decompose lignin in wood and non-wood plant matter, as well as to convert and mineralize synthetic lignin, lignin model compounds and xenobiotics. The species secretes a number of lignin-modifying enzymes (LMEs), for example isozymes of manganese and lignin peroxidases, and at least one laccase. As the fungus secretes also  $H_2O_2$ -producing glyoxal oxidase and a set of cellulases and hemicellulases (CAZymes) it is capable of degrading all the main components of wood. Because of such large capability for conversion of wood, lignocelluloses, and even harmful organic compounds, the Finnish isolate 79 (FBCC43) is selected as a model organism for studying fungal biodegradation of wood. In this project, brown rot fungal species are studied in parallel to gain more knowledge on the variations of fungal wood decay strategies.

Our first goal has been to gain about 17 x coverage of the genome of *P. radiata* by direct 454 pyrosequencing and Titanium chemistry, and completion of the draft sequencing data into scaffolds by mate-pair library sequencing using SOLiD is ongoing. First, particular emphasis has been attained for completion and annotation of the mitochondrial genome (see abstract by Ilona Oksanen et al.).

Besides the ongoing whole genome sequencing, our aim is to study gene and protein expression, when the wood-decaying white and brown rot fungi are growing on wood and various plant lignocelluloses. Expression of LME marker genes for lignin and wood decay will be followed in different time points by real-time quantitative RT-PCR, and transcriptome analyzed by cDNA and RNA sequencing. Finally, secretome proteomes during wood and lignocellulose decay will be mined by peptide sequencing. In addition, molecular systematics of the obviously polyphyletic genus *Phlebia* is studied in order to better classify the various species, and to describe their lignocellulose-decomposition types.

## Genetic Variation and Evolution of Perchlorate Reductase and Chlorite Dismutase: The Central Enzymes Involved in Microbial Dissimilatory (Per)chlorate Reduction

**Anna L. Engelbrektson**\* (aengelbrektson@berkeley.edu), Kathryne G. Byrne-Bailey, Eunice Moon-Lim, Antinea H. Chair, Suzanna Repo, Steven E. Brenner, and John D. Coates

Department of Plant and Microbial Biology, University of California, Berkeley, California

Perchlorate (ClO<sub>4</sub><sup>-</sup>) is a world-wide environmental contaminant, posing a significant health threat. Numerous perchlorate reducing microorganisms from across the phylum Proteobacteria have been isolated. The two proteins essential for this metabolism are

perchlorate reductase (Pcr: pcrABCD) and chlorite dismutase (Cld: cld). Despite their metabolic interdependency these two genes have separate regulatory systems for expression. Nothing is known about the evolution and acquisition of these genes in the environment. DNA and amino acid sequence data for the Pcr subunits, Cld (using Pfam (PF06778) "cld-like gene"), as well as the 16S rRNA has been collected from a wide number of environmental isolates including Azospirillum, and Dechloromonas species, as well as sequenced bacterial and archaeal genomes from species not experimentally confirmed to perform this metabolism. Interestingly, Pcr were not found in many of the same genomes as Cld hinting at a potential multi-functionality of the Cld-like proteins. Phylogenetic trees for each subunit were used to compare gene disparity across different organisms and their evolutionary relationship relative to the 16S rRNA. Sequence similarity did not mirror the evolutionary relationship established by the 16S rRNA gene indicating gene acquisition through horizontal transfer with each of the separate subunits taking divergent evolutionary paths. This was also supported by phylogenetic trees where known (per)chlorate reducing bacteria ((P)CRB) claded independently from non/untested-(P)CRB bacteria. Studying Pcr and Cld in environmental isolates as well as environmental samples creates a framework for investigating horizontal transfer of these genes, whether different types of the subunits exist in certain niches and whether genes are lost, gained or swapped. Furthermore, these studies may indicate how these proteins evolved and their origins may be established.

#### **Rare Variant Detection Bacterial Samples: Feasibility Study**

**V.Y. Fofanov**,<sup>1</sup> J. Liu,<sup>1</sup> J. Howard,<sup>2</sup> T. Constantin<sup>1</sup>\* (tudor@eurekagenomics.com), M. Shin,<sup>1</sup> H. Koshinsky,<sup>1</sup> and Y. Fofanov<sup>1,2</sup>

<sup>1</sup>Eureka Genomics Corp., Hercules, California, and <sup>2</sup>University of Houston, Houston, Texas

Bacterial stocks from different origins can be differentiated based on the presence of rare variants and could act as a sample's fingerprint and be important in investigation and prosecution of bioterrorism attacks or attempts. Existing approaches cannot detect unknown variants if present in less than 10% of the sample. The recent advances in High Throughput Sequencing (HTS) technologies may have sufficient throughput to make feasible the detection of such rare variants.

We have conducted several computational simulations, as well as a number of deep sequencing experiments on Illumina's GAIIx using *Enterobacteria phage phiX174* (NC\_001422), *Pseudomonas aeruginosa* (NC\_008463), and PCR amplified regions from *Mycobacterium tuberculosis*. Experimental data suggests that while HTS can produce sufficient sequence information required to successfully detect rare variants present in as low as 0.01% of the sample, distinguishing true rare variants from false positive rare variants remains a major limiting factor. However, carefully controlling the factors contributing to false positive rare variant detection, including systematic sources of error such as machine sequencing error, alignment algorithm errors, and bias induced by features of the reference sequence itself, allowed us to attach sensitive and specific identification of rare variants present in as low as 0.1% of the sample.

## Effect of Read Length, Coverage, and Fragment Size on the Quality of de novo Assemblies of Bacterial Genomes

**V.Y. Fofanov**, J. Liu, M. Shin\* (maria@eurekagenomics.com), N. Bulsara, T. Constantin, and H. Koshinsky

Eureka Genomics Corp., Hercules, California

High Throughput Sequencing (HTS) has significantly increased the pace of genomic data generation; billions of bases of sequence information can be generated in a single run. One of the common applications of HTS is the generation of assembled draft reference for bacterial genomes. A common question is: "How much sequence data and of what type will result in the best assembly?" Or in other words, "For a given amount of data, what level of coverage, read length, read type and insert (gap) size will result in the assembly that most closely resembles the genome of the organism?" We have performed an empirical study to <u>quantify</u> the effects of the amount and type (read length and, in the case of paired reads, insert size) of the sequence data on the quality and length of the scaffolds and genome assembly.

Specifically, we investigated the effects of read lengths (34 - 100 bases), coverage (15X – 105X) and read types (single reads and paired reads with either short (200 bp average) or long (5 kb average) insert length) on the quality of assemblies for *Pseudomonas aeruginosa* (low number of repeated regions – generally considered easy to assemble) and *Streptococcus pneumoniae* (high number of repeated regions – generally considered difficult to assemble). The experimental data allowed us to quantify and rank the contribution of read length, coverage, and read type on the length (measured by N50) and quality (measured by numbers of single nucleotide errors and mis-assemblies) of bacterial assemblies.

## High Fiber Degrading Activity in the Particle-associated Microbiota of the Hoatzin Crop

**Filipa Godoy-Vitorino**,<sup>1</sup> Stephanie Malfatti<sup>1</sup>, Maria A. Garcia-Amado<sup>2</sup>, Maria Gloria Dominguez-Bello<sup>3</sup>, Philip Hugenholtz<sup>1</sup>\* (phugenholtz@lbl.gov), and Susannah G. Tringe<sup>1</sup>\* (sgtringe@lbl.gov)

<sup>1</sup>Microbial Ecology Program, DOE Joint Genome Institute, Walnut Creek, California; <sup>2</sup>Instituto Venezolano de Investigaciones Científicas (IVIC), Caracas, Venezuela; and <sup>3</sup>Department of Biology, University of Puerto Rico, Rio Piedras Campus, San Juan, Puerto Rico

The hoatzin (*Opisthocomus hoazin*) is a South American herbivorous bird that has an enlarged crop analogous to the rumen, where foregut microbes degrade the otherwise indigestible plant materials, providing energy to the host. The crop harbors an impressive array of microorganisms with potentially novel cellulolytic enzymes.

Our study describes the composition of the particle-associated microbiota in the hoatzin crop, combining a survey of 16S rRNA genes in 7 adult birds and metagenome sequencing of two animals.

Our pyrotag survey demonstrates that the particle-associated microbiota have a variety of bacterial phyla (~31 with  $\geq 0.1\%$  relative abundance) including the dominant Bacteroidetes, Firmicutes, Actinobacteria and Proteobacteria. Members of Prevotellaceae, are the most abundant and ubiquitous taxa, suggesting that the degradation of hemicellulose is an important activity in the crop.

Nonetheless, preliminary results from the metagenome of the particle-associated microbiota of two adult birds show that the crop microbiome contains a high number of genes encoding cellulases (such as GH5) more abundant than those of the termite gut, as well as genes encoding endo-1,4-b-galactanase (hemicellulases).

These preliminary results show that the carbohydrate-active enzyme genes in the crop metagenome could be a source of biochemical catalysts able to deconstruct plant biomass.

### Agave: A Biofuel Feedstock for Arid and Semi-arid Environments

Stephen Gross<sup>1</sup>\* (smgross@lbl.gov) and Axel Visel<sup>1,2</sup>

<sup>1</sup>DOE Joint Genome Institute, Walnut Creek, California, and <sup>2</sup>Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California

Efficient production of plant-based, lignocellulosic biofuels relies upon continued improvement of existing biofuel feedstock species, as well as the introduction of new feedstocks capable of growing on marginal lands to avoid conflicts with existing food production and minimize use of water and nitrogen resources. To this end, species within the plant genus Agave have recently been proposed as new biofuel feedstocks. Many agave species are adapted to hot and arid environments generally unsuitable for food production, yet have biomass productivity rates comparable to other second-generation biofuel feedstocks such as switchgrass and *Miscanthus*. Agaves achieve remarkable heat tolerance and water use efficiency in part through a Crassulacean Acid Metabolism (CAM) mode of photosynthesis, but the genes and regulatory pathways enabling CAM and thermotolerance in agaves remain poorly understood. We seek to accelerate the development of agave as a new biofuel feedstock through genomic approaches using massively-parallel sequencing technologies. First, we plan to sequence the transcriptome of A. tequilana to provide a database of protein-coding genes to the agave research community. Second, we will compare transcriptome-wide gene expression of agaves under different environmental conditions in order to understand genetic pathways controlling CAM, water use efficiency, and thermotolerance. Finally, we aim to compare the transcriptome of A. tequilana with that of other agave species to gain further insight into molecular mechanisms underlying traits desirable for biofuel feedstocks. These genomic approaches will provide sequence and gene expression information critical to the breeding and domestication of Agave species suitable for biofuel production.

## Using Existing Liquid-Handling Robotic Technology to Construct Indexed Unamplified Illumina Tru-Seq Shotgun Libraries

Christopher A. Hack\* (cahack@lbl.gov) and Jan-Fang Chang

DOE Joint Genome Institute, Walnut Creek, California

Recently, Illumina released its TruSeq kits for indexed library construction. In the wake of that release, several companies (including Illumina) have announced plans to produce automated liquid-handling platforms specifically designed to construct libraries using the TruSeq reagents and modified versions of the TruSeq protocol. These platforms vary in cost, throughput, and amount of required user intervention. Here we present the automated construction of TruSeq libraries using an existing robotic platform: the Beckman-Coulter Biomek FX. With this platform, we have constructed 24 libraries in parallel, and the method could theoretically be used to construct up to 96 libraries at one time. Using deep-

well reagent blocks instead of reservoirs and employing disposable pipette tips minimizes the risk of sample cross-contamination. 12 unique adapters with index sequences provided by Illumina are used in library construction enabling up to 12 libraries to be pooled in a single lane of an Illumina chip, and it is possible to create enough unique index sequences to enable the pooling of 96 samples. The libraries are created without PCR amplification, reducing construction time and eliminating any sequencing bias introduced at the amplification step. Furthermore, the high stability of the reagents used permits the construction method to proceed with minimal user intervention; the entire process time of sheared fragments to final library output is less than 5 hours, with only one user intervention step required once the process has begun. It is hoped that using robotic technology already in hand will save the facility library construction platform costs and enable us to implement automated indexed unamplified Illumina TruSeq library construction quickly into our production line. The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

#### **Structural Comparisons of Plant Glycosyltransferases**

Sara Fasmer Hansen\* (sfhansen@lbl.gov), Ryan McAndrew, Paul Adams, Peter McInerney, Masood Hadi, and Henrik Vibe Scheller

Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California

Plant cell walls are composed primarily of structural polysaccharides such as cellulose, hemicelluloses and pectins. To assemble these polysaccharides, the plant needs an extensive biosynthetic machinery and it has been estimated that over 2000 gene products are involved in making and maintaining the wall (Carpita *et al.*, 2001; Dhugga, 2001). The polysaccharides are mainly synthesized by glycosyltransferases (GTs) – a family of enzymes that transfer a sugar residue from an activated donor substrate, usually a nucleotide sugar donor, to an acceptor such as a growing oligosaccharide, forming a glycosidic bond. GTs generally displays exquisite specificity for both the sugar donor and the acceptor substrates (Breton *et al.*, 2006), and are highly stereo- and regiospecific.

Functional prediction of a putative GT based on sequence similarities is problematic and many closely related sequences have different catalytic activites. In spite of the tremendous variety of reactions catalyzed by GTs, they appear to share a limited number of protein fold types and only two structural folds, GT-A and GT-B, or variants thereof, have been identified for the nucleotide sugar-dependent GTs solved to date. However, for many GT families – and particularly those specific to plants – no structure has been solved, so it is not clear if other fold types exist or only variants of the known patterns.

We believe that crystallization and structural comparison of the catalytic domains could help to find conserved motifs involved in both donor and acceptor substrate recognition of the many GT sequences. We have selection of a broad group of rice and Arabidopsis GT candidates potentially involved in polysaccharide assembly. Using bioinformatics and modeling, secondary structures were predicted for optimal construction of truncation variants suitable for crystallization. Furthermore, several protein expression vectors containing fusion protein tags for improvement of solubility and expression level of target proteins was selected for expression and production of active and/or soluble forms of protein for crystallization.

This project is funded by The Carlsberg Foundation and by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 with Lawrence Berkeley National Laboratory.

References:

- 1) Breton et al. (2006) Glycobiology, 16, 29R-37R.
- 2) Carpita et al. (2001) Plant Mol Biol, 47, 1-5.
- 3) Dhugga (2001) Curr Opin Plant Biol, 4, 488-93.

## Comparative Gene Expression of the *Caldicellulosiruptor* Genus Using RNAseq

**Loren J. Hauser**<sup>1</sup>\* (hauserlj@ornl.gov), Sara Blumer-Schuette,<sup>2</sup> Ira Kataeva,<sup>3</sup> Sung-Jae Yang,<sup>3</sup> Farris Poole,<sup>3</sup> Daniel Quest,<sup>1</sup> Inci Ozdemir,<sup>2</sup> Andrew Frock,<sup>2</sup> Erika Lindquist,<sup>4</sup> Tanya Woyke,<sup>4</sup> Bob Cottingham,<sup>1</sup> Michael W. W. Adams,<sup>3</sup> and Robert M. Kelly<sup>2</sup>

<sup>1</sup>Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>2</sup>North Carolina State University, Raleigh, North Carolina; <sup>3</sup>University of Georgia, Athens, Georgia; and <sup>4</sup>DOE Joint Genome Institute, Walnut Creek, California

All known members of the *Caldicellulosiruptor* genus grow optimally between 65°C to 80°C and can anaerobically degrade plant biomass using various and complementary strategies. They are prime candidates for use in an industrial consolidated bioprocessing facility to produce second generation biofuels from complex plant material such as switchgrass. In collaboration with the Department of Energy Joint Genome Institute (JGI) we have recently completed sequencing and annotating the genomes of eight members of this genus. In addition, we have generated RNAseq data from four members grown on a variety of carbon sources including, glucose, maltose, cellobiose, starch, crystalline cellulose (Avicel), and dilute acid pre-treated switchgrass. Two of the primary advantages of RNAseq are its dynamic range and sensitivity. Greater than 98.5% of all protein coding genes had some detectable expression in all growth states and varied in expression level up to 10<sup>6</sup> fold. The expression levels of some genes, when grown on different carbon sources, varied by over  $10^3$  fold. As expected, the genes encoding ABC sugar transporters, cellulases and other glycosyl hydrolases were amongst the genes with the greatest changes in expression levels when grown on sugars versus complex carbon sources such as switchgrass. However, there were a number of other genes, such as members of a CRISPR cluster and some genes involved in fatty acid metabolism, that had unexpected changes in expression when grown on different carbon sources. We are developing an analysis pipeline to process and visualize the data and will also compare them with the results from DNA microarray analyses. RNAseq analyses will also include identifying the 5' end of transcription units, defining operons, identifying co-regulated genes and operons, and predicting transcription factor binding sites. Preliminary analysis has identified putative promoters embedded in genes, which allows the definition of unconventional operons and regulons. A thorough analysis will undoubtedly reveal additional unique biological phenomenon.

### **Microbial Community Profiling of Restored Wetland Sediments**

**Shaomei He**<sup>1</sup>\* (She@lbl.gov), Mark Waldrop,<sup>2</sup> Lisamarie Windham-Myers,<sup>2</sup> Tanja Woyke,<sup>1</sup> and Susannah G. Tringe<sup>1</sup>

<sup>1</sup>DOE Joint Genome Institute, Walnut Creek, California, and <sup>2</sup>U.S. Geological Survey, Menlo Park, California

Wetland restoration has the benefit of reversing land subsidence on peat islands in areas drained for agriculture, thereby reducing risk of levee failure. In addition, it provides

wildlife habitat, and the high primary production and slow decomposition rates found in restored wetlands may result in a net negative carbon flux beneficial for sequestration of atmospheric  $CO_2$ . Despite these potential benefits, one major concern is the emission of methane that could potentially offset the greenhouse gas benefits of carbon captured due to primary production. In wetland ecosystems, microorganisms play key roles in important processes, such as methane production and oxidization. Therefore, we are interested in the microbes found in restored wetland ecosystems and the processes they mediate. In this preliminary study, we collected belowground samples from a restored wetland largely vegetated with cattails and tules from a U.S. Geological Survey pilot-scale restoration project on Twitchell Island in the Sacramento/San Joaquin Delta. We collected samples at two different depths, and extracted DNA from different plant biomass types. Pyrosequencing of amplified V8 regions of 16S rRNA genes was used to generate microbial community profiles. For all samples analyzed, microbial communities were primarily governed by plant biomass types. For the same plant biomass type, a moderate influence of depth was observed. Archaeal species closely affiliated with characterized methanogens were identified from both sample depths, as well as methanotrophic bacteria. The highest archaeal abundance was observed in a tule rhizome (accounting for  $\sim 24\%$  of the total microbial community). Many abundant microbial species were also similar to those previously observed in similar environments, such as peat bogs, marshes, rice paddies, anaerobic digesters, methanogenic consortia, soils and sediments. Many of these microbial species show very high 16S rRNA gene sequence similarity to characterized microorganisms that perform lignocellulose decomposition, sugar fermentation, denitrification, Fe(III) reduction, sulfate reduction, sulfur oxidation, methanogenesis and methane oxidation. Therefore, this primary investigation enables a glimpse of the microbial community composition and the likely biological processes that these microbes mediate. Further metagenomic and metatranscriptomic analyses are planned to reveal the microbial functions important for long-term carbon sequestration and nutrient cycling by the restored wetland microbial community.

### Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen

**Matthias Hess**<sup>1,2,3</sup>\* (matthias.hess@tricity.wsu.edu), Alexander Sczyrba,<sup>2,3</sup> Rob Egan,<sup>2,3</sup> Tae-Wan Kim,<sup>4</sup> Harshal Chokhawala,<sup>4</sup> Gary Schroth,<sup>5</sup> Shujun Luo,<sup>5</sup> Douglas S. Clark,<sup>4,6</sup> Feng Chen,<sup>2,3</sup> Tao Zhang,<sup>2,3</sup> Roderick I. Mackie,<sup>7</sup> Len A. Pennacchio,<sup>2,3</sup> Susannah G. Tringe,<sup>2,3</sup> Axel Visel,<sup>2,3</sup> Tanja Woyke,<sup>2,3</sup> Zhong Wang,<sup>2,3</sup> and Edward M. Rubin<sup>2,3</sup>

<sup>1</sup>Applied Microbial Genomics & Ecology, Washington State University, Richland, Washington; <sup>2</sup>DOE Joint Genome Institute, Walnut Creek, California; <sup>3</sup>Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California; <sup>4</sup>Energy Biosciences Institute, University of California, Berkeley, California; <sup>5</sup>Illumina Inc., Hayward, California; <sup>6</sup>Department of Chemical and Biomolecular Engineering, University of California, Berkeley, California; and <sup>7</sup>Department of Animal Sciences, Institute for Genomic Biology and Energy Biosciences Institute, University of Illinois, Urbana, Illinois

The paucity of enzymes that efficiently deconstruct plant polysaccharides represents a major bottleneck for industrial-scale conversion of cellulosic biomass into biofuels. Cow rumen microbes specialize in degradation of cellulosic plant material, but most members of this complex community resist cultivation. To characterize biomass-degrading genes and genomes, we sequenced and analyzed 268 gigabases of metagenomic DNA from microbes adherent to plant fiber incubated in cow rumen. From these data, we identified 27,755 putative carbohydrate-active genes and expressed 90 candidate proteins, of which 57% were enzymatically active against cellulosic substrates. We also assembled 15

uncultured microbial genomes, which were validated by complementary methods including single-cell genome sequencing. These data sets provide a substantially expanded catalog of genes and genomes participating in the deconstruction of cellulosic biomass.

#### iPlant EOT: Novel Education, Outreach, and Training Programs

Uwe Hilgert\* (hilgert@cshl.edu), Jason Williams, Cornel Ghiban, Eun-Sook Jeong, Mohammed Khalfan, and **David Micklos** 

Dolan DNA Learning Center, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

The iPlant Collaborative (iPlant; http://www.iplantcollabroative.org) is a project funded by the National Science Foundation to develop computer (cyber) infrastructure that provides plant researchers access to the large-scale datasets and high-powered informatics tools that drive modern biology. A key part of iPlant's mission involves education, outreach, and training to infuse iPlant-generated tools into innovative curricula and informal science programs.

**DNA Subway** The first educational product released by iPlant, DNA Subway (http://www.dnasubway.org) presents complex scientific tools and data in an intuitive and appealing interface, and makes high-level genome analysis broadly available to students and educators. "Riding" different lines, users can annotate up to 150,000 basepairs of DNA, prospect plant genomes for gene and transposon families, construct phylogenetic trees, identify plants through DNA Barcoding, and, soon, analyze 2<sup>nd</sup>-generation sequencing transcriptome data.

**DNA Barcoding** This project develops simplified workflows to engage students in plant identification through short, standardized DNA barcoding regions such as the *rbcL* chloroplast gene. iPlant provides cyberinfrastructure support for students who will amplify and sequence plant DNA and post the data online. Students then will analyze their data using DNA Subway bioinformatics tools.

**Orphaned Data** This project will support distributed research by connecting researchers who have under-analyzed datasets with faculty and students who wish to work with "real" data. An online "marketplace" will index projects with descriptions of data available and suggestions for how they can be used for class or independent research projects. The marketplace will also include networking tools to allow collaborators to coalesce around datasets.

## The Serpula lacrymans Wood Decay Transcriptome

Nils Högberg\* (nils.hogberg@mykopat.slu.se), A. Kohler, F. Martin, and J. Stenlid

Uppsala BioCenter, Department of Forest Mycology & Pathology, Uppsala, Sweden

Wood is a composite material consisting of cellulose, hemicellulose and lignin. Brown rot fungi are capable of excessive removal of cellulose and hemicellulose while lignin is modified and left like a weak amorphous skeleton. The *Serpula lacrymans* wood decay transcriptome include of 517 synchronized genes, which were more than 4-fold differentially regulated after 10 days compared to a glucose treatment. A major part of these genes were involved in carbohydrate metabolism and oxidative processes. Six genes are up-regulated > 100 fold including the cellulases/hemicellulases GH61, GH43, GH5 the latter two with cellulose binding modules, the oxidative enzymes iron reductase and FAD-

dependent pyridine nucleotide-disulphide oxidoreductase and two GH74 endoglucanases. In spite of the reduced numbers of plant cell wall degrading enzymes in the S. lacrymans genome, our data show that the remaining enzymes in this group have an active role in brown rot wood decay. Brown rot fungi have been hypothesized to utilize the Fenton reaction in order to produce hydroxyl radicals from hydrogen superoxide and reduced iron. In S. lacrymans the iron reductase and pyridine nucleotide-disulphide oxidoreductase are appropriate sources for Fenton chemistry substrates reduced iron and hydrogen superoxide respectively. The cellulose binding motif in the iron reductase enables substrate specificity which ensure hydroxyl radical activity in spite of their short half-life of these compounds. The synergistic action of genes in the transcriptome is further illustrated by the genes that show a more than 10-fold up-regulation which including a large number of carbohydrate active enzymes involved in plant cell wall degradation within the glycosyl hydrolase groups GH1, 3, 5, 10, 12, 28, 35, 43, 61, 74 and 115 and a number of genes with the capacity to form hydrogen superoxide including Alcohol oxidase and Aryl-alcohol dehydrogenase. Lignin is left as a modified residue by brown rot fungi, in S. lacrymans Copper radical oxidases Aromatic-ring hydroxylase and several copies of Cytochrome P450 enzymes that are up regulated more than 10-fold have the capacity to modify lignin. Brown rot fungi colonize wood via the ray cells with a high lipid content, accordingly a number of lipases are as well up regulated. Sugar transporters are another group of upregulated enzymes translocating glucose from the substrate to other parts of the mycelium. The wood degradation transcriptome of S. lacrymans and P. placenta contrast each other since the Postia genome lack the iron reductase with a Carbohydrate binding module 1, both critical parts of the Serpula transcriptome. Further experimentation within the fields of iron acquisition, symbiosis with pine seedlings and fruitbody formation is currently conducted which will increase the knowledge about gene expression in *S. lacrymans*.

### **Characterization of the Bacterial Metagenome in an Industrial Algae Bioenergy Production System**

Shi Huang,<sup>1</sup> Scott Fulbright,<sup>2</sup> Xiaowei Zeng,<sup>1</sup> Tracy Yates,<sup>3</sup> Greg Wardle,<sup>3</sup> Stephen Chisholm<sup>2</sup>\* (chisholm@colostate.edu), Jian Xu,<sup>1</sup> and **Peter Lammers**<sup>4</sup>

<sup>1</sup>Qingdao Institute of BioEnergy and BioProcess Technology, Chinese Academy of Sciences, Qingdao, Shandong Province, China; <sup>2</sup>Department of Bioagricultural Sciences and Pest Management, and Graduate Program in Cell and Molecular Biology, Colorado State University, Fort Collins, Colorado; <sup>3</sup>Solix Biofuels, Fort Collins, Colorado; and <sup>4</sup>Algal Bioenergy Program, New Mexico State University, Las Cruces, New Mexico

Cultivation of oleaginous microalgae for fuel generally requires growth of the intended species to the maximum extent supported by available light. The presence of undesired competitors, pathogens and grazers in cultivation systems will create competition for nitrate, phosphate, sulfate, iron and other micronutrients in the growth medium and potentially decrease microalgal triglyceride production by limiting microalgal health or cell density. Pathogenic bacteria may also directly impact the metabolism or survival of individual microalgal cells. Conversely, symbiotic bacteria that enhance microalgal growth may also be present in the system. Finally, the use of agricultural and municipal wastes as nutrient inputs for microalgal production systems may lead to the introduction and proliferation of human pathogens or interfere with the growth of bacteria with beneficial effects on system performance. These considerations underscore the need to understand bacterial community dynamics in microalgal production systems in order to assess microbiome effects on microalgal productivity and pathogen risks.

Here we focus on the bacterial component of microalgal production systems and describe a pipeline for metagenomic characterization of bacterial diversity in industrial cultures of an oleaginous alga, *Nannochloropsis salina*. Environmental DNA was isolated from 12 marine algal cultures grown at Solix Biofuels, a region of the 16S rRNA gene was amplified by PCR, and 16S amplicons were sequenced using a 454 automated pyrosequencer. The approximately 70,000 sequences that passed quality control clustered into 53,950 unique sequences. The majority of sequences belonged to thirteen phyla. At the genus level, sequences from all samples represented 169 different genera. About 53% of all sequences could not be identified at the genus level and were classified at the next highest possible resolution level. Of all sequences, 79.92% corresponded to 169 genera and 70 other taxa. We apply a principal component analysis across the initial sample set to draw correlations between sample variables and changes in microbiome populations.

## **Testing an Enrichment Culture Metagenome Binning Heuristic** with a Dechlorinating Microbial Consortium

**Laura A. Hug**,<sup>1</sup> Paul J. McMurdie,<sup>2</sup> Alison S. Waller,<sup>3</sup> Susan Holmes,<sup>4</sup> Alfred M. Spormann,<sup>2,5</sup> Elizabeth A. Edwards<sup>1,3</sup>\* (elizabeth.edwards@utoronto.ca)

<sup>1</sup>Department of Cell and Systems Biology, University of Toronto, Toronto, Ontario, Canada; <sup>2</sup>Department of Civil and Environmental Engineering, Stanford University, Stanford, California; <sup>3</sup>Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario, Canada; and <sup>4</sup>Department of Statistics and <sup>5</sup>Chemical Engineering, Stanford University, Stanford, California

Chlorinated solvents are recalcitrant groundwater contaminants that pose environmental and human health hazards. Several bacterial groups are known to partially degrade chlorinated compounds, often leading to increased concentrations of daughter products, including vinyl chloride (VC), a known human carcinogen. Dehalococcoides spp. are key players in the complete reductive dechlorination of tetrachloroethene (PCE) and trichloroethene (TCE) to the non-toxic end product ethene. The enrichment culture KB-1 is used industrially for bioaugmentation and bioremediation of chlorinated ethenes at contaminated sites. The KB-1 consortium contains six to ten dominant organisms, including *Dehalococcoides* (Dhc) strains responsible for the critical vinyl chloride (VC) to ethene detoxification step via reductive dehalogenation. A metagenome was generated by the Joint Genome Institute (JGI) to provide a blueprint for elucidating the metabolic interrelationships within KB-1. In total, 95 MB of Sanger sequence from a short insert (3) kb) library and a fosmid (40 kb) library was assembled de novo to 6361 contigs. We developed a custom binning heuristic method using k-means on the contig trinucleotide frequencies, and implemented novel methods to allow scaffolding information and organismal abundance to inform the binning process. The extent of coverage of the dominant organisms' genomes was estimated. Manual gap closure of Dehalococcoides contigs resulted in a draft assembly of a core Dehalococcoides genome from the dominant strain in KB-1.

### **Comparative Metagenomics of Alaskan Permafrost Soils**

**Jenni Hultman**<sup>1</sup>\* (jjhultman@lbl.gov), Rachel Mackelprang,<sup>3</sup> Olivia U. Mason,<sup>1</sup> Regina Lamendella,<sup>1</sup> Krystle Chavarría,<sup>1</sup> Tijana Glavina del Rio,<sup>3</sup> Rex Malmstrom,<sup>3</sup> Tanja Woyke,<sup>3</sup> Eddy Rubin,<sup>3</sup> Mark Waldrop,<sup>2</sup> and Janet K. Jansson<sup>1,3</sup>

#### **Poster Presentations**

<sup>1</sup>Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; <sup>2</sup>U.S. Geological Survey, Menlo Park, California; <sup>3</sup>DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, California

The fate of organic carbon reserves currently sequestered in permafrost is uncertain yet critically important for addressing terrestrial feedbacks to climate change. Warming increases the probability of thermokarst formation and increases  $CO_2$  and  $CH_4$  flux to the atmosphere. However, we understand little of the underlying microbial controls of nitrogen or carbon cycling in permafrost soils. Metagenomic and 16S rRNA gene pyrotag sequencing was used to study both in situ microbial communities and their functions in permafrost soils and in incubation experiments. Samples were obtained from four sites in Alaska that differed in their soil carbon chemistry: 1) a stable, low productivity black spruce forest, 2) a thermokarst bog at the Bonanza Creek LTER Station outside of Fairbanks, 3) a mineral permafrost soil at Coldfoot and 4) an organic soil at Hess Creek. Permafrost samples were taken from both the active and the permafrost layers. In total, over 100,000 16S rRNA genes were sequenced and >100 Gb of metagenome data was obtained. The results of the pyrotag sequencing indicated that the communities in the permafrost layers and bog were more similar to each other than to those in the active layers. An uncultured representative of the Chloroflexi phylum was found to be cosmopolitan in the bog and permafrost layer samples. Therefore, we sorted single cells from a sample where this OTU was highly represented and subjected these to single cell sequencing. For the metagenome data we tested different methods to assemble the Illumina GAII data and applied tetranucleotide clustering of the assembled contigs to construct draft genomes, including that of a novel methanogen. The metagenome data was also screened to determine which functional genes were abundant in the different samples. In addition, we incubated subsamples at 5°C with and without substrates to monitor the response of the communities to thaw and found that several genes involved in nitrogen cycling were enriched upon thaw. In addition, several genes responsible for degradation of labile carbon substrates were enriched. These data are the first to highlight specific functional changes and community responses of uncultivated microbes in permafrost. Our findings lay the groundwork for using metagenomic data to gain a better understanding of the potential impact climate change will have on microbial community processes in the arctic.

## Metagenome Analysis of High-temperature Chemotrophic Microbial Communities Provides a Foundation for Dissecting Microbial Community Structure and Function

**W. Inskeep**<sup>1</sup>\* (binskeep@montana.edu), M. Kozubal,<sup>1</sup> J. Beam,<sup>1</sup> Z. Jay,<sup>1</sup> R. Jennings,<sup>1</sup> H. Bernstein,<sup>2</sup> R. Carlson,<sup>2</sup> D. Rusch,<sup>3</sup> S. Tringe,<sup>4</sup> M. Romine,<sup>5</sup> R. Brown,<sup>5</sup> M. Lipton,<sup>5</sup> and J. Fredrickson<sup>5</sup>

<sup>1</sup>Department of Land Resources and Environmental Sciences and Thermal Biology Institute, Montana State University, Bozeman, Montana; <sup>2</sup>Department of Chemical and Biological Engineering, Montana State University, Bozeman, Montana; <sup>3</sup>J. Craig Venter Institute, Rockville, Maryland; <sup>4</sup>DOE Joint Genome Institute, Walnut Creek, California; and <sup>5</sup>DOE Pacific Northwest National Laboratory, Richland, Washington

Microbial communities are a collection of interacting populations, each comprised of numerous individuals. However, a significant fraction of our knowledge base in microbiology originates from organisms grown and studied in pure culture, in the absence of other members of the community who may compete for resources or provide necessary co-factors and or substrates. Moreover, many of the organisms studied in pure culture have not represented the numerically dominant members of microbial communities found in situ. The advent of molecular tools (and -omics technologies) has provided opportunities for assessing the predominant and relevant indigenous organisms, as well as their likely function within a connected network of different populations (i.e., community). Hightemperature microbial communities are often considerably less diverse than mesophilic environments and constrained by dominant geochemical attributes such as pH, dissolved oxygen, Fe, sulfide, and or trace elements including arsenic and mercury. Consequently, the goal of our work is to utilize high-temperature geothermal environments including acidic Fe-oxidizing communities as model systems for understanding microbial interactions among community members. Recent metagenomic sequencing of hightemperature, acidic Fe-mats of Norris Geyser Basin, Yellowstone National Park (YNP) reveal communities dominated by novel archaea, bacterial members of the deeply-rooted Order Aquificales, and less-dominant Bacillales and Clostridiales. Phylogenetic and functional analysis of metagenome sequence is providing an excellent foundation for hypothesizing the role of individual populations in a network of interacting community members, and for testing specific hypotheses regarding the importance of biochemical pathways responsible for material and energy cycling. For example, we are using metagenome sequence in combination with information available from reference strains to identify protein-coding sequence of importance in the oxidation and or reduction of Fe, S, O, and As, as well as central C metabolism (including fixation of  $CO_2$ ). Genes coding for proteins with hypothetical or putative roles in electron transfer, and C-cycling are being investigated using quantitative-reverse transcriptase-PCR (Q-RT-PCR) to evaluate functional capacity quantitatively in both pure-culture and subsequent mixed communities. Future transcriptomic and proteomic analyses will be coupled with detailed studies focused on the position of different organisms (spatial context) during Fe-mat development, as well as the role of  $O_2$  flux across Fe-oxidizing boundary layers. Depositional studies have been conducted to correlate Fe-oxide deposition rates with O<sub>2</sub> flux rates measured using O<sub>2</sub>microelectrodes. Consensus genome sequence of 5-6 dominant community members is being used to develop population specific metabolic models that can then be coupled with appropriate assumptions to develop community network models. Proteomic results will be used to assess and confirm the importance of specific proteins and to make improvements to individual and or community models. Application of genomic, proteomic, and metabolic information to dissect microbial community structure and function is tractable within high-temperature geothermal systems in part due to the relative simplicity of the community and stability of several key geochemical variables (i.e. pH, Fe, O<sub>2</sub>).

## What Is the Role of Microaerophilic Bacteria during Cellulose Degradation? Learning from a Model

Jantiya Isanapong,<sup>1</sup>W. Sealy Hambright,<sup>1</sup> Atcha Boonmee,<sup>2</sup> and **Jorge L.M. Rodrigues**<sup>1</sup>\* (jorge@uta.edu)

<sup>1</sup>University of Texas, Arlington, Texas, and <sup>2</sup>Khon Kaen University, Khon Kaen, Thailand

Wood-feeding termites degrade 1 billion tons of plant material every year, performing an important function in the global carbon cycle. For degradation to occur, termites depend on a microbial community, which works in complete synchrony to transform cellulose and hemicellulose into oligosaccharides, H<sub>2</sub>, and CH<sub>4</sub>. Being composed of more than 250 bacterial species, our knowledge about the different members of this microbial community remains limited. We chose the <u>Verrucomicrobium</u> sp. strain TAV2 to gain insights into the role of microaerophiles surrounding the gut wall of the termite *Reticulitermis flavipes*.

Genome analysis of strain TAV2 revealed the presence of a *cbb3*-type cytochrome oxidase gene, responsible for capturing free  $O_2$  and maintaining optimum sub-oxic conditions. The transcriptional and proteomic profiles were compared for TAV2 cells grown under two different oxygen concentrations, 2% and 20%. A custom-designed microarray containing 4,022 coding sequences was used competitive hybridizations. High throughput proteomics was performed with ion trap mass spectrometers - MS operating in tandem. Data analyses were performed with the softwares GeneSpring GX11 and DAnTE v.1.2.

A total of 75 genes were observed as differentially expressed (P < 0.05) in our transcriptional analysis, representing 1.9% of all genes present in the microarray. When cells were grown with 2% O<sub>2</sub> concentration, a condition thought to represent the surroundings of the termite gut wall, 49 and 26 genes were up and down regulated, respectively. Our proteomic analysis indicated that 55 proteins were significantly expressed at 2% O<sub>2</sub> concentration, while 30 proteins were only detected at 20% O<sub>2</sub> ((P < 0.05). When a principal component analysis was performed for the proteomic dataset, we were able to explain 72.3% of the variation, with O<sub>2</sub> being the important cause for proteomic differential expression. Positive correlation between log transformed values for transcriptional and proteomic data were observed for specific pathways such as carbohydrate metabolism and enzymes such as the acetyl xylan esterase, which is involved in hemicellulose degradation. In addition, peptides corresponding to the *cbb3* cytochrome oxidase were detected in higher numbers in cells growing under 20% O<sub>2</sub> condition.

## Insights from the Genome of the Cellulolytic Thermophile *Clostridium clariflavum* DSM 19732

**Javier A. Izquierdo**<sup>1,2</sup>\* (Javier.Izquierdo@Dartmouth.EDU), Lynne A. Goodwin,<sup>3</sup> Tanja Woyke,<sup>3</sup> Karen W. Davenport,<sup>3</sup> Shunsheng Han,<sup>3</sup> Lee R. Lynd<sup>1,2</sup>

<sup>1</sup>Thayer School of Engineering, Dartmouth College, Hanover, New Hampshire; <sup>2</sup>DOE BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>3</sup>DOE Joint Genome Institute, Walnut Creek, California

*Clostridium clariflavum* is a thermophilic anaerobe isolated from thermophilic sludge able to utilize a variety of cellulosic substrates. We have recently characterized lignocellulolytic enrichments from thermophilic compost dominated by C. clariflavum strains. Given their predominance in these consortia and the ability of these organisms to utilize xylan, we have sequenced the genome of the type strain to further understand the physiology of lignocellulose utilization in this organism. A total of 72 glycosyl hydrolases have been identified, most of which have their closest match in C. thermocellum, its closest sequenced relative. A variety of novel chimeric glycosyl hydrolases with multiple functions have been identified, including a family 48 and family 9 multi-catalytic enzyme of critical relevance. We have also detected a carbohydrate sensing mechanism similar to the system recently proposed for C. thermocellum with potential cellulose-, xylan- and pectin-active modules. A variety of novel cellulosomal structural proteins have also been identified, including novel anchoring proteins and untethered proteins with unique CBM modules. In contrast with C. thermocellum, we have found broader diversity of xylanases and of enzymes involved in 5-carbon sugar metabolism. In addition, the genome has allowed us to identify novel central metabolism pathways within the Cluster III cellulolytic clostridia. Our exploration of the C. clariflavum genome provides relevant insights into a variety of novel approaches in cellulolytic clostridia that enable them to break down hemicellulose and other components of woody plant biomass.

## **Integration of Flux-Balance Analysis and Pathway Databases**

Peter D. Karp\* (pkarp@ai.sri.com) and Mario Latendresse

SRI International, Menlo Park, California

We describe new computational techniques for generating metabolic flux models from pathway databases. The Pathway Tools [1] software is a software package for creating, updating, visualizing, and analyzing Pathway/Genome Databases (PGDBs) for organisms with sequenced genomes. We have recently developed software for generating a linear programming model of a metabolic reaction network that is stored within a PGDB. Pathway Tools automatically invokes the SCIP solver on that model. The resulting optimized fluxes are then displayed on a metabolic pathway map for the PGDB by Pathway Tools, to accelerate a user's understanding of the predicted fluxes.

Benefits of this approach are that the metabolic flux model is closely coupled with an integrated genomic/metabolic knowledge base, and with other computational tools for manipulating that knowledge base. For example, users can visualize reactions, metabolites, pathways, and genome information using the rich visualization capabilities of Pathway Tools. Users can update metabolic reactions, substrates, and pathway definitions using the interactive editors within Pathway Tools, and those updates are reflected in the flux-balance model that is generated from the PGDB. In addition, metabolic model debugging tools within Pathway Tools can be applied to the metabolic flux model. Example debugging tools include tools for element balancing of metabolic reactions, and for detecting dead-end metabolites in the metabolic network.

In addition we have developed novel methods for completing a metabolic model. We have extended the gap-filling work of Maranas and colleagues to yield a multiple gap-filling approach. Using a meta-optimization procedure that is also automatically generated from a PGDB, our software will extend an incomplete metabolic model by postulating reversals of unidirectional reactions in the metabolic model, and by postulating additions of new reactions to the metabolic model from the MetaCyc database. These two approaches extend metabolic models to produce biomass compounds that they were previously unable to synthesize. Additionally, our software will gap-fill the nutrient compounds, that is, adding additional nutrient compounds that will produce biomass compounds can still not be produced. Finally, the software will identify which biomass compounds can still not be produced even after the preceding types of gap filling, thus further focusing the user's model debugging efforts. Taken together, these techniques can radically shorten the time required to develop FBA models from months to days.

1. Karp, P.D., Paley, S.M., Krummenacker, M., Latendresse, M., Dale, J.M., Lee, T., Kaipa, P., Gilham, F., Spaulding, A., Popescu, L., Altman, T., Paulsen, I., Keseler, I.M., and Caspi, R. (2010) "Pathway Tools version 13.0: Integrated Software for Pathway/Genome Informatics and Systems Biology," *Briefings in Bioinformatics* 11:40-79.

## Genome-enabled Investigations of H<sub>2</sub> Production by the Photosynthetic Bacterium *Rhodobacter sphaeroides*

**Wayne S. Kontur**<sup>1,3</sup>\* (wskontur@wisc.edu), Eva C. Ziegelhoffer, <sup>1</sup> Melanie A. Spero,<sup>1,3</sup> Saheed Imam,<sup>1,3</sup> Daniel R. Noguera,<sup>2,3</sup> and Timothy J. Donohue<sup>1,3</sup>

<sup>1</sup>Department of Bacteriology and <sup>2</sup>Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Wisconsin, and <sup>3</sup>DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Wisconsin

*Rhodobacter sphaeroides* is a purple non-sulfur photosynthetic bacterium that can utilize intracellular reductant during photoheterotrophic growth to produce the potential biofuel hydrogen gas  $(H_2)$  via its nitrogenase enzyme. To gain insight into the genes important to the onset of  $H_2$  production and those that impact the amount of  $H_2$  produced, we have performed global transcript analyses on cells producing no measurable  $H_2$  and cells fed different organic substrates that produce differing amounts of  $H_2$ .

Our analysis shows that there are only a small number of differences in RNA accumulation between H<sub>2</sub>-producing and non-H<sub>2</sub>-producing cells. Using a 2-fold change in RNA abundance as a cut-off, the transcripts that show differential accumulation between H<sub>2</sub>producing and non-H<sub>2</sub>-producing cells are mostly derived from genes involved in three metabolic pathways. The genes involved in nitrogenase assembly and function (including those predicted to transfer electrons to the enzyme) and the genes involved in uptake hydrogenase assembly and function are significantly higher in transcript abundance in the H<sub>2</sub>-producing cells. In contrast, most genes located in the two operons that code for enzymes of the Calvin-Benson-Bassham (CBB) cycle are significantly lower in transcript abundance in the H<sub>2</sub>-producing cells.

The global transcript analysis of cultures producing differing amounts of  $H_2$  from different organic substrates reveals the relationships between  $H_2$  production capacity and transcript level for the genes of these three pathways. Not surprisingly, nitrogenase transcript levels are positively correlated to  $H_2$  production capacity for low  $H_2$  producing cells. However, we find that transcript level plateaus for cells with high  $H_2$  production capacities. Though uptake hydrogenase RNA levels are higher in the  $H_2$  producing cultures than the non- $H_2$ producing cultures, we find that RNA levels are negatively correlated to  $H_2$  production capacity. The genes for the CBB cycle are also negatively correlated with  $H_2$  production capacity of a culture. Finally, we note that RNA levels of the genes that code for the enzymes involved in polyhydroxybutyrate (PHB) synthesis (a carbon and electron reserve polymer whose production can detract from  $H_2$  production) do not correlate with the amounts of PHB or  $H_2$  produced by the cultures. Our transcriptome results, along with analyses of mutants with defects in the above pathways, have helped us test predictions for how the interplay between these pathways affects the distribution of cellular reductant and ultimately determines  $H_2$  production capacity.

In conjunction with JGI, we are also analyzing the genome of an uncharacterized mutant strain of *R. sphaeroides*, Ga, that exhibits an elevated H<sub>2</sub>-production capacity compared to its wild-type parent (strain 2.4.1). Preliminary analysis suggests that there are  $\sim 100$  small-scale differences (SNPs and indels) between the strains, along with a few large-scale deletions in the Ga genome. Further investigation will allow us to determine whether the differences in H<sub>2</sub> production capacity in Ga and a library of other wild type strains is due to differences in pathways already known to affect H<sub>2</sub> production, or some pathway(s) whose relationship to H<sub>2</sub> production is as yet uncharacterized.
# High Throughput Production and Characterization of Cellulytic and Hemicellulytic Enzymes

**LauraLynn Kourtz<sup>1</sup>**\* (lkourtz@c56technologies.com), David Mead,<sup>2,3</sup> Colleen Drinkwater,<sup>2,3</sup> Julie Boyum,<sup>2,3</sup> Jan Deneke,<sup>2,3</sup> Krishne Gowda,<sup>1</sup> Ronald Godiska,<sup>1</sup> Eric Steinmetz,<sup>1</sup> and Phil Brumm<sup>1,3</sup>

<sup>1</sup>C56 Technologies Inc., Middleton, Wisconsin; <sup>2</sup>Lucigen Corp., Middleton, Wisconsin; and <sup>3</sup>DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Wisconsin

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

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

### Soil Fungal Communities and Their Responses to Climate Change Factors in Large-Scale Field Experiments

**Cheryl R. Kuske**<sup>1</sup>\* (Kuske@lanl.gov), La Verne Gallegos-Graves,<sup>1</sup> Carolyn F. Weber,<sup>1</sup> Stephanie A. Eichorst,<sup>1</sup> Monica Moya Balasch,<sup>1</sup> Andrea Porras-Alfaro,<sup>2</sup> Gary Xie,<sup>1</sup> Kuan-Liang Liu,<sup>1</sup> Lawrence O. Ticknor,<sup>1</sup> Rytas Vilgalys,<sup>3</sup> R. David Evans,<sup>4</sup> Bruce A. Hungate,<sup>5</sup> Robert B. Jackson,<sup>3</sup> J. Patrick Megonigal,<sup>6</sup> Christopher W. Schadt,<sup>7</sup> and Donald R. Zak<sup>8</sup>

<sup>1</sup>Los Alamos National Laboratory, Los Alamos, New Mexico; <sup>2</sup>Western Illinois University, Macomb, Illinois; <sup>3</sup>Duke University, Durham, North Carolina; <sup>4</sup>Washington State University, Pullman, Washington; <sup>5</sup>Northern Arizona University, Flagstaff, Arizona; <sup>6</sup>Smithsonian Environmental Research Center, Edgewater, Maryland; <sup>7</sup>Oak Ridge National Laboratory, Oak Ridge, Tennessee; and <sup>8</sup>University of Michigan, Ann Arbor, Michigan

Elevated atmospheric  $CO_2$  generally increases plant productivity and subsequently increases the availability of cellulose in soil to microbial decomposers. As key cellulose degraders, soil fungi are likely to be one of the most impacted and responsive microbial groups to elevated atmospheric  $CO_2$ . However, we do not understand how soil fungal communities are distributed *in situ*, which fungi contribute most to cellulose degradation, or how they respond to elevated  $CO_2$ . Using a combination of Sanger and 454 pyrotag sequencing, we investigated the diversity and composition of fungi in soils from six large DOE climate change experiments (FACE and OTC experiments) that span very different terrestrial ecosystems. We determined the relative impacts of ecosystem type, elevated atmospheric  $CO_2$ , and other soil and treatment factors on the soil fungal community using the large subunit rRNA gene (LSU) as a molecular survey tool. To support taxonomic analysis of soil LSU sequences, we generated a ~7000 member sequence dataset and used this dataset to calibrate the accuracy of a new fungal classifier program. Using the GH7 cellobiohydrolase I gene *(cbhI)* as a proxy for the soil fungal cellulolytic community, we conducted a parallel DNA-based survey across five of the ecosystem experiments. Currently, we are determining patterns of expression of this gene in soil RNA at the Duke Forest pine FACE site.

### Sequence and Annotation of Genome of Shiitake Mushroom *Lentinula edodes*

**H.S. Kwan**<sup>\*</sup> (hoishankwan@cuhk.edu.hk), C.H. Au, M.C. Wong, J. Qin, I.S.W. Kwok, W.W.Y. Chum, P.Y. Yip, K.S. Wong, L. Li, Q.L. Huang, W.Y. Nong, and M.K. Cheung

The Chinese University of Hong Kong, Hong Kong SAR, People's Republic of China

Lentinula edodes (Shiitake/Xianggu) is an important cultivated edible mushroom. Understanding the genomics and functional genomics of L. edodes allows us to improve its cultivation and quality. Sequencing the genome provides a comprehensive understanding of the biology of the mushroom. We can also develop many molecular genetic markers for breeding and genetic manipulation. We can identify genes encoding various bioactive proteins and pathways leading to bioactive compounds. We sequenced the genome of L. edodes monokaryon L54A using Roche 454 and ABI SOLiD genome sequencing. Sequencing reads of about 1400Mbp were *de novo* assembled into a 39.8Mb genome sequence. We compiled the genome sequence into a searchable database with which we have been annotating the genes and analyzing the metabolic pathways. Gene ortholog groups of *L. edodes* genome sequence were compared across genomes of several fungi, including mushrooms, identifying gene families unique to mushroom-forming fungi. In addition, we have been using a battery of molecular techniques to annotate the genome of Lentinula edodes. We used RNA arbitrarily primed-PCR, SAGE, LongSAGE, EST sequencing and cDNA microarray to analyze genes differentially expressed in different growth stages and conditions. We are learning more about the molecular biology of this mushroom, including the fruiting body development, its lignocellulolytic systems and other important metabolic pathways.

### Integrated Metagenomics and Metaproteomics Analyses of the Gut Microbiome in Twins with Crohn's Disease

**Regina Lamendella**<sup>1</sup>\* (rlamendella@lbl.gov), Alison Erickson,<sup>3</sup> Brandi Cantarel,<sup>2</sup> Jonas Halfvarson,<sup>5</sup> Nathan VerBerkmoes,<sup>3</sup> Manesh Shah,<sup>3</sup> Youseff Darzi,<sup>4</sup> Jeroen Raes,<sup>4</sup> Claire Fraser-Liggett,<sup>2</sup> Robert Hettich,<sup>3</sup> and Janet Jansson<sup>1,6</sup>

<sup>1</sup>Lawrence Berkeley National Laboratory, Berkeley, California; <sup>2</sup>The University of Maryland School of Medicine, Baltimore, Maryland; <sup>3</sup>Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>4</sup>Flemish Institute for Biotechnology, Univ. of Brussels, Belgium; <sup>5</sup>Örebro University Hospital, Örebro, Sweden; <sup>6</sup>DOE Joint Genome Institute, Walnut Creek, California

The causes and etiology of Crohn's disease (CD) are currently poorly understood, however both host genetics and environmental factors are thought to play a role. The current hypothesis is that a general breakdown in the balance between protective and harmful bacteria in the gut, "dysbiosis", sparks the inflammation associated with CD. In this study, we examined the extent to which shifts in the gut microbial community structure, gene content, and expressed proteins are relevant to the etiology of CD. Shotgun metagenomics and metaproteomics were used to obtain the complement of genes and expressed proteins, respectively, in fecal samples collected from a Swedish twin cohort, comprised of four healthy individuals, six individuals with inflammation in the ileum (ileal CD, ICD), and two individuals with inflammation in the colon (colonic CD, CCD). The data were analyzed with a variety of statistical approaches including non-metric multidimensional scaling (nMDS), partial least squares (PLS) models, species indicator tests, and pathway comparisons (iPATH and DAVID). Preliminary results showed distinct clustering of individual metaproteomes by disease status. Species indicator tests and PLS models revealed several specific microbial populational shifts, including a significant reduction of Fecalibacterium in the ICD cohort. By comparing the biochemical pathways that are depleted in ICD patients that also are depleted in several commensal gut members, it was possible to depict some of the gut processes that are otherwise carried out by these members of the gut microbiota. Interestingly, the members of the bacterial community that appear to be depleted in the ICD cohort, play important roles in acetate and butyrate production. Additionally, marked shifts in the abundance of specific human proteins were correlated with disease status. For example, ICD individuals had significantly higher amounts of several digestive enzymes, which is consistent with previous metabolite data suggestive of malabsorption in the gut of CD patients. Additionally, ICD patients possessed a higher number of immunological proteins, supporting a heightened immune activity often associated with CD. This study highlights the utility of integrating taxonomic, metagenomic, and metaproteomic profiles to correlate specific changes in community structure and function with disease status.

#### **UV Decontamination of MDA Reagents for Single Cell Genomics**

**Janey Lee**<sup>1</sup>\* (jlee2@lbl.gov), Damon Tighe,<sup>1</sup> Alexander Sczyrba,<sup>1</sup> Christian Rinke,<sup>1</sup> Scott Clingenpeel,<sup>1</sup> Rex Malmstrom,<sup>1</sup> Stephanie Malfatti,<sup>1</sup> Len Pennacchio,<sup>1</sup> Ramunas Stepanauskas,<sup>2</sup> Jan-Fang Cheng,<sup>1</sup> and Tanja Woyke<sup>1</sup>

<sup>1</sup>DOE Joint Genome Institute, Walnut Creek, California, and <sup>2</sup>Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, Maine

Single cell genomics, the amplification and sequencing of genomes from single cells, can provide a glimpse into the genetic make-up and thus life style of the vast majority of uncultured microbial cells, making it an immensely powerful and increasingly popular tool. This is accomplished by use of multiple displacement amplification (MDA), which can generate billions of copies of a single bacterial genome producing microgram-range DNA required for shotgun sequencing. Here, we would like to address one challenge inherent in such a sensitive method and propose a solution for the improved recovery of single cell genomes. While DNA-free reagents for the amplification of a single cell genome are a prerequisite for successful single cell sequencing and analysis, DNA contamination has been detected in various reagents, which poses a considerable challenge. Our study demonstrates the effect of UV radiation in efficient elimination of exogenous contaminant DNA found in MDA reagents, while maintaining Phi29 activity. Consequently, we also find that increased UV exposure to Phi29 does not adversely affect genome coverage of MDA amplified single cells. While additional challenges in single cell genomics remain to be resolved, the proposed methodology is relatively quick and simple and we believe that its application will be of high value for future single cell sequencing projects.

#### **DUK – A Fast and Efficient Kmer Matching Tool**

Mingkun Li\* (mli@gmail.com), Alex Copeland, and James Han

DOE Joint Genome Institute, Walnut Creek, California

A new tool, DUK, is developed to perform matching task. Matching is to find whether a query sequence partially or totally matches given reference sequences or not. Matching is similar to alignment. Indeed many traditional analysis tasks like contaminant removal use alignment tools. But for matching, there is no need to know which bases of a query sequence matches which position of a reference sequence, it only need know whether there exists a match or not. This subtle difference can make matching task much faster than alignment. DUK is accurate, versatile, fast, and has efficient memory usage. It uses Kmer hashing method to index reference sequences and Poisson model to calculate p-value. DUK is carefully implemented in C++ in object oriented design. The resulted classes can also be used to develop other tools quickly. DUK have been widely used in JGI for a wide range of applications such as contaminant removal, organelle genome separation, and assembly refinement. Many real applications and simulated dataset demonstrate its power.

### The Root-associated Microbiota: Applying Plant Genetics to Extended Phenotypes

**Derek S. Lundberg\*** (derekLS@email.unc.edu), Sarah Lebeis, Sur Herrera Paredes, and Jeffrey Dangl

University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

Plant roots, through their physical structure and the release of larger molecules and ions, create a local environment that is chemically and microbially distinct from the bulk soil in which they grow. The advent of culture-independent methods and modern sequencing technologies allow us to study root-associated microbial communities with greater detail. The biological significance of the majority of the microbial associations is unknown, as is the extent and means by which the host plant genetics controls the composition of root-associated microbial communities. These questions have significant scientific value and also vast economic value for crop plants, because symbiotic microorganisms enable plants to acquire essential nutrients, resist pathogens, and overcome abiotic stresses.

Microbiota composition, in plants and animals, shows correlation with host genotype when great evolutionary distances between host species are considered. Using the genetic power of Arabidopsis thaliana, as well as related Brassicaceae species and the model grass Brachypodium distachyon (being sequenced by JGI), we are testing the hypothesis that the microbiota of a plant is an extended phenotype sufficiently dependent on the plant's genotype to be mapped to quantitative trait loci or individual genes. We grow multiple individuals of inbred plant genotypes in controlled laboratory conditions, using surface-sterilized seeds with homogenized wild soils - including soils characterized in great depth by JGI and James Tiedje et al. Individual plants are harvested at flowering and at a senescent developmental stage and environmental DNA is made from both rhizosphere soil and total root systems. We describe communities using 454-generated sequences of small subunit rRNA amplicons. Having described the root-associated communities of hundreds of individual plants, we confirm the expected result that all rhizospheres, regardless of soil type or genotype, assemble a microbial community distinct from bulk soil. Soil type causes the most radical differences between rhizosphere samples, while the intra-specific genotype has only a minor effect for rhizospheres. Current data

suggests that microbial endophyte communities are more dependent on host plant genotype that rhizosphere communities, and imminent data should confirm this. Several microbes are consistently more common on and in roots than in bulk soil, demonstrating reproducible symbiosis. Further, we have succeeded in culturing some of these enriched symbiotic microbes and have plans to generate draft genomes. We are beginning to generate full metagenomic and metatranscriptomic datasets to compare four Arabidopsis accessions with three other members of the Brassicaceae family.

#### Time-dependent Profiles of Transcripts Encoding Lignocellulosemodifying Enzymes of the White Rot Fungus *Phanerochaete carnosa* Grown on Multiple Wood Substrates

Jacqueline MacDonald\* (jacqueline.macdonald@utoronto.ca) and Emma Master

Department of Chemical Engineering, University of Toronto, Toronto, Ontario, Canada

The wood-decaying white-rot fungi produce sets of enzymes that can degrade all the main components of lignocellulose and so are a valuable source of enzymes used in the production of renewable chemicals and liquid fuel from wood. While most white-rot have been isolated primarily from hardwoods (angiosperms), Phanerochaete carnosa has been isolated almost exclusively from softwoods (gymnosperms). It is anticipated that by elucidating the enzyme activities which facilitate softwood decay by *P. carnosa*, new enzyme formulations will be identified that result in more efficient utilization of this resource. Previous transcriptome analysis of P. carnosa identified transcripts that are enriched during growth on wood substrates. To gain a greater understanding of wood decay by this fungus, we used quantitative (q) RT-PCR to quantify transcripts encoding manganese peroxidases (MnP), lignin peroxidases, mannanase, xylanase, acetyl xylan esterase, glucuronovl esterase, and cellobiohydrolase at five time points during cultivation on balsam fir, lodgepole pine, white spruce, or sugar maple. The transcript profiles are consistent with a concerted response to wood species and a sequential decay strategy in which lignin is decayed early on. Compared to the model hardwood-degrading Phanerochaete chrvsosporium, P. carnosa produces a greater proportion of transcripts encoding proteins involved in lignin decay, particularly manganese peroxidase. We also evaluated three internal standards for qRT-PCR, and found transcripts encoding chitin synthase to be more consistently expressed than those encoding actin or GAPDH.

#### **Bioinformatic Characterization of Oxalate Decarboxylases and Formate Dehydrogenases in the Agaricomycotina (Basidiomycota)**

Miia Mäkelä\* (miia.r.makela@helsinki.fi), Ilona Oksanen, Annele Hatakka, and Taina Lundell

University of Helsinki, Department of Food and Environmental Sciences, Division of Microbiology, Fungal Biotechnology Laboratory, Helsinki, Finland

Organic acids secreted by basidiomycetous fungi are important metabolites as they play several roles in fungal growth, defence reactions, and nutrient uptake. Oxalic acid, which is a toxic compound in high concentrations, is the most commonly secreted fungal acid. In wood-decaying brown-rot fungi, oxalate is proposed to assist in the decomposition of cellulose, whereas in white-rot fungi, oxalate facilitates the reactions catalyzed by ligninmodifying oxidative enzymes. However, high oxalate levels are shown to inhibit the decomposition reactions. For these reasons, specific oxalate-degrading enzyme oxalate

Posters alphabetical by first author. \*Presenting author

decarboxylase (ODC, EC 4.1.1.2) is recognized as one of the key enzymes in wood and lignin degradation by saprobic fungi.

Oxalate decarboxylase (ODC, EC 4.1.1.2) is a primarily intracellular enzyme produced by certain bacteria and fungi and it catalyzes the decarboxylation of oxalate into formate and  $CO_2$  in a highly specific reaction. ODC is a Mn-containing multimeric enzyme of the functionally diverse cupin protein superfamily. Cupin proteins share similar primary and tertiary structure with two conserved His-containing,  $Mn^{2+}$ -binding motifs separated by an intermotif region that varies in length. ODCs belong to the bicupin subclass due to a duplication of the cupin domain.

In basidiomycetous fungi, ODC has been proposed to act sequentially with its reaction product, formate, degrading intracellular enzyme formate dehydrogenase (FDH, EC 1.2.1.2). FDHs are a heterogeneous group of prokaryotic and eukaryotic enzymes that catalyze the oxidation of formate to  $CO_2$  and H<sup>+</sup>. In aerobic organisms, the FDHs are mainly NAD<sup>+</sup>-dependent enzymes resulting in the formation of CO<sub>2</sub> and NADH. In filamentous fungi, it has been hypothesized that the NADH produced by FDH may be used for ATP synthesis during the fungal vegetative growth.

This study was conducted as a part of the DOE JGI SAP project on comparative genomics of saprobic Agaricomycotina fungi. We looked for the protein model sequences containing conserved bicupin domains characteristic to ODCs within the genomes of 21 basidiomycetous fungal species belonging to the subphylum Agaricomycotina (http://genome.jgi-psf.org/programs/fungi/index.jsf). FDH models were collected from the same annotated genomes by using FDH as a search term. We found 75 putative ODC and 43 putative FDH encoding gene models from 19 fungal species. Typically, several ODCs with varying polypeptide length (400-500 amino-acids) and with an average of 443 amino-acids were found within one fungal species. In the case of FDH, a lower number of isozymes were detected within a single genome. Protein phylogeny and selection pressure of ODCs and FDHs are analyzed to give better evolutionary insight for their functions and eco-physiology in wood and plant material degradation.

### The Role of RWA Proteins in *O*-acetylation of Cell Wall Polysaccharides

**Yuzuki Manabe**<sup>1</sup>\* (YManabe@lbl.gov), Yves Verhertbruggen,<sup>1</sup> Emilie A. Rennie,<sup>1</sup> Soe M. Htwe,<sup>1</sup> Dominique Loque,<sup>1</sup> Majse Nafisi,<sup>3</sup> Caroline Orfila,<sup>3</sup> J. Paul Knox,<sup>4</sup> Sascha Gille,<sup>2</sup> Markus Pauly,<sup>2</sup> Yumiko Sakuragi,<sup>3</sup> and Henrik Vibe Scheller<sup>1,2</sup>

<sup>1</sup>Feedstocks Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, California; <sup>2</sup>Department of Plant & Microbial Biology, University of California, Berkeley, California; <sup>3</sup>Department of Plant Biology and Biotechnology University of Copenhagen, Denmark; and <sup>4</sup>Centre for Plant Sciences, University of Leeds, United Kingdom

Many different glycans in plant cell walls are acetyl esterified; however, the enzymes involved in the acetylation have not been identified.

The Cas1p protein in the fungus *Crypotococcus neoformans* is required for *O*-acetylation of glucuronoxylomannan in the capsule. Homologs of Cas1p are also present in animals, but their function is not known. Plant genomes also encode proteins that are related to Cas1p. Thus, *Arabidopsis thaliana* has four homologues of the *Cas1p* gene. Arabidopsis mutants with insertion in the respective genes were identified, and we found that at least one of the mutants, designated *reduced wall acetylation 2 (rwa2)* had decreased levels of acetylated cell wall polymers. Two independent alleles of *rwa2* mutants were examined

and the plants were shown to contain about 20% lower amounts of wall bound acetic acid compared to wild type. The same level of acetate deficiency was found in different pectic polymers and in xyloglucan. Thus, the *rwa2* mutation affects different polymers to the same extent.

There was no visible difference observed between wild type and *rwa2* mutants at any developmental stages. However, the *rwa2* mutants displayed increased resistance toward *Botrytis cinerea*, a necrotrophic fungus. The other mutants, *rwa1*, *rwa3*, and *rwa4*, did not have measureable changes in acetylation, presumably due to genetic redundancy. Double, triple and quadruple *rwa* mutants are currently being investigated. RWA double mutants display no or minimal morphological phenotype and the acetylation level is reduced for up to 25%. In contrast, three of triple mutants have severe growth deterioration and their phenotypes are quite distinctive from each other.

The data indicate that RWA proteins function at a biochemical step prior to the actual acetyl transferase reaction. Bioinformatic studies have identified the DUF231 class of proteins in plants that may function in complexes with RWA proteins and mediate polymer specific acetylation.

#### Defining the Maize Transcriptome de novo Using Deep RNA-Seq

Jeffrey Martin\* (jamartin@lbl.gov), Stephen Gross, Cindy Choi, Tao Zhang, Erika Lindquist, Chia-Lin Wei, and Zhong Wang

DOE Joint Genome Institute, Walnut Creek, California

*de novo* assembly of the transcriptome is crucial for functional genomics studies with bioenergy crops, since many of them lack high quality reference genomes. In a previous study we successfully *de novo* assembled simple eukaryote transcriptomes exclusively from short Illumina RNA-Seq data. However, extensive alternative splicing, present in most of the higher eukaryotes, poses a significant challenge for current short read assembly processes. Gene duplications retained from ancestral polyploidization events, common in plant genomes, also present challenges in assembly of individual transcripts from distinct genes. Here we present preliminary results which greatly improved the assembly of the maize transcriptome, using combined experimental and informatics strategies to resolve transcript variants.

We chose the maize transcriptome as a test case since the reference genome can be used for assessing the quality of the assembled transcript variants. Our experimental strategies include ultra deep sequencing and multiple libraries with various insert lengths. We generated 78 gigabases (306 million read pairs) of both stranded and non-stranded RNA-Seq data by sequencing three libraries made from a seedling mRNA sample. The first library was a 180bp insert library; the second was a 250bp tight-insert library and was sequenced 2x151bp and subsequently the two read pairs were joined to form 250bp reads; while the third library was a 500bp tight-insert library to provide long-range connectivity. We further improved our published *Rnnotator* pipeline to assemble the reads from all libraries into transcripts. By comparing these *de novo* assembled transcripts to the reference-based gene models we evaluated the performance of our transcript variants and transcripts from duplicate genes. In addition, we also evaluated the potential of combining reference-based and *de novo* assembly approaches to leverage the strengths of both strategies to further improve transcriptome annotation. In summary, we expect our strategies will be generically applicable to many plant transcriptome studies. The maize gene models derived in this study can serve as a valuable resource for the maize research community.

#### A Reasonable Resequencing Pipeline for Next Generation Sequencing Data at the JGI

Joel Martin\* (j\_martin@lbl.gov), Wendy Schackwitz, Anna Lipzen, and Len Pennacchio

DOE Joint Genome Institute, Walnut Creek, California

The continued acceleration of throughput and quality from next generation sequencing platforms provides increased opportunities for detection of short insertions and deletions (INDELS), single nucleotide variation(SNV) and structural variation (SV) between strains and among populations of microbes, fungi and plants. The challenge of analyzing this increased volume of data continues apace, exacerbated by the multiform characteristics of the different species sequenced at the JGI in pursuit of bioenergy, understanding global carbon cycling and biogeochemistry. This requires an automated pipeline to ensure standard and systematic results while allowing the flexibility of detailed manual inspection and analysis of individual projects. We present the results of microbes, plants and fungi as analyzed with our current pipeline, compare false positive/negative estimates between various aligners and detection tools for SNVs, INDELs and SVs. An overview will also be given of visualization and navigation tools available to collaborators for working with the data files delivered by the resequencing group.

# **'Omics' Analyses of the Deep-sea Microbial Community Response to the Deepwater Horizon Oil Spill**

**Olivia U. Mason**<sup>1\*</sup> (oumason@lbl.gov), Terry C. Hazen,<sup>1</sup> Patrick Chain,<sup>2</sup> Eric A. Dubinsky,<sup>1</sup> Julian Fortney,<sup>1</sup> James Han,<sup>3</sup> Jenni Hultman,<sup>1</sup> Regina Lamendella,<sup>1</sup> Rachel Mackelprang,<sup>3</sup> Lauren M. Tom,<sup>1</sup> Susannah G. Tringe,<sup>3</sup> Tanja Woyke,<sup>3</sup> Edward M. Rubin,<sup>1,3</sup> and Janet K. Jansson<sup>1,3</sup>

<sup>1</sup>Lawrence Berkeley National Laboratory, Berkeley, California; <sup>2</sup>Lawrence Livermore National Laboratory, Livermore, California; and <sup>3</sup>DOE Joint Genome Institute, Walnut Creek, California

The Deepwater Horizon oil spill is the largest spill in US history. To assess the response of the deep-sea microbial communities in the Gulf of Mexico to this large-scale environmental perturbation two plume samples (~1.6 and 10 km SW from the wellhead) and one uncontaminated, control sample collected from plume depth (~40 km WSW from the wellhead) were analyzed. Microbial diversity was assessed using 454-pyrotag sequencing of 16S rRNA genes. Metagenomic, transcriptomic, and single cell genome sequencing was carried out using the Illumina sequencing platform. Pyrotag sequence analysis revealed that microbial diversity was significantly lower in the plume, with the order Oceanospirillales comprising 80% of the proximal and 90% of the distal plume sample, compared to 3% of the control sample. Analysis of assembled metagenomic sequences revealed that proteobacterial genes dominated all samples, particularly the genus *Colwellia* and the order *Oceanospirillales*. Numerous COGs were significantly different between the contaminated and uncontaminated samples. For example, a methylaccepting chemotaxis protein, a protein shown to be involved in bacterial chemotactic response to hydrocarbons, was the most statistically significant (higher relative abundance) in both plume samples relative to the control. Further, unassembled reads were compared

to a novel database of proteins involved in hydrocarbon degradation using the tblastn algorithm. Proteins involved in degradation of simple aromatics and alkanes were more abundant in the plume interval compared to the uncontaminated sample. Proteins involved in methane oxidation (*Pmo*) were lower in the proximal sample, but higher in the distal plume sample relative to the control. Analysis of both the assembled and unassembled transcriptomic data confirmed that proteins involved in hydrocarbon degradation that were identified as most abundant in the plume relative to the control were, in fact, expressed *in situ*. Further, methyl accepting chemotaxis proteins were expressed in the plume interval. Genomic analysis of *Oceanospirillales* single cells obtained by flow cytometry and sequenced using the Illumina sequencing platform validated our metagenome and transcriptome derived findings in that both single cell genomes encode hydrocarbon degradation and chemotaxis. In conclusion, pyrotag data suggested *Oceanospirillales* dominated the plume interval. Analysis of the plume metagenome and transcriptome revealed that Bacteria in the plume exhibited the potential for, and in fact expressed, chemotactic behavior and the ability to degrade hydrocarbons.

Finally, single cell genome sequencing of two *Oceanospirillales* obtained from the plume interval linked function with phylogeney and confirmed that this order was indeed likely involved in degrading hydrocarbons from the Deepwater Horizon oil spill.

#### The JGI Annotation Pipeline for Genomes and Metagenomes

**K. Mavromatis**, M. Huntemann\* (MHuntemann@lbl.gov), P. Williams, A. Pati, N. Ivanova, and N.C. Kyrpides

DOE Joint Genome Institute, Walnut Creek, California

The JGI annotation pipeline is applied on genomes and metagenomes sequenced at the DOE-Joint Genome Institute, and is offered as a service to the community through the IMG web submission portal. It utilizes state of the art methods for *Quality Assessment* of sequences and tools for the *prediction of genes and other features*. These tools are constantly evaluated using simulated sets and the pipeline is updated in order to handle data from new sequencing technologies. The annotated sequences are subsequently integrated in the Integrated Microbial Genomes system where the *functional annotation* takes place, using a hierarchical set of rules, and become available to the users.

#### Pangenomes: A New Paradigm in Genome Analysis

K. Mavromatis\* (KMavromatis@lbl.gov), H. Marcel, N. Ivanova N, and N.C. Kyrpides

DOE Joint Genome Institute, Walnut Creek, California

Sequencing of a large number of related microbial species has made clear that no single genome can describe a species while the processes that are involved in such comparative analyses are computationally expensive. The concept of a pangenome refers to a composite entity, composed of species representing the sum of all genes present in the genomes of different strains belonging to a given species. It consists of the non-redundant union of features for the selected group of genomes, which can be further divided, in the core (i.e. the genes present in all of the participating strains), the variable part (the genes present in some but not all of the strains) and the unique part (genes that are found in one organism only). This approach has the clear advantage of data compression, since similar features are collapsed and represented by one. It also facilitates their comparative analysis by allowing the rapid identification of common genes, the determination of conserved

Posters alphabetical by first author. \*Presenting author

syntenous regions among the strains, the direct comparison of their metabolic content and provides insights into their evolution. By keeping pointers to the genes that are collapsed in the pangenome we provide gene centric analysis such as sequence comparisons, identification of small sequence variations e.g. SNPs etc. We have developed methods that generate pangenomes based on the identification of orthologous genes and conserved chromosomal regions, and tools for their visualization and exploration. These datasets and tools are available through the Integrated Microbial Genomes system (IMG).

#### Isotopic Analysis of RNA Microarrays Shows Microbial Resource Use Profiles Decoupled from Phylogeny

**Xavier Mayali**<sup>1</sup>\* (mayali1@llnl.gov), Peter K. Weber,<sup>1</sup> Eoin L. Brodie,<sup>2</sup> Shalini Mabery,<sup>1</sup> Paul Hoeprich,<sup>1</sup> and Jennifer Pett-Ridge<sup>1</sup>

<sup>1</sup>Physical and Life Science Directorate, Lawrence Livermore National Laboratory, Livermore, California, and <sup>2</sup>Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California

Most microorganisms remain uncultivated, and typically their ecological roles must be inferred from diversity and genomic studies. To directly measure functional roles of uncultivated microbes, we developed chip-SIP, a high-sensitivity, high-throughput stable isotope probing (SIP) method performed on a phylogenetic microarray. We incubated a microbial community with isotopically labeled substrates, hybridized community rRNA to a microarray, and measured isotope incorporation—and therefore substrate use—by secondary ion mass spectrometer imaging (NanoSIMS). Chip-SIP analysis of an estuarine community quantified amino acid, nucleic acid or fatty acid incorporation by 81 taxa. The resulting resource use profile demonstrates that bacterial functional capacity can be decoupled from phylogeny. This approach provides a means to test genomics-generated hypotheses about biogeochemical function in natural environments.

### A Phylogenetic Classification of Bacterial Multiheme Cytochromes c Reveals Diverse Families and Horizontal Gene Transfer between Distant Phyla

**Ryan A. Melnyk**<sup>1</sup>\* (melnyk@berkeley.edu), Hans K. Carlson,<sup>1</sup> Kelly C. Wrighton,<sup>1,2</sup> and John D. Coates<sup>1</sup>

<sup>1</sup>Department of Plant & Microbial Biology, University of California, Berkeley, and <sup>2</sup>Department of Environmental Science, Policy, and Management, University of California, Berkeley

Bacterial multiheme cytochromes c (MHCs) are a heterogeneous group of proteins that transport electrons in various respiratory processes. MHCs are characterized by the presence of multiple conserved heme-binding motifs and significant sequence divergence elsewhere, which has impeded the understanding of the evolution of the superfamily as a whole. Here we present the analysis of the MHC complements of 1041 bacterial proteomes, in which we employed a bioinformatic screen to identify 2789 MHCs. Use of a spectral clustering algorithm (SCPS) enabled the grouping of the majority (82%) of the protein sequences into 15 monophyletic clusters with homologous domains. Several of these clusters were phylogenetically diverse, containing divergent protein sequences from ten or more phyla, suggesting an ancient origin of several MHC subfamilies. Within these clusters,  $\delta$ - and  $\epsilon$ -Proteobacterial MHCs were often more closely related to the Firmicutes MHCs rather than the MHCs of the other Proteobacterial clades, a distribution that is

similar to the distribution of the two bacterial systems for cytochrome *c* biogenesis. We explore one 'case study' of MHCs and biogenesis proteins transferred between the  $\delta$ -Proteobacteria *Geobacter* genus and the Firmicutes *Thermincola potens* strain JR, and identify several unidentified protein families with conserved sequence and synteny between the two groups, predicting a possible role as novel factors in MHC biogenesis or function. Our results suggest that the evolution of modern MHC families was not strictly vertical, but fraught with transfer between divergent taxonomic groups. Preference for one biogenesis system may be a factor in shaping horizontal gene transfer, as related MHCs often come from organisms with similar biogenesis systems. Additionally, it seems that duplication of MHCs and rapid expansion of unique families is happening in multiple lineages, in a molecular example of convergent evolution.

#### The RNA-Seq Analysis Pipeline on Galaxy

Xiandong Meng<sup>1,2</sup>\*(xiandongmeng@lbl.gov), Jeffrey Martin,<sup>1,2</sup> and Zhong Wang<sup>1,2</sup>

<sup>1</sup>DOE Joint Genome Institute, Walnut Creek, California, and <sup>2</sup>Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California

We have developed a pipeline for standard RNA-Seq analysis and deployed it onto the JGI Galaxy server. The pipeline contains four modules: QC, Counting, Statistics and de novo Assembly. This web-based bioinformatics pipeline enables biologists with little programming expertise to perform standard RNA-seq data analysis. Furthermore, biologists can tweak the exposed options of each module to fine-tune their analysis. The pipeline is preconfigured to run on the JGI's high performance computing resources for optimal performance.

#### Functional Screening of Metagenomic Libraries for Glycosyl Hydrolase Activity

Keith Mewis\* (kmewis@interchange.ubc.ca), Marcus Taupp, and Steven Hallam

University of British Columbia, Vancouver, British Columbia, Canada

Cellulose, the most abundant source of organic carbon on the planet, has wide-ranging industrial applications with increasing emphasis on biofuel production. Environmental genomics, or metagenomics, have provided a new tool to bridge the cultivation gap in search of novel bioconversion enzymes. Here we present a solution-based screen for glycosyl hydrolase activity using both a chromogenic and fluorescent substrate to screen for cellulase and 1,4-xylanase activity, respectively. This screen is sensitive, quantitative, and automated with a throughput of more than one hundred 384-well microtiter plates per day. We applied this method to large-insert metagenomic libraries derived from two different environments, yielding more than 50 clones showing activity from at least 6 different CAZy (http://www.cazy.org) glycosyl hydrolase families (GHFs). These include a number of novel enzymes, bringing function to genes previously identified through *in silico* methods. Future directions will involve the screening of additional environments and the use of new substrates to functionally identify further GHF genes and their relationships to eachother.

### The Chimeric Genome of Emiliania huxleyi

John Miller\* (jjmiller@umd.edu) and Charles Delwiche

University of Maryland, College Park, Maryland

*Emiliania huxleyi* is a haptophyte alga and a prominent member of the marine phytoplankton in many areas of the world. Haptophytes -- like the other "chromalyeolate" taxa, including dinoflagellates, cryptomonads, and heterokonts -- possess secondary plastids of red algal origin and are expected to have chimeric genomes. To test this hypothesis, bipartitions were counted from a group of phylogenetic protein trees. This allowed data to be quantified from trees sharing as few as one taxon. *Emiliania* grouped most frequently with heterokonts (291). When sequences linking *Emiliania* with heterokonts were concatenated, this haptophyte alied itself to the phototrophic heterokonts with strong bootstrap support. The oomycetes were the outgroup to the *Emiliania*/phototrophic heterokont bipartition. The second most frequent lineage for *Emiliania* to group with was the green lineage (i.e., Viridiplantae)(139). Concatenated trees linking *Emiliania* to the Viridiplantae placed *E. huxlevi* as an outgroup to the green lineage with high bootstrap support. E. huxlevi only grouped with rhodophytes 39 times, but grouped with Cyanidioscyzon merolae more frequently (39) than any single organism of the Viridiplantae. *Emiliania* was linked to both the heterokonts and the green lineage in 96 trees. When these alignments were concatenated, *Emiliania* was alied to the heterokonts with the green lineage sister to the Emiliania/heterokont bipartition. The larger number of proteins linking *Emiliania* to the green lineage compared to those linking it to the heterokonts and the green lineage together may imply evolutionary differences like differential lineage sorting or separate plastid acquisitions. When poorly aligned sequences were removed, the same relationships occurred. When only trees in which the branch linking E. huxleyi to the clade of interest had bootstrap values above 70 were included, E. huxlevi again grouped with the heterokonts far more frequently than with any other bipartition, but with a disproportionate loss of trees supporting an E. huxlevi/Viridiplantae bipartion. Additional support for the chimeric nature of the E. huxleyi genome was provided by the non-random distribution of genes within the E. *huxleyi* genome linking it to the heterokonts and separately to the Viridiplantae implying that genes from different sources are syntenous. This study was limited by the lack of data from cryptomonads, dinoflagellates, and rhodophytes in comparison to the relatively rich datasets for heterokonts and the Viridiplantae. Haptophytes may be more closely related to another chromalyeolate clade, such as the cryptomonads, than to heterokonts but because of the lack of cryptomonad data such a pattern would have been difficult to detect in this study. Although it is possible that the *Emiliania* genome contains genes derived from a green alga, it is also possible that the large green signal represents a primary plastid lineage signal. If so, one would predict that this apparent signal will disappear as more rhodophyte genomes become available. Despite these limitations, the data show clearly that the *E. huxlevi* genome is chimeric, and that it has large contributions linking it to heterokonts and primary plastid containing lineages.

### Genome-wide Assessment of Switchgrass (*Panicum virgatum*) Genetic Diversity and Population Structure Using *de novo* Genotyping-by-sequencing (GBS)

Geoffrey P. Morris\* (gmorris@uchicago.edu), Paul Grabowski, and Justin Borevitz

Department of Ecology & Evolution, University of Chicago, Chicago, Illinois

Switchgrass is an emerging bioenergy crop and model for the genetics of ecophysiological adaptation. Characterizing genetic diversity and population structure of switchgrass is an important step towards the identification of habitat-specific ecotypes, unusual or divergent lines, and functional loci through genome-wide association studies. Because traditional genotyping methods are laborious, previous studies have been limited in terms of number of loci and subject to ascertainment bias. Therefore, we have employed new de novo genotyping-by-sequencing methods (Illumina sequencing of multiplexed, reducedrepresentation libraries) to characterize switchgrass varieties from across its range. We obtained ~210 million Illumina reads from eight cultivated switchgrass varieties and 23 wild accessions from extreme micro-environments in a Great Lakes dune system. Using *de novo* assembly, read mapping, and custom genotyping algorithms we identified  $\sim$ 5000 single nucleotide polymorphisms. While we find a signal of isolation-by-distance across the continental range, the local population also harbors high levels of genetic diversity which warrants further investigation. This strategy of short-read sequencing from multiplexed, reduced-representation libraries is effective at assessing genetic diversity and population structure in organisms without a reference genome.

#### Understanding a Three-way Symbiosis Using Transcriptome Analyses of the Fungal Partner *Curvularia Protuberata*

Mustafa R. Morsy\* (mrmorsy@noble.org) and Marilyn J. Roossinck

The Samuel Roberts Noble Foundation, Ardmore, Oklahoma

In Yellowstone National Park, tropical panic grass (*Dichanthelium lanuginosum*) can survive soil temperatures up to 55 °C in geothermal areas, due to a mutualistic association with the endophytic fungus *Curvularia protuberata* carrying a mycovirus, Curvularia thermotolerant virus (CThTV). Curing the *C. protuberata* of CThTV abolishes the fungal ability to provide plants with heat tolerance showing that the three partners are required for survival at elevated temperatures. The fungus carrying the virus provides heat stress tolerance not only to the native host but also to many other crop plants including eudicot crop plants. No information is known about the molecular mechanism of the heat tolerance acquired by this three-way symbiosis. The genome of *C. protuberata* has not been sequenced. To gain insights into the biology of *C. protuberata* and the role of the CThTV in increasing plant thermotolerance, with the JGI we generated a large numbers of ESTs from mycelial cultures of *C. protuberata* with or without CThTV growing under control and heat stress conditions. We will present a preliminary analysis of generated ESTs as well as metabolome comparisons between the wild type fungus containing the virus and a virus free isolate.

#### Novel Thermophilic Microorganisms and Cellulases for Improving Second-generation Biofuel Technologies

**Senthil Kumar Murugapiran**<sup>1</sup>\* (Senthil.Murugapiran@unlv.edu), Jeremy A. Dodsworth,<sup>1</sup> Jessica Guy,<sup>1</sup> Joseph Peacock,<sup>1</sup> Tanja Woyke,<sup>2</sup> Susannah G. Tringe,<sup>2</sup> and Brian P. Hedlund<sup>1</sup>

<sup>1</sup>School of Life Sciences, University of Nevada, Las Vegas, Nevada, and <sup>2</sup>DOE Joint Genome Institute, Walnut Creek, California

The main goal of second-generation biofuels is to replace valuable foodstuffs such as corn and cane sugar with lignocellulosic feedstock as carbon sources for microbial fermentation to yield bioethanol. We set up eight different *in situ* cellulolytic enrichments in Great Boiling Spring (GBS), Nevada, including two incubation temperatures (77°C and 85°C), two cellulose feedstocks (corn stover and aspen sawdust), and two incubation locations within the spring (sediment and water column). Each enrichment was used as inoculum for lab cultivation of cellulolytic thermophiles and 16S rRNA gene pyrotag analysis. Subsequently, four enrichments were chosen for metagenomic analysis. Aerobic isolates have been identified by 16S rRNA gene sequencing as: Geobacillus vulcani, G. stearothermophilus, G. thermoleovrans, and G. lituanicus, however, none of the *Geobacillus* isolates are active on crystalline cellulose. The anaerobic pure cultures have been identified as *Thermotoga petrophila*, which is known to be active on soluble cellulose. Additional more strongly cellulolytic lab enrichments are the focus of current study. 16S rRNA pyrosequencing results showed that 676 microbial species-level operational taxonomic units and 55 microbial phyla were represented. The sixteen most abundant phyla represented 86-99% of all species in each of the enrichments. Remarkably, five of the most abundant phyla have never been cultivated in the laboratory and may include novel cellulolytic thermophiles. Unifrac analysis showed that temperature, cellulose feedstock, and incubation location, in that order, control the microbial community composition of the cellulolytic enrichments. In addition, our analysis shows that phyla containing known cellulolytic organisms are enriched at both 77°C and 85°C, including *Thermotogae*, *Dictyoglomus*, and *Firmicutes*. Candidate phyla OP9, C2, and OP10 are also enriched with cellulose, however it is possible that these organisms may be involved in consortial catabolism of cellulose but not be primary cellulolytic organisms. More than 100 contigs encoding putative cellulase genes are present in metagenomes from cellulolytic enrichments. Identification of novel cellulolytic thermophilic microorganisms and thermotolerant enzymes may help overcome a major hurdle in the conversion of complex lignocellulose feedstocks into fermentable sugars or directly into biofuels by reducing the energy and water requirements for this intricate process.

#### Serratia spp – Their Role in Antagonism against Oilseed Rape Pathogens and in Plant Growth Promotion

Saraswoti Neupane\* (Saraswoti.Neupane@slu.se), Nils Högberg, Sadhna Alström, Björn Andersson, and Roger Finlay

Uppsala BioCenter, Department of Forest Mycology & Pathology, Uppsala, Sweden

In Sweden, poor emergence and seedling establishment in oilseed rape production is a major challenge due to harsh winters and infestation by soil borne pathogens in intensive cultivation. *Rhizoctonia solani* and *Verticillium dahliae* are, two major soil-borne pathogens of oilseed rape, responsible for substantial yield losses. The lack of effective fungicides and resistant cultivars to control these has encouraged research into finding

alternative means. The use of plant growth promoting and pathogen inhibiting (antagonistic) bacteria is an environmentally friendly alternative approach. In this study, four *Serratia* spp, indigenous to oilseed rape, were evaluated for their plant growth stimulating activity and control of soil borne pathogens using *in vitro*, greenhouse and field experiments. We observed *Serratia* spp have the ability to inhibit *R. solani* growth and to stimulate plant growth but the degree of antagonism and plant growth stimulation varied among strains. We also observed a significant effect on seed germination, seedling establishment and yield increment in field trials. The genetic regulation underlying antagonism against plant pathogens and plant growth stimulation is currently being studied by using combined, comparative genomic and transcriptomic approaches.

#### Genome Sequencing of *Nitrosomonas sp.* AL212, an Ammonia-Oxidizing Bacterium Adapted for Growth at Low Ammonia Concentrations

**Jeanette M. Norton**<sup>1</sup>\* (jeanette.norton@usu.edu), Yuichi Suwa,<sup>2</sup> Annette Bollmann,<sup>3</sup> Martin G. Klotz,<sup>4</sup> Lisa. Y. Stein,<sup>5</sup> Hendrikus J. Laanbroek,<sup>6</sup> Daniel J. Arp,<sup>7</sup> and Lynne A. Goodwin<sup>8</sup>

<sup>1</sup>Utah State University, Logan, Utah; <sup>2</sup>Chuo University, Tokyo, Japan; <sup>3</sup>Miami University, Oxford, Ohio; <sup>4</sup>University of Louisville, Louisville, Kentucky; <sup>5</sup>University of Alberta, Edmonton, Alberta, Canada; <sup>6</sup>Netherlands Institute of Ecology, Utrecht, Netherlands; <sup>7</sup>Oregon State University, Corvallis, Oregon; and <sup>8</sup>DOE Joint Genome Institute, Los Alamos National Laboratory, Los Alamos, New Mexico

Ammonia-oxidizing bacteria (AOB) in the genus Nitrosomonas have several distinct lineages that are associated with ecophysiological traits. The Nitrosomonas oligotropha lineage (also known as cluster 6A) includes the type species Nitrosomonas oligotropha and *Nitrosomonas ureae* and several *Nitrosomonas* isolates including AL212 (Suwa et al. 1994) and IS-79 (Annette Bollmann). This group is distinguished by higher substrate affinity (low Km), lower growth rates and increased sensitivity to high ammonia concentrations. These traits contrast to the ammonia tolerance of strains related to N. europaea and N. eutropha. Similar oligotrophic AOB have been found to be widely distributed in the environment. Interest in the sequencing and study of this group of organisms remains high as these have been detected in drinking water supply and wastewater treatment systems. Nitrosomonas strain AL212 was isolated from sewage sludge (Suwa et al 1994) and genomic DNA was harvested at the lab of Yuichi Suwa and sent for sequencing at the Joint Genome Institute (DOE). Draft sequence results were available in December 2008. With the harvesting of additional DNA, the genome sequence was finished and closed in December 2010. The Nitrosomonas sp. AL212 consists of a chromosome and two plasmids totaling 3.3 Mb at 44.7 % GC with 2983 candidate proteinencoding gene models. The complete genome revealed important differentiations in key functional genes and genome structure. Several functional gene clusters are more closely related to AOB from the *Nitrosospira* lineage than to other sequenced *Nitrosomonas* spp. (for example the top KEGG BLAST hits numbered 286 to Nitrosomonas eutropha, 440 to Nitrosomonas europaea and 847 to Nitrosospira multiformis). Our initial focus is on the gene clusters encoding ammonia oxidation, carbon fixation, urea hydrolysis and nitrite reduction. Three nearly identical copies of the *amo* operon, two types of the RuBisCo encoding gene cluster (green-like and red-like) and several nitrite reductases were identified and examined. Implications for evolutionary lines of descent and functional diversity in the AOB are described.

Acknowledgments/References:

 Bollmann, A., and H.J. Laanbroek. 2001. Continuous culture enrichments of ammoniaoxidizing bacteria at low ammonium concentrations. FEMS Microbiol. Ecol. 37:211-221.
Suwa, Y., Y. Imamura, T. Suzuki, T. Tashiro, and Y. Urushigawa. 1994. Ammoniaoxidizing bacteria with different sensitivities to (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> in activated sludges. Water Res. 28:1523-1532.

#### Unraveling Plant Xylan Biosynthesis Using Co-expression, Co-localization, and Protein-Protein Interaction Analysis

**Ai Oikawa**<sup>1</sup>\* (aoikawa@lbl.gov), Casper Nicholas Søgaard,<sup>2</sup> Yumiko Sakuragi,<sup>2</sup> Emilie A. Rennie,<sup>1</sup> Stephanie N. Morrison,<sup>1</sup> Peter McInerney,<sup>1</sup> Masood Z. Hadi,<sup>1</sup> Hiren J. Joshi,<sup>1</sup> Joshua L. Heazlewood,<sup>1</sup> and Henrik Vibe Scheller<sup>1</sup>

<sup>1</sup>Joint BioEnergy Institute, Emeryville, California, and <sup>2</sup>The Department of Plant Biology and Biotechnology, University of Copenhagen, Copenhagen, Denmark

Xylans constitute the major non-cellulosic component of plant biomass. Xylan biosynthesis is particularly pronounced in cells with secondary walls, implying that the synthesis network consists of a set of highly expressed genes in such cells. To improve the understanding of xylan biosynthesis, we performed a comparative analysis of coexpression networks between Arabidopsis and rice as reference species with different wall types. Many co-expressed genes were represented by orthologs in both species, which implies common biological features, while some gene families were only found in one of the species, and therefore likely to be related to differences in their cell walls.

In addition to the information from co-expression, knowledge of sub-cellular localization of the corresponding proteins contributes to our understanding of protein function and putative interactions. Biosynthesis of hemicelluloses, including xylan, takes place in the endomembrane system, particularly in the Golgi apparatus. To predict the subcellular location of the identified proteins, we developed a new method, PFANTOM (glant protein <u>family information-based predictor</u> for endo<u>m</u>embrane), which was shown to perform better for proteins in the endomembrane system than other available prediction methods. By the combined *in silico* approaches of expression profiling and localization prediction, we identified novel putative xylan synthesis components in the Golgi apparatus, including glycosyl<u>t</u>ransferases and glycosyltransferase-like proteins.

The co-expression and co-localization of the glycosyl<u>t</u>ransferases suggested the possibility that some of them form a xylan synthase complex. Based on this hypothesis, we investigated the interaction of several of the xylan synthase candidates in Arabidopsis by BiFC (<u>bi</u>molecular <u>f</u>luorescence <u>c</u>omplementation) analysis. The observed protein-protein interactions support the presence of one or more protein complexes involved in xylan biosynthesis. This work was supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 with Lawrence Berkeley National Laboratory.

#### References

An integrative approach to the identification of Arabidopsis and rice genes involved in xylan and secondary wall development. (2010) Oikawa A, Joshi HJ, Rennie EA, Ebert B, Manisseri C, Heazlewood JL, Scheller HV. *PLoS One.* 5(11):e15481.
Visual mapping of cell wall biosynthesis. (2011) Sakuragi Y, Nørholm MH, Scheller HV. *Methods Mol Biol.* 715:153-67.

### The Mitochondrial Genome of the White Rot Fungus *Phlebia radiata* (Basidiomycota, Agaricomycotina) for Dissection of the Wood-degradation Biology

**Ilona Oksanen**<sup>1</sup>\* (ilona.oksanen@helsinki.fi), Pia K. Laine,<sup>2</sup> Jaana Ekojärvi,<sup>1</sup> Miia R. Mäkelä,<sup>1</sup> Lars Paulin,<sup>2</sup> and Taina K. Lundell<sup>1</sup>

<sup>1</sup>Fungal Biotechnology Laboratory, Department of Food and Environmental Sciences, University of Helsinki, Helsinki, Finland, and <sup>2</sup>DNA Sequencing and Genomics Laboratory, Institute of Biotechnology, University of Helsinki, Helsinki, Finland

Before being able to improve the industrial enzymatic cellulose degradation and conversion leading to more sustainable production of biofuels from plant waste and novel raw materials, we need to understand the biochemical pathways connected to high activities of the plant lignin and cellulose degrading enzymes. Thus, we aim to dissect the wood degradation biology of our model fungus, Phlebia radiata Fr. strain 79 (FBCC 43) isolated in Finland that is capable of degrading all wood components efficiently proven by several physiological, biochemical, and genetic studies. We will analyze the whole transcriptome and proteome of P. radiata during wood degradation (see abstract by Ekojärvi et al.). As a background for transcriptomics and proteomics and finally for gene silencing, we will sequence the whole genome of *P. radiata*. Thus far, we have sequenced the mitochondrial genome and the nuclear genome is in production. Single stranded DNA (sstDNA) was sequenced using the 454 sequencing technology with GS FLX Titanium chemistry resulting in high average sequence coverage of  $\sim 300x$ . The draft of the mitochondrial genome is a large, single 156 kbp circular molecule with GC percentage of 31 and containing a large amount of non-coding regions. The program tRNAscan-SE identified 28 tRNA genes corresponding to all 20 amino acids. Eleven of the tRNA genes are anticlockwise orientated. Complete annotation of the mitochondrial genome, comparative genomics, and evolutionary analyses are under way.

### Genomes On-line Database (GOLD), Project Catalog, Metadata Standards and Ontologies for Genomes and Metagenomes

I. Pagani\* (IPagani@lbl.gov), N.C. Kyrpides, and K. Liolios

DOE Joint Genome Institute, Walnut Creek, California

The Genomes On Line Database (GOLD) is a comprehensive resource for centralized monitoring of genome and metagenome projects worldwide. Both complete and ongoing projects, along with their associated metadata, can be accessed in GOLD through precomputed tables and a search page. As of March 2011, GOLD contains information for almost 10000 sequencing projects, of which more than 1600 have been completed and their sequence data deposited in a public repository. GOLD continues to expand, moving toward the goal of providing the most comprehensive repository of metadata information related to the projects and their organisms/environments in accordance with the Minimum Information about a (Meta)Genome Sequence (MIGS/MIMS) specification. This expansion leads GOLD to explore mechanisms that standardize the description of genomes and facilitate and easy exchange and integration of genomic metadata. GOLD is available at: http://www.genomesonline.org.

# **Comparative Study of Basidiomycete Telomeres and Subtelomeric Regions**

Gúmer Pérez, Antonio G. Pisabarro, and Lucía Ramírez\* (lramirez@unavarra.es)

Genetics and Microbiology Group, Public University of Navarre, Pamplona, Spain

Telomeres and the subtelomeric regions are usually scarcely studied in genome sequence projects because of their repetitive nature and the occurrence of chromosomal rearrangements breaking down the synteny between these regions in closely related species. Fungal telomeres have been widely studied in Ascomycetes. In basidiomycetes, however, despite a number of fungal genomes (many of them corresponding to generi involved white biotechnology processes) have been sequenced, their telomeric and subtelomeric regions are unknown. The study of these regions is of great importance since they have been described as harbouring secondary metabolite clusters.

We have analysed the telomeric and subtelomeric regions in *Pleurotus ostreatus* using molecular and bioinformatics tools and, here we perform a comparative study of these sequences in other sequenced basidiomycetes: *Agaricus bisporus, Ceriporiopsis subvermispora, Heterobasidion annosum, Phanerochaete chrysosporium* and *Postia placenta* using a bioinformatics approach with the purpose of: i) determining the arrangement of genomes in putative linkage groups in species with no genetic maps available, ii) studying the synteny of the linkage groups arisen after informatics analysis of these genomes with the from *P. ostreatus*, and iii) determining the presence of similar genes the in subtelomeric regions of different genomes with different evolutionary history.

#### Comparative Whole Deep Sequencing Transcriptome Analysis of the Dikaryotic Lifestyle and the Lignocellulolytic Strategies in the Model White Rot *Pleurotus ostreatus*

Antonio G. Pisabarro\* (gpisabarro@unavarra.es), Gúmer Pérez, Francisco Santoyo, and Lucía Ramírez

Genetics and Microbiology Research Group, Public University of Navarre, Pamplona, Spain

The new sequencing technologies are more powerful tools for transcriptome studies than the classic microarrays mainly because of their larger dynamic range of transcript abundance estimation. We have performed a whole transcriptome analysis (WTA) of the model white rot basidiomycete *Pleurotus ostreatus* using the Applied Biosystems Solid platform. The genome of P. ostreatus has been recently sequenced by the DOE Joint Genome Institute. The genomes of the two nuclei present in the dikaryotic strain N001 have been sequenced and assembled independently making *P. ostreatus* the first organism for which the two haplotypes have been effectively sequenced in a given individual. We have used these genome sequences as template for the annotation of the WTA data produced by monokaryons derived from each of the two N001 nuclei and for the N001 dikaryon itself, cultured under different conditions. This analysis has revealed the differences in the transcriptome landscape between monokaryons and dikaryons challenged by common environmental (culture) conditions. Besides that, the analysis also reveals the differential expression of genes involved in the degradation of lignocellulose by *P. ostreatus* and permits to compare the strategies used by this and other white and rot fungi whose genome is available.

### Lipidomics of the Unicellular Green Alga Dunaliella salina

**J.E.W. Polle**<sup>1</sup>\* (jpolle@brooklyn.cuny.edu), F. Joseph,<sup>1</sup> C. Fisher-Ramos,<sup>1</sup> V. Samburova,<sup>2</sup> B. Zielinska,<sup>2</sup> M.S. Lemos,<sup>3</sup> S.R. Hiibel,<sup>3</sup> and J.C. Cushman<sup>3</sup>

<sup>1</sup>Brooklyn College, The City University of New York, Brooklyn, New York; <sup>2</sup>Desert Research Institute, Reno, Nevada; and <sup>3</sup>Department of Biochemistry and Molecular Biology, University of Nevada, Reno, Nevada

The alga Dunaliella salina belonging to the Chlorophyta is a model organism for research on hypersaline stress responses and adaptation as it thrives in hyper-saline brines and accumulates equimolar concentrations of osmoprotectants, including glycerol, which can comprise 30% of cellular biomass. In addition to studying cellular responses to acclimation to varying saline concentrations, cells of *D. salina* accumulate high beta-carotene concentrations, up to 10% of the biomass, in lipid globules in the chloroplast when cells are exposed to environmental stress such as nutrient limitation and high irradiances. Currently, the genome of D. salina is in the sequencing pipeline at JGI. In the context of the expected genomic and transcriptomic data originating from the genome sequencing project, we now focus on the lipidome of *D. salina* using UHPLC-TOF-MS. In addition, we investigate other Dunaliella species. A total of 18 different strains of Dunaliella were evaluated for both total lipid and free fatty acid content using two different methods, LC-MS-MS and GC/MS, respectively, with algae grown under nutrient rich conditions. The two methods were found to be in excellent agreement with one another with both reporting an average fatty acid content of 1.0% of algal dry weight with linolenic acid (C18:3). linoleic acid (C18:2), oleic acid (C18:1), and palmitic acid (C16:0) being the major species listed in decreasing relative abundance. mRNA expression and lipid profiling studies are underway to evaluate the effect of nutrient deprivation on lipid content. Because the LC-MS-MS provides both structural characterization of lipids and their quantitative analysis, this method will be used to analyze algal species subjected to nutrient deprivation stress to evaluate their use as potential biofuel feedstocks.

### Metatranscriptomic Inventory of Rhizosphere Soils in an Arid Grassland Under Global Environmental Change Scenarios

**Amy Jo Powell**<sup>1,9</sup>\* (ajpowel@sandia.gov), Don Natvig,<sup>2</sup>\* Andrea Porras-Alfaro,<sup>2,6</sup> Joanna Redfern,<sup>2</sup>\* Miriam Hutchinson,<sup>2</sup> Kylea Odenbach,<sup>1</sup>\* Susannah Tringe,<sup>3</sup> Edward Kirton,<sup>3</sup> Eric Ackerman,<sup>1</sup> Blake Simmons,<sup>1,4</sup> Scott Collins,<sup>2</sup> Robert Sinsabaugh,<sup>2</sup> Diego A. Martinez,<sup>4</sup> Chris Detter,<sup>3,5</sup> Ralph A. Dean,<sup>6</sup> Jon Magnuson<sup>7</sup>, and Randy Berka<sup>8</sup>

<sup>1</sup>Sandia National Laboratories, Albuquerque, New Mexico; <sup>2</sup>University of New Mexico/Sevilleta Long Term Ecological Research Program, Albuquerque, New Mexico; <sup>3</sup>DOE Joint Genome Institute, Walnut Creek, California; <sup>4</sup>The Joint BioEnergy Institute, Emeryville, California; <sup>5</sup>Los Alamos National Laboratory, Los Alamos, New Mexico; <sup>6</sup>Western Illinois University, Macomb, Illinois; <sup>7</sup>North Carolina State University/Center for Integrated Fungal Research, Raleigh, North Carolina; <sup>8</sup>The Broad Institute, Cambridge, Massachusetts; <sup>9</sup>Pacific Northwest National Laboratory, Richland, Washington; and <sup>10</sup>Novozymes, Bagsvaerd, Denmark

We report the initial results of a pilot metagenomic inventory of rhizosphere soils in a native Chihuahuan Desert grassland dominated by the long-lived perennial C4 grass, blue-grama (*Bouteloua gracilis*). The study site is located at the Sevilleta National Wildlife Refuge, central New Mexico, and is part of the Sevilleta Long-Term Ecological Research program. The goal of these experiments is to assess microbial community composition, structure, richness and metabolic potential using high throughput sequencing of total RNA

from rhizopshere soil. In this pilot effort, we compared two samples taken from plots that are part of a long-term, multi-factor, global change experiment. Samples were collected from the rhizosphere of blue grama plants subjected to three treatments: fire, increased nighttime temperatures, and nitrogen addition. Control samples were taken from adjacent unburned and untreated sites. Our results reveal both substantial overlap and compelling differences in the two samples, in terms of community composition and structure. At the kingdom level, sequences from both samples are dominated by eubacteria, approaching 90% of the total. Among the eukaryotes, fungi were the most highly represented ( $\sim$ 58%), followed by sequences from the Viridiplantae ( $\sim 13\%$ ) and Metazoans ( $\sim 8\%$ ); approximately 20% of the reads received no taxonomic rank beyond eukaryote. The taxonomic distributions of fungal BLAST hits were similar to our recent PCR-based analyses of rhizosphere fungal communities at nearby sites. These analyses employed fungal-specific primers to the rDNA region, and resulted in sequences dominated by Ascomycota (>80%). In striking contrast, the current study suggests that in the control/no treatment sample, the Basidiomycota accounted for approximately 60% of fungal sequences, while Ascomycota sequences represented  $\sim 30\%$  of the fungal total, with  $\sim 10\%$ belonging to other major fungal lineages. The experimental sample had approximately equal proportions of Ascomycota and Basidiomycota (41 and 42%, respectively), suggesting the treatments affect fungal composition. Further, these findings suggest that state-of-the-art PCR-based methods for characterizing fungal communities are biased against the Basidiomycota and other lineages. While the present metatranscriptomic analysis reveals significantly greater richness within the Ascomycota than our previous PCR-based studies (*i.e.*, more distinct taxa), these results nonetheless showed consistencies across the different methods; both approaches revealed that members of the Pleosporales are dominant. Our total RNA sequencing results also revealed dramatic prokaryote community composition and structure differences between the two samples. Most notably, sequences from the cyanobacteria are underrepresented in the experimental sample. Our preliminary results indicate that genomics-based approaches are both credible and underrepresented tools for studying the effects of global environmental change on microbial communities in rhizosphere soils. Further *in silico* analyses of the two samples discussed here and sequencing of sample replicates will be performed to validate the observed patterns.

### Bypassing Signal Activation in the System-wide Mapping of Genes Regulated by Response Regulators

Lara Rajeev\* (lrajeev@lbl.gov), Eric G Luning, Paramvir S Dehal, Morgan N Price, Adam P Arkin, and Aindrila Mukhopadhyay

Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California

Two component regulatory systems, comprised of sensor histidine kinases and response regulators, are central to the regulation of stress responses in bacteria. Environmental bacteria especially encode large numbers of putative two component systems and the genes regulated by these systems represent the regulatory networks that impact important natural phenomena such as metal, sulfur, nitrogen and carbon cycling. However, due to lack of knowledge regarding the environmental cues that activate signal transduction, and paucity of methods for high throughput genetic manipulation, these valuable networks remain largely unmapped in most bacteria. Here we used an *in vitro* array-based DAP-chip (<u>DNA Affinity P</u>urified-chip) method to systematically map the genes regulated by all DNA binding response regulators in the model sulfate reducing bacterium, *Desulfovibrio vulgaris* Hildenborough. Our results from the DAP-chip measurements show at least 200

genes, representing approximately 84 operons, to be regulated by 24 response regulators in *D. vulgaris* Hildenborough, of which only one has characterized orthologs. Our results have allowed us to identify the response regulators involved in the regulation of flagella and pili assembly, lactate utilization, exopolysaccharide synthesis, lipid biosynthesis, and in the responses to low potassium, phosphate starvation and nitrite stresses among others. Gene sets regulated by multiple response regulators forming regulatory networks were also discovered. Finally, using the identified gene sets and orthologs in closely related bacteria, we predicted and experimentally verified binding motifs for 15 of these response regulators. These functional predictions may be applied to related species as well, since the binding site motifs appear conserved for several response regulators.

# CheA-3 is Essential for Chemotaxis Towards Electron Acceptors in *Desulfovibrio vulgaris* Hildenborough

**Jayashree Ray**<sup>1</sup>\* (jray@lbl.gov), Kimberly L. Keller,<sup>2</sup> Bernhard Knierim,<sup>3</sup> Manfred Auer,<sup>3</sup> Judy D. Wall,<sup>2</sup> and Aindrila Mukhopadhyay<sup>1</sup>

<sup>1</sup>Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; <sup>2</sup>Biochemistry Department, University of Missouri, Columbia, Missouri; and <sup>3</sup>Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California

Chemotaxis is essential for survival of anaerobic bacterium *Desulfovibrio vulgaris* Hildenborough in the environments with limiting growth nutrients and stressful conditions. Multiple sets of chemotaxis genes including three *cheA* homologs were identified in the genome sequence of the D. vulgaris. Each CheA is a histidine kinase (HK) and part of a two component signal transduction system. Using lactate as electron donor and carbon source and sulfate as the electron acceptor, D. vulgaris shows significant swarming on soft agar plates, the diameter of which appears to be dependent on the sulfate concentration. In order to investigate the role of the putative chemotaxis clusters in *D. vulgaris* growth, motility and swarm phenotype, we generated gene disruption mutants in the cheA loci in all three clusters. Under the optimal growth lactate sulfate conditions, the  $\Delta cheA3$  mutant displayed a complete loss of swarming while deletions in the other two *cheA* loci did not impact this phenotype. Complementing  $\Delta cheA3$  using a plasmid borne copy of cheA3 restores the phenotype of swarming on lactate sulfate as well as motility towards sulfate. Electron microscopy revealed that  $\Delta cheA3$  contains flagella and is not impacted in either growth in liquid medium nor motility as observed on wet mounts. Examination of  $\Delta pilA$ and  $\Delta fliA$  reveals that only  $\Delta fliA$  is defective in swarming suggesting that, though flagellated, the  $\Delta cheA3$  strain is unable to engage its flagella for swarming. Taken together, the results in our study indicate that wild type D. vulgaris requires motility towards electron acceptors in its environment and the CheA3 kinase in D. vulgaris is responsible for modulating this motility. CheA3 shows significant similarity to the Shewanella oneidensis CheA3 and the Vibrio cholerae CheA2 that are responsible for chemotaxis in the respective organisms.

#### Expansion of the Genomic Encyclopedia of Bacteria and Archaea

**Christian Rinke**<sup>1</sup>\* (CRinke@lbl.gov), Alex Sczyrba,<sup>1</sup> Stephanie Malfatti,<sup>1</sup> Janey Lee,<sup>1</sup> Jan-Fang Cheng,<sup>1</sup> Ramunas Stepanauskas,<sup>2</sup> Jonathan A. Eisen,<sup>1,3</sup> Steven Hallam,<sup>4</sup> William P. Inskeep,<sup>5</sup> Brian P. Hedlund,<sup>6</sup> Stefan M. Sievert,<sup>7</sup> Wen-Tso Liu,<sup>8</sup> George Tsiamis,<sup>9</sup> Philip Hugenholtz,<sup>10</sup> and Tanja Woyke<sup>1</sup>

<sup>1</sup>DOE Joint Genome Institute, Walnut Creek, California; <sup>2</sup>Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, Maine; <sup>3</sup>Department of Evolution and Ecology, University of California, Davis, California; <sup>4</sup>Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada; <sup>5</sup>Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, Montana; <sup>6</sup>School of Life Sciences, University of Nevada, Las Vegas, Nevada; <sup>7</sup>Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts; <sup>8</sup>Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois; <sup>9</sup>Department of Environmental and Natural Resources Management, University of Ioannina, Agrinio, Greece; and <sup>10</sup>Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, Australia

To date the vast majority of bacterial and archaeal genomes sequenced are of rather limited phylogenetic diversity as they were chosen based on their physiology and/ or medical importance. The Genomic Encyclopedia of Bacteria and Archaea (GEBA) project (Wu et al. 2009) is aimed to systematically filling the gaps of the tree of life with phylogenetically diverse reference genomes. However more than 99% of microorganisms elude current culturing attempts, severely limiting the ability to recover complete or even partial genomes of these largely mysterious species. These limitations gave rise to the GEBA uncultured project. Here we propose to use single cell genomics to massively expand the Genomic Encyclopedia of Bacteria and Archaea by targeting 80 single cell representatives of uncultured candidate phyla which have no or very few cultured representatives. Generating these reference genomes of uncultured microbes will dramatically increase the discovery rate of novel protein families and biological functions, shed light on the numerous underrepresented phyla that likely play important roles in the environment, and will assist in improving the reconstruction of the evolutionary history of *Bacteria* and Archaea. Moreover, these data will improve our ability to interpret metagenomics sequence data from diverse environments, which will be of tremendous value for microbial ecology and evolutionary studies to come.

#### The Yellowstone Metagenomic Analysis Workshop

Frank F. Roberto<sup>1</sup>\* (Francisco.Roberto@inl.gov) and William P. Inskeep<sup>2</sup>

<sup>1</sup>Idaho National Laboratory, Idaho Falls, Idaho, and <sup>2</sup>Thermal Biology Institute, Montana State University, Bozeman, Montana

The Idaho National Laboratory (INL) and the Thermal Biology Institute (TBI), in conjunction with the Yellowstone National Park Research Coordination Network (RCN-YNP) at Montana State University hosted a metagenomic analysis workshop Jan. 14-16, 2011 in Jackson, Wyoming. Over 25 invitees from participating universities, national labs, private industry, JGI and the J. Craig Venter Institute (JCVI)convened to discuss recent progress in analyzing data generated by the Yellowstone Metagenome Project (YMP) at JGI. Four speakers described the status of analyses examining phototrophic microbial mat communities, metatranscriptomic studies comparing 454 pyrosequencing and SOLiD sequencing at Mushroom Spring, diversity of Aquificales lineages across different habitats in Yellowstone, and distribution and function of thermophilic archaea in a range of hot springs. These studies employed techniques such as k-means clustering, AMPHORA, fragment recruitment, and principal components analysis to compare metagenomic datasets resulting from the YMP. Breakout sessions during the workshop provided participants with overviews of new and improved tools for metagenomic analysis, including the UCSC Archaeal Genome Browser, the Advanced Recruitment Viewer and Multi-Dimensional Scatterplot Viewer from JCVI, and features of the IMG/M Compare Genes and ClaMS utilities at JGI. A report summarizing the workshop will be published later this year.

# Towards Automating High-throughput Combinatorial DNA Assembly

**Rafael D. Rosengarten**<sup>1</sup>\* (rdrosengarten@lbl.gov), Peter McInerney,<sup>2</sup> Huu M. Tran,<sup>2</sup> Jay D. Keasling,<sup>1</sup> and Nathan J. Hillson<sup>1</sup>

<sup>1</sup>Fuels Synthesis and <sup>2</sup>Technology Divisions, Joint BioEnergy Institute, Emeryville, California

A principal goal of microbial metabolic engineering is to create host strains that produce useful biologically-derived compounds, such as substitutes for petrochemicals and liquid fuels. Novel strains are typically generated by modifying genes and/or gene expression through the introduction of DNA devices, i.e. plasmids, genomic integration/deletion cassettes, etc. Just one notable example is the overexpression of the mevalonate pathway in combination with heterologous expression of the plant cytochrome P450 monooxygenase in E. coli and in S. cerevisiae to drive in vivo terpene synthesis. In many cases, there is no way to know a priori what combination of genes and regulatory elements will yield optimal titers, and the labor and time required to generate all possible combinations of components in these engineered pathways can become prohibitive. In the mevalonate pathway example, there are at least eight genes, as well as several different promoters, ribosomal binding sites, and terminators. There are over 1,700 possible ways to combine these particular parts. We are motivated to develop processes to automate the high-throughput construction of genetic devices while remaining flexible to accommodate specific users' experimental requirements. Our approach is to integrate biological computer aided design (BioCAD) tools, scarless combinatorial DNA assembly methods, and liquid handling robotics and microfluidics technologies. We have validated at the lab bench the design capabilities of the software Device Editor and 15, and used these tools to instruct robotic generation, purification, and assembly of DNA parts. Test-beds include a subset of 48 pathway variants, 160 upregulator devices, and an over-expression library of over 30,000 gene combinations. One challenge inherent in the construction of both large pathways and libraries is the assessment of pathway completeness and library diversity. Next generation sequencing will play a critical role in deconvoluting cloned DNA species, thus allowing determination of the success of our processes. These ongoing efforts aim to expand the scale—from tens to hundreds to thousands of constructs—and quicken the pace of DNA assembly, while maintaining maximal flexibility for users.

# *Pleurotus ostreatus* Heme Peroxidase Inventory: From the Genome Sequence to the Enzyme Molecular Structure

**Francisco Javier Ruiz-Dueñas** (fjruiz@cib.csic.es), Elena Fernández, María Jesús Martínez, Ángel T. Martínez\* (atmartinez@cib.csic.es)

Centro de Investigaciones Biológicas, CSIC, Madrid, Spain

The white-rot fungus *Pleurotus ostreatus* has been described as a ligninolytic organism able to degrade lignin selectively. The limited attack to cellulose makes this and other fungi from the genus *Pleurotus* very interesting in different biotechnological applications related to the use of plant biomass, including the integrated lignocellulose biorefineries for the future production of chemicals, materials and biofuels. This fact, together with the understanding of the regulation of the ligninolytic system constituted by heme peroxidases

and other enzymes, as well as the increasing interest in *Pleurotus* as an edible mushroom (with a world production near that of *Agaricus bisporus*) were the main reasons to include this fungus among the organisms of interest to be sequenced by the US Department of Energy. The Joint Genome Institute (JGI, DOE, USA) has sequenced the whole genome of *Pleurotus ostreatus* (dikaryotic strain N001) in a project coordinated by A.G. Pisabarro from the Public University of Navarre (Spain). The strategy followed consisted of sequencing the DNA of two monokaryons (PC15 and PC9) which had been previously obtained from the dikaryotic strain N001. Then, the sequences of both monokaryons were released to different collaborators after assembly by the Stanford Human Genome Center and subsequent annotation by JGI. The 35.6 and 34.3 Mbp assemblies of PC9 (v1.0) and PC15 (v2.0) contain 572 and 12 nuclear scaffolds, and are predicted to have approximately 12206 and 12330 gene models, respectively.

An exhaustive screening of the genome sequence of PC9 and PC15 was performed to search for nucleotide sequences of heme peroxidases in this white-rot fungus. After sequence identification and manual curation of the corresponding genes and cDNAs, the deduced amino acid sequences were converted into structural homology models using crystal structures of reference proteins, deposited in the RCSB Protein Data Bank, as templates and the programs implemented by the automated protein homology-modeling server "Swiss-Model". A comparative study of these sequences and their structural models with those of known fungal peroxidases revealed the complete inventory of heme peroxidases of this fungus. This consists of: i) cytochrome c peroxidase (1 model) and ligninolytic peroxidases (9 models), including five manganese peroxidases and four versatile peroxidases but not lignin peroxidase (in agreement with previous expression/production studies and DNA hybridization analysis), as representative of the "classical" superfamily of plant, fungal, and bacterial peroxidases; and ii) members of two relatively "new" peroxidase superfamilies, namely heme-thiolate peroxidases (3 models), here described for the first time in a fungus from the genus *Pleurotus* and initially represented by the *Leptoxyphium fumago* chloroperoxidase, related enzymes and aromatic peroxygenases recently described in several agaric basidiomycetes with very interesting catalytic properties (e.g. aromatic oxygenation), and dye-decolorizing peroxidases (4 models), already known in *P. ostreatus* but still to be thoroughly explored and characterized. At present, these heme peroxidases are being expressed in *Escherichia coli* and their stability and catalytic properties studied with the aim of determining their substrate specificity and real biotechnological potential.

#### **RNA-SEQ** and Homolog Searching in *Pleurotus ostreatus*

**Francisco Santoyo**\* (Francisco.santoyo@unavarra.es), Antonio G. Pisabarro, Gúmer Pérez, and Lucía Ramírez

Genetics and Microbiology Research Group, Public University of Navarre, Pamplona, Spain.

*Pleourotus ostreatus* is an edible basidiomycete of great interest in the field of food and bioremediation. In 2009 genome sequences of the two haplotypes of the strain N001 of Pleurotus ostreatus var florida were released. Sequences of both haplotypes could be studied separately permitting the analysis of two gene complements of a sexually mature basidiomycete simultaneously for the first time.

Considering that P.ostreatus is a white rot fungus with an important battery of lignine degrading enzymes, we performed RNA-SEQ experiments in the haplotypes and the dicaryotic strains grown in liquid and solid media at three temperatures to learn about the abilities of this fungus to secrete lignine degradation enzymes under different

environmental conditions. For this purpose, we carried out a search of alleles corresponding to each haplotype using the Blast software. In each haplotype two sequences were considered alleles if their e-value were less then e-20 and their identity percentages were greater than 60%.

The RNA-SEQ experiments were carried out using the SOLID platform from Applied Biosystems,. Data were analyzed with the software TopHat. The results obtained showed important differences in the transcription level of both protoclones in some genes involved in lignin degradation pathways under different environmental conditions. In this way, we estimated that the genic action in trancript level in the LACC6 gene is sobredominance, whereas in the LACC5 gene is intermediate dominance.

Finally, some of the results obtained from interesting genes were tested with quantitative PCR.

#### Microbial Community Characterization of Leaf-Cutter Ant Refuse Dumps through 16S rRNA Full-length and Pyrosequencing Analysis

**Jarrod J. Scott**<sup>1,2,3</sup>\* (scott2@wisc.edu), Garret Suen,<sup>1,2</sup> Frank O. Aylawrd,<sup>1,2</sup> Joseph A. Moeller,<sup>1,2</sup> Susannah G. Tringe,<sup>4</sup> Kerrie W. Barry,<sup>4</sup> and Cameron R. Currie<sup>1,2,3</sup>

<sup>1</sup>DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, Wisconsin; <sup>2</sup>Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin; <sup>3</sup>Smithsonian Tropical Research Institute, Balboa, Ancon, Panama; and <sup>4</sup>DOE Joint Genome Institute, Walnut Creek, California

Leaf-cutter ants use fresh plant material to grow a mutualistic fungus that serves as the ants' primary food source. Within fungus gardens, various plant compounds are metabolized and transformed into nutrients suitable for ant consumption. This symbiotic association produces a large amount of refuse consisting primarily of partly degraded plant material. A leaf-cutter ant colony is thus divided into two spatially and chemically distinct environments that together represent a plant biomass degradation gradient. Little is known about the microbial community structure within a single refuse dump or between dumps of different colonies.

Using full-length and pyrotag 16S rRNA analysis, coupled with a variety of community metrics, we assessed and compared the microbiota of refuse dumps from three mature colonies of the leaf-cutter ant *Atta colombica*. We sampled five strata per each dump. In total, we generated 15,917 16S rRNA clones from the full-length clone libraries across all 15 samples. Removal of short and low-quality sequences yielded 14 627 high-quality sequences (~1360 bp average length). Library sizes ranged from 549 to 2618 clones ( $\mu = 975 \pm 138$  s.e. per sample). In addition, we generated a total 507,877-pyrotag sequences across all 15 samples. Quality screening resulted in a total of 361,442 clones (~290 bp average length). Library sizes ranged from 476 to 60,842 clones ( $\mu = 24,096 \pm 3678$  S.E. per sample).

The full-length dataset revealed the presence of 23 Bacterial and 2 Archaeal phyla in addition to sequences with no identifiable classification. However, 13 of the identified phyla were each present in less than 0.1% of the total clones. The remaining 12 phyla accounted for 97.7% of total clone content. Roughly, 80% of total diversity was present in three phyla, namely Proteobacteria (40.5% of all sequences; mean (±S.E.) abundance in 15 pooled was  $41.2 \pm 3.4\%$ , range = 4.0 - 58.7%), Actinobacteria (24.3%;  $\mu = 25.5 \pm 5.4\%$ ,

range = 6.2 - 91.8%), and Bacteroidetes (14.8%;  $\mu = 16.1 \pm 2.6\%$ , range = 2.6 - 34.5%). Similarly, the pyrotag dataset revealed the presence of 24 Bacterial and 1 Archaeal phyla in addition to sequences with no identifiable classification. However, 17 of the identified phyla were each present in less than 1% of the total clones. The remaining eight phyla accounted for 94.0% of total clone content. Roughly, 80% of total diversity was present in three phyla, namely Proteobacteria (52.6% of all sequences; mean (±S.E.) abundance in 15 was 50.8 ± 3.4%, range = 32.3-68.8%), Actinobacteria (13.9%;  $\mu = 16.1 \pm 2.6\%$ , range = 1.1 - 42.6%), and Bacteroidetes (14.2%;  $\mu = 16.1 \pm 3.5\%$ , range = 0.4 - 41.6%). Our results also show that communities change significantly through the refuse dumps and that communities are more similar based on depth rather than host nest. Finally, our comparison of these communities using full-length and pyrotag sequencing revealed consistent patterns of community composition and structure.

#### Metagenomic Analysis of the Asian Longhorned Beetle Gut Microbial Consortium Reveals Potential Contribution to Lignocellulose Digestion

**Erin Scully**<sup>1</sup>\* (eds14@psu.edu), Scott Geib,<sup>2</sup> John Carlson,<sup>3</sup> Ming Tien,<sup>4</sup> and Kelli Hoover<sup>5</sup>

<sup>1</sup>Interdisciplinary Graduate Degree Program in Genetics, Pennsylvania State University, University Park, Pennsylvania; <sup>2</sup>USDA-ARS U.S. Pacific Basin Agricultural Research Center, Hilo, Hawaii; <sup>3</sup>Department of Forestry, Huck Institute of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania; <sup>4</sup>Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania; and <sup>5</sup>Department of Entomology, Pennsylvania State University, University Park, Pennsylvania;

The Asian longhorned beetle (ALB; *Anoplophoraglabripennis*) is a destructive, woodboring pest that is capable of subsisting in 24 different healthy, deciduous tree species. ALB's ability to thrive in healthy host trees is particularly impressive considering that most other wood-boring insects feed on stressed or dying trees whose woody components have been pre-digested by wood-degrading fungi. In contrast, ALB feeds and grows in a harsh environment devoid of nutritional resources by using intractable components, including lignin and cellulose, for energy. Our lab recently demonstrated that the lignin macromolecule is rapidly degradedduring passage through the ALB gut, indicating that the gut microbiome may serve as a novel source of efficient lignin-degrading enzymes that could be exploited for industrial biofuels production. In concert, the gut harbors a diverse microbial community hypothesized to provide key enzymes for efficient lignocellulose digestion and nutrient provisioning.

DOE has recognized the potential of the gut microbiota to provide significant contributions to biofuels production. Through our collaboration with the Joint Genome Institute, we sequenced 1.25 million reads from the ALB gut metagenome using 454 Titanium chemistry. Although a small portion of our reads assembled to generate 26,000 contigs ranging in size from 100 to 31,000 bp, the majority of reads did not assemble, leaving over 600,000 singletons. Due to the complexity of the community and the danger of generating chimeric contigs during assembly, we are focusing the majority of our annotation efforts on individual reads. Preliminary annotation of predicted protein coding regions using MG-RAST indicated that a majority of reads were associated with carbohydrate metabolism, including a strong representation of reads predicted to encode cellulases, xylanases, and pectinases. Althougha subset of the reads were also associated with degradation of phenylpropanoid containing compounds, including reads classified as lignostilbene-alpha, beta-dioxygenases, no reads associated with degradation of polymeric lignin were

detected. However, annotation using a more comprehensive database may allow us to detect reads that could be involved rapid oxidative depolymerization of lignin. Also, given that ALB also harbors a unique soft rot fungal isolate belonging to the *Fusarium solani* species complex, it is possible that lignin depolymerization in the gut involves a novel mechanism that will require further analysis of the genome of this fungal isolate.

### Efficient Graph Based Assembly of Short-Read Sequences on a Hybrid Core Architecture

Alex Sczyrba<sup>\*1,2</sup> (asczyrba@lbl.gov), Abhishek Pratap,<sup>1,2</sup> Shane Canon,<sup>2,3</sup> James Han,<sup>1,4</sup> Alex Copeland,<sup>1,2</sup> Zhong Wang,<sup>1,2</sup> Tony Brewer,<sup>5</sup> David Soper,<sup>5</sup> Mike D'Jamoos,<sup>5</sup> Kirby Collins,<sup>5</sup> and George Vacek<sup>5</sup>

<sup>1</sup>DOE Joint Genome Institute, Walnut Creek, California; <sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, California; <sup>3</sup>National Energy Research Scientific Computing Center (NERSC), Oakland, California; <sup>4</sup>Lawrence Livermore National Laboratory, Livermore, California; and <sup>5</sup>Convey Computer Corp., Richardson, Texas

Advanced architectures can deliver dramatically increased throughput for genomics and proteomics applications, reducing time-to-completion in some cases from days to minutes. One such architecture, hybrid-core computing, marries a traditional x86 environment with a reconfigurable coprocessor, based on field programmable gate array (FPGA) technology. In addition to higher throughput, increased performance can fundamentally improve research quality by allowing more accurate, previously impractical approaches.

We will discuss the approach used by Convey's de Bruijn graph constructor for short-read, *de-novo* assembly. Bioinformatics applications that have random access patterns to large memory spaces, such as graph-based algorithms, experience memory performance limitations on cache-based x86 servers. Convey's highly parallel memory subsystem allows application-specific logic to simultaneously access 8192 individual words in memory, significantly increasing effective memory bandwidth over cache-based memory systems. Many algorithms, such as Velvet and other de Bruijn graph based, short-read, *denovo* assemblers, can greatly benefit from this type of memory architecture. Furthermore, small data type operations (four nucleotides can be represented in two bits) make more efficient use of logic gates than the data types dictated by conventional programming models.

JGI is comparing the performance of Convey's graph constructor and Velvet on both synthetic and real data. We will present preliminary results on memory usage and run time metrics for various data sets with different sizes, from small microbial and fungal genomes to very large cow rumen metagenome. For genomes with references we will also present assembly quality comparisons between the two assemblers.

#### Transcriptional Response of *Rhodopseudomonas palustris* Strainspecific Gene Encoding Stress Response and Solute Transport

**Shaneka S. Simmons**<sup>1</sup>\* (angelique.c.lee@jsums.edu), Leslie M. Perry,<sup>2</sup> Sarah Martinez,<sup>2</sup> Hari H.P. Cohly,<sup>1</sup> Michael S. Allen,<sup>2</sup> and Raphael D. Isokpehi<sup>1</sup>

<sup>1</sup>Center for Bioinformatics & Computational Biology, Department of Biology, Jackson State University, Jackson, Mississippi, and <sup>2</sup>Department of Biological Sciences, University of North Texas, Denton, Texas

*Rhodopseudomas palustris* are metabolically versatile species that can convert atmospheric carbon dioxide into biomass, recycle aromatic polymers of ligin, produce hydrogen gas for energy production, and fix atmospheric nitrogen, R. palustris strains are also able to degrade a wide range of toxic organic compounds, and may be of use in bioremediation of polluted sites. The availability of genome sequences of 7 R. palustris strains provides rich genomic data to determine the genomic basis for diverse phenotypes of the bacteria. We are particularly interested in the function of the universal stress protein (Pfam00582) superfamily, which encompasses a conserved group of proteins that provide cells with the ability to respond to environmental stresses. Homologs of genes encoding the USP domain have been identified in archeae, bacteria and eukaryotes. Usp-containing organisms encode a small, single Usp protein (~14-15 kDA), a larger version (~30 kDa) consisting of two Usp domains in tandem, or as one or two Usp domains together with other functional domains. Functional domains commonly fused to Usp domains include antiporter, voltage channels, amino acid permeases, and protein kinase domains. A total of 61 architectures in 872 species, representing 9002 sequences are documented for universal stress protein in Pfam. Rhodopseudomonas palustris genes encoding the Universal Stress Protein domain were selected and prioritized according to amino acid length and presence of unique functional annotation. Here we identify a novel Universal Stress Protein (USP, PF00582) Domain, in the C terminus of Rhodopseudomonas palustris MSF-1 protein (535aa) that shares similarity with only 5 other bacterial proteins. Major Facilitator Superfamily (MSF-1, PF07690) is one of the two largest, ubiquitous families of membrane transporters found on Earth present in bacteria, archaea, and eukarya. Members of MSF-1 function as single polypeptide secondary carriers capable of solute uniport, solute/cation symport, solute/cation antiport and/or solute/solute antiport with inwardly and/or outwardly directed polarity. Transporters are involved in antibiotic efflux, nutrient capture, environmental sensing, protein secretion, toxin production, photosynthesis, and oxidative phosphorylation with transport of diverse substrates including ions, sugars, sugar phosphates, drugs, neurotransmitters, nucleosides, amino acids and peptides in response to the chemiosmotic ion gradient. Standard PCR was conducted to confirm that the gene with domains for MSF-1 and USP was uniquely present in *R. palustris* BisB18. Our investigation confirmed bioinformatics prediction. Furthermore, Quantitative Reverse-Transcriptase Polymerase Chain Reaction (QRT-PCR) was conducted to measure transcriptional response of the USP gene to NaCl stress. Upon salt-induced osmotic stress, a four-fold increase in cDNA expression of the USP gene was observed. Further research are planned to (i) determine response to other types of environmental stress conditions; (ii) elucidate transport properties of the encoded protein; (iii) interactions of the two protein domains.

### Genome Sequence-enabled Identification of Avirulence Genes in *Cronartium quercuum* f.sp. *fusiforme*

**Katherine E. Smith**<sup>1,2</sup>\* (smithk@ufl.edu), Claire Anderson,<sup>1</sup> Jason A. Smith,<sup>1</sup> C. Dana Nelson,<sup>2</sup> and John M. Davis<sup>1</sup>

<sup>1</sup>School of Forest Resources and Conservation, University of Florida, Gainesville, Florida, and <sup>2</sup>Southern Institute of Forest Genetics, U.S. Forest Service, Saucier, Mississippi

The *Cronartium quercuum* f.sp. *fusiforme* (*Cqf*) whole genome sequencing project (currently in the assembly stage at the Joint Genome Institute, JGI) will enable identification of avirulence genes in the most devastating pine fungal pathogen in the southeastern USA. Amerson and colleagues (manuscript in prep) have mapped 9 resistance genes in loblolly pine suggesting that at least 9 corresponding avirulence genes should exist in the fungus. Identification of these avirulence genes would greatly facilitate

resistant pine genotype selection for deployment to forest plantations. For example, the *Cqf* avirulence gene, *Avr1*, specifically interacts with the *Fr1* resistance gene. Six Random Amplified Polymorphic DNA (RAPD) markers and eight Amplified Fragment Length Polymorphism (AFLP) markers have been identified that are significantly linked to *Avr1* and these define its location to an 8.62 cM interval. AFLP fragments were PCR amplified from the parents of a mapping population that segregates for *Avr1* and from bulks of progeny that either contain, or lack, the avirulence locus (or allele). One hundred and fourteen AFLP primer combinations were tested using this bulked segregant analysis approach and 3946 amplified fragments were scored, 198 of which are polymorphic between the parental isolates. AFLP markers identified as linked to *Avr1* have been screened in a mapping population using converted markers and placed on the genetic map. Our next task is to locate these DNA sequences in the assembled genome of Cqf and determine their function in this pathosystem.

### The Complete Genome Sequence of *Fibrobacter succinogenes* S85 Reveals a Cellulolytic and Metabolic Specialist

**Garret Suen**,<sup>1,2</sup> Paul J. Weimer,<sup>3</sup> David M. Stevenson,<sup>3</sup> Frank O. Aylward,<sup>1,2</sup> Julie Boyum,<sup>4</sup> Jan Deneke,<sup>4</sup> Colleen Drinkwater,<sup>4</sup> Natalia N. Ivanova,<sup>5</sup> Natalia Mikhailova,<sup>5</sup> Olga Chertkov,<sup>6</sup> Lynne A. Goodwin,<sup>5,6</sup> Cameron R. Currie,<sup>1,2</sup> David Mead,<sup>1,4,7</sup> and Phillip J. Brumm\*<sup>1,7</sup> (pbrumm@c56technologies.com)

<sup>1</sup>DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, Wisconsin; <sup>2</sup>Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin; <sup>3</sup>U.S. Dairy Forage Research Center, U.S. Department of Agriculture–Agricultural Research Services (USDA–ARS), Madison, Wisconsin; <sup>4</sup>Lucigen Corp., Middleton, Wisconsin; <sup>5</sup>DOE Joint Genome Institute, Walnut Creek, California; <sup>6</sup>Biosciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico; and <sup>7</sup>C5-6 Technologies, Middleton, Wisconsin

Fibrobacter succinogenes is an important member of the rumen microbial community that converts plant biomass into nutrients usable by its host. This bacterium, which is also one of only two cultivated species in its phylum, is an efficient and prolific degrader of cellulose. Specifically, it has a particularly high activity against crystalline cellulose that requires close physical contact with this substrate. However, unlike other known cellulolytic microbes, it does not degrade cellulose using a cellulosome or by producing high extracellular titers of cellulase enzymes. To better understand the biology and cellulolytic degrading machinery of F. succinogenes, we sequenced the genome of the type strain S85 to completion. A total of 3,085 open reading frames were predicted from its 3.84 Mbp genome, which consists of a single circular chromosome. Analysis of sequences predicted to encode for carbohydrate-degrading enzymes revealed an unusually high number of genes (114) that were classified into 46 different families of glycoside hydrolases, carbohydrate binding modules (CBMs), carbohydrate esterases, and polysaccharide lyases. Of the 30 identified cellulases, none contain CBMs in families 1, 2, and 3, which are typically associated with crystalline cellulose degradation. Polysaccharide hydrolysis and utilization assays showed that F. succinogenes was able to hydrolyze a number of polysaccharides, but was only able to utilize the hydrolytic products of cellulose. This suggests a model in which F. succinogenes uses its array of hemicellulosedegrading enzymes to remove these polysaccharides in order to gain access to cellulose. This is reflected in its genome, as F. succinogenes lacks many of the genes necessary to transport and metabolize non-cellulose polysaccharides. The F. succinogenes genome reveals a bacterium that specializes on cellulose as its sole energy source, and provides insight into a novel strategy for cellulose degradation.

#### The Microbiomes of Five Fungus-associated Insect Herbivores Reveal Communities that Share Microbial Diversity and Function

**Garret Suen**<sup>1,2</sup>\* (gsuen@wisc.edu), Jarrod J. Scott,<sup>1,2,3</sup> Frank O. Aylward,<sup>1,2</sup> Joseph A. Moeller,<sup>1,2</sup> Sandra M. Adams,<sup>1,2</sup> Aaron S. Adams,<sup>4</sup> Peter H.W. Biedermann,<sup>5</sup> Susannah G. Tringe,<sup>6</sup> Kerrie W. Barry,<sup>6</sup> Stephanie Malfatti,<sup>6</sup> Lynne A. Goodwin,<sup>6</sup> Michael Poulsen,<sup>7</sup> Duur Aanen,<sup>8</sup> Kenneth G. Raffa,<sup>4</sup> and Cameron R. Currie<sup>1,2,3</sup>

<sup>1</sup>DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, Wisconsin; <sup>2</sup>Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin; <sup>3</sup>Smithsonian Tropical Research Institute, Balboa, Ancon, Panama; <sup>4</sup>Department of Entomology, University of Wisconsin-Madison, Madison, Wisconsin; <sup>5</sup>Institute of Ecology and Evolution, Division Behavioral Ecology, University of Bern, Bern, Switzerland; <sup>6</sup>DOE Joint Genome Institute, Walnut Creek, California; <sup>7</sup>Department of Biology, Section for Ecology and Evolution, University of Copenhagen, Copenhagen, Denmark; and <sup>8</sup>Laboratory of Genetics, Wageningen University, Wageningen, Netherlands

Fungus-associated insect herbivores are widespread and dominant herbivores worldwide. This includes bark beetles in North America, leaf-cutter ants in South America, Ambrosia beetles in Europe, Asia, and Australia, and fungus-growing termites in Africa and Asia. Their success is arguably due to their close association with fungi that allows them to overcome host plant defenses. As a result, they are ecosystem engineers that contribute substantially to carbon cycling through their decomposing activities. To gain insight into these insect-fungus systems, we sequenced 16S pyrotag libraries and community metagenomes from samples of the insect-fungus interface, totaling 4.64 Gbp and 1.8 million predicted proteins. Our 16S rRNA analysis revealed that these systems are dominated by bacteria in the Gammaproteobacteria, which accounted for at least 75% of the total 16S sequences in each sample. Phylogenetic binning analysis of the associated community metagenomes also confirmed these findings, and further showed that the majority of the Gammaproteobacteria are those from the family *Enterobacteriaceae*. Because these insect-fungus systems are associated with high plant biomass degrading capacity, we hypothesized that bacteria in these systems may contribute to the overall deconstruction of plant cell wall polysaccharides. A carbohydrate-active enzyme (CAZymes) analysis revealed that these bacteria harbor a large amount of enzymes that hydrolyze hemicelluloses, starch, and pectin. Interestingly, only a handful of enzymes implicated in cellulose degradation were found, suggesting that these insect-fungal systems may use other means to deconstruct this carbohydrate such as the associated fungus. Further comparisons of these community metagenomes with other plant-biomass degrading microbiomes like the cow rumen, termite hindgut, and Tammar wallaby foregut shows that the insect-fungus metagenomes are more similar to each other with respect to bacterial diversity, overall gene content, metabolism, and CAZymes. This may indicate convergent evolution of these communities, likely due to the shared similarity of an insect associating with a fungus. Given recent interest in enzymes efficient at degrading plant cell wall polysaccharides for biofuel production, these data provide insights into how natural and highly evolved herbivores leverage symbiotic microbial communities to achieve largescale and rapid plant biomass deconstruction.

### Amazon Rainforest Microbial Observatory: Deforestation of the Largest CO<sub>2</sub> Sequestration Terrestrial Ecosystem in the World Causes Losses in Microbial Community Spatial Structure

James M. Tiedje<sup>1</sup>\* (tiedjej@msu.edu), Brendan J. Bohannan,<sup>2</sup> Klaus Nüsslein,<sup>3</sup> Vivian H. Pellizari,<sup>4</sup> Kyung-Hwa Baek,<sup>3</sup> Brigitte J. Feigl,<sup>4</sup> Ederson da C. Jesus,<sup>1</sup> Babur Mirza,<sup>5</sup> Rebecca Muller,<sup>2</sup> Fabiana da S. Paula,<sup>4,5</sup> Siu M. Tsai,<sup>4</sup> and **Jorge L.M. Rodrigues**<sup>5</sup>

<sup>1</sup>Center for Microbial Ecology, Michigan State University, East Lansing, Michigan; <sup>2</sup>University of Oregon, Eugene, Oregon; <sup>3</sup>University of Massachusetts, Amherst, Massachusetts; <sup>4</sup>University of Sao Paulo, Sao Paulo, Brazil; and <sup>5</sup>University of Texas, Arlington, Texas

The Amazon forest is the largest tropical forest ecosystem on Earth, influencing hydrological and climatological cycles and balancing the flux of atmospheric gases of global proportions. Being regarded as a hotspot for biodiversity, the Amazon forest is virtually unexplored for its microbiological component. We established the first microbial observatory in the Amazon forest in order to: (1) identify its bacterial taxa, (2) characterize the distance-decay relationship for microbial communities, and (3) quantify the taxonomic alteration after deforestation has taken place.

A spatially-explicit soil sampling strategy covered an area of approximately 100 km<sup>2</sup> in the State of Rondonia, Brazil, at scales from centimeters to 10 kilometers. This biogeographical approach was designed for three different land use types: primary rainforest, pastures established at different ages, and secondary rainforest. Twelve soil core samples were taken at each corner of nested quadrants inside a 100 m<sup>2</sup> area, with the sampling quadrant repeated a 1 km and 10 km distances for each land use. Total soil DNA was analyzed by pyrosequencing with primers targeting the V4 region of the 16S rRNA gene. Sequences were processed through the pyrosequencing pipeline of the Ribosomal Database Project, and grouped into operational taxonomic units based on 97% sequence identity.

Microbial biogeography did respond to land use change in a way similar to plants. Our results indicate that any two forest microbial communities have a statistically significant decrease in similarity with increased distances between them. This distance-decay relationship is disrupted with the conversion of forest to pasture, indicating loss of spatial structure. While this biogeographical pattern remained similar to what was previously observed for plants, microbial diversity responded to land use change in a way different from plants. Land use change did not significantly reduce the number of bacterial taxa or reduce phylogenetic diversity. We observed an increase in phylogenetic diversity with increasing pasture age. The trends of increased phylogenetic diversity following land use change from forest to pasture are also supported by changes in functional gene diversity for both ammonia oxidizing bacteria and archaea, and methane oxidizing bacteria. *Proteobacteria* and *Acidobacteria* were found in higher proportions in primary and secondary forests, while *Firmicutes* and *Actinobacteria* had their highest abundances in pasture samples. Pasture abandonment with conversion to secondary forest restored the microbial community to a composition similar to that observed for primary forest soil.

# **Proteomics of Global Climate Change: The Response of Marine Organisms to Warming and Ocean Acidification**

Lars Tomanek\* (ltomanek@calpoly.edu)

Department of Biological Sciences, Environmental Proteomics Laboratory, California Polytechnic State University, San Luis Obispo, California

Increasing temperatures and ocean acidification challenge marine organisms in unprecedented ways, possibly increasing levels of physiological stress that can affect populations over the long run. We used a comparative proteomics approach to study the effects of a changing climate on the cellular physiology of mussels (blue mussels Mytilus trossulus and M. galloprovincialis) and eastern oysters (Crassostrea virginica) in response to acute and chronic heat stress and hypercapnic conditions mimicking more acidic seawater. Our results indicate that acute heat stress activates the synthesis of a battery of molecular chaperones, the proteasome degradation pathway, and a broad suite of cytoskeletal proteins in gill tissue of both *Mytilus* species. However, acute heat stress limits the abundance of oxidative stress proteins while moving energy metabolism from pro-oxidant NADH-producing to anti-oxidant NADPH-producing pathways in the coldadapted *M. trossulus* but not the warm-adapted *M. galloprovincialis*. Changes in the NAD-dependent deacetylase (sirutin-2), an indicator of lifespan, suggest that the changes in energy metabolism may be caused by the acetylation status of enzymes. Chronic heat stress shows changes opposite from those of acute heat stress in the proteome of both *Mytilus* species: an up-regulation in NADH-producing energy metabolism with increasing temperatures. Chronic hypercapnic conditions that lower the pH of seawater by 0.4 pH units increase the abundance of several oxidative stress proteins in mantle tissue of the eastern oyster. These and a number of other studies suggest that oxidative stress is a universal cellular stress accompanying a range of chemical and physical stressors. Our results show that proteomics, based on 2D gel electrophoresis and tandem mass spectrometry, is an ideal method to detect global changes in protein abundance in order to generate new hypotheses about how climate change affects the cellular physiology of marine organisms.

#### The Microbial Energy Processes Gene Ontology Project

**Trudy Torto-Alalibo**<sup>1</sup>\* (trudy@vbi.vt.edu), Biswarup Mukhopadhyay,<sup>1</sup> T. M. Murali,<sup>2</sup> Brett M. Tyler,<sup>1</sup> and Joao C. Setubal<sup>1</sup>

<sup>1</sup>Virginia Bioinformatics Institute, Blacksburg, Virginia, and <sup>2</sup>Virginia Polytechnic Institute and State University, Blacksburg, Virginia

The MENGO project is a community-oriented multi-institutional collaborative effort that aims to develop new Gene Ontology (GO) terms to describe microbial processes of interest to bioenergy. Such terms will aid in the comprehensive annotation of gene products from diverse energy-related microbial genomes. The Gene Ontology consortium was formed in 1998 to create universal descriptors, which can be used to describe functionally similar gene products and their attributes across all organisms. MENGO, an interest group of the GO consortium, solicits help from the bioenergy community in developing GO terms relevant for energy-related processes. The MENGO interest group will host a workshop right after the JGI Users meeting, on March 25<sup>th</sup> at the Joint Genome Institute in Walnut Creek. This workshop will introduce participants to the Gene Ontology. Additionally, we will have an open forum to seek input from participants on general concepts relevant to bioenergy processes, which will subsequently inform MENGO term development. Funding for the MENGO project is provided by the Department of Energy as part of the Systems Biology Knowledgebase program.

### Assembly of 67 Haloarchaeal Genomes Using Paired Illumina Reads

A. Tritt<sup>1</sup>\* (atritt@ucdavis.edu), D. Larsen,<sup>1</sup> J. Eisen,<sup>1,3,4</sup> A. Darling,<sup>1</sup> and M. Facciotti<sup>1,2</sup>

<sup>1</sup>UC Davis Genome Center, University of California, Davis, California; <sup>2</sup>Department of Biomedical Engineering, University of California, Davis, California; <sup>3</sup>Section of Evolution and Ecology, College of Biological Sciences, University of California, Davis, California; and <sup>4</sup>Department of Medical Microbiology and Immunology, School of Medicine, University of California, Davis, California

Unlike many archaea, the Halobacteriacea, are easily and safely cultured in laboratory. This phylogenetic clade thus offers a good model system for studying the archaea in general and more specifically, life in hypersaline environments.

To date, thirteen haloarchaeal genomes have been published. Here, we present sixty-seven draft genomes from the Halobacteriacea. We sequenced isolates from eighteen total genera, of which ten had no previously sequenced representative. Newly sequenced genera include *Halosarcina, Halosimplex, Natronolimnobius, Halovivax, Natronobacterium, Natrinema, Halococcus, Halobiforma, Natronorubrum,* and *Natronococcus.* We also sequenced more deeply the following genera: *Haloferax, Haloarcula, Halorubrum, Haloterrigena, and Natrialba.* 

All isolates were sequenced on an Illumina GA II and/or an Illumina HiSeq 1000. To this end, we tested multiple assembly strategies, and developed a highly efficient and effective pipeline for assembling Illumina short-read data. Briefly, before assembly, raw reads were corrected for errors using Reptile (3). Contigs were built and scaffolded using IDBA (2), and then extended and scaffolded again using SSPACE (1). Furthermore, through sequencing large-insert mate-pair libraries, we demonstrate complete scaffolding of some genomes using Illumina short-read data alone. Finally, draft genomes were annotated on the RAST server (4) for future analysis of genomic diversity. Other future directions include sequencing of additional large insert libraries to facilitate assembling all genomes down to complete scaffolds.

References:

1) Boetzer M, Henkel CV, Jansen HJ, Butler D, and Pirovano W, *Scaffolding pre*assembled contigs using SSPACE. Bioinformatics, 2010

2) Peng Y, Leung H, Yiu S and Chin F, *IDBA – A Practical Iterative de Bruijn Graph De Novo Assembler*. RECOMB, 2010. 6044: 426-440.

3) Xiao Y, Dorman KS and Aluru S, *Reptile: representative tiling for short read error correction*. Bioinformatics, 2010. 26: 2526-2533.

4) Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, and Zagnitko O, *The RAST Server: Rapid Annotations using* 

#### **Studying Bioenergy-relevant Traits Using Natural and Induced Variation in the Model Grass** *Brachypodium distachyon*

**Ludmila Tyler**<sup>1,2</sup>\* (ltyler@berkeley.edu), Michael A. Steinwand,<sup>1</sup> Cynthia Cass,<sup>3</sup> John Sedbrook,<sup>3</sup> and John P. Vogel<sup>1</sup>

<sup>1</sup>USDA-ARS Western Regional Research Center, Albany, California; <sup>2</sup>Department of Plant and Microbial Biology, University of California, Berkeley, California; and <sup>3</sup>School of Biological Sciences, Illinois State University, Normal, Illinois

Short stature, simple growth requirements, abundant genomic resources, and diverse germplasm make *Brachypodium distachyon* a good model for less-tractable grasses, including the emerging bioenergy crops switchgrass and Miscanthus. Both biomass quantity and quality affect a plant's suitability for use as a feedstock in biofuel production. However, relatively little is known about the genes which control plant architecture and cell wall composition in the grasses. To improve our understanding of these bioenergyrelevant traits and their genetic basis, we surveyed more than 170 natural accessions for variation in growth habit and height. For a subset, we measured stem densities and tested for compositional differences using near-infrared spectroscopy (NIRS). A simultaneous saccharification and fermentation assay utilizing the cell-wall-degrading, ethanolproducing bacterium *Clostridium thermocellum* was also developed. This assay is being used to assess the conversion characteristics of natural accessions, as well as ethylmethane-sulfonate and fast-neutron mutants identified via NIRS screening. Based on the phenotypic results, we have chosen a core set of natural accessions for genome resequencing and have begun mapping several of the mutations. A summary of these characterizations will be presented.

# Gene Expression in Representatives of *Methylophilaceae* from Lake Washington, Assessed via RNAseq Analysis

A. Vorobev\* (voralex@u.washington.edu), D.A.C. Beck, M.G. Kalyuzhnaya, and L. Chistoserdova

University of Washington, Seattle, Washington

Previous collaborations with the JGI have indicated that *Methylophilaceae* must play a significant role in environmental cycling of single carbon compounds in Lake Washington and likely in other freshwater environments. We sequenced and analyzed the genomes of three different strains of methylotrophic bacteria within the family *Methylophilaceae*, all isolated from Lake Washington sediment: Methylotenera mobilis JLW8, Methylotenera versatilis 301 and Methylovorus glucosetrophus SIP3-4. Sequence analysis revealed significant genomic divergence among the three strains, suggesting that specialized lifestyles must be responsible for their environmental fitness. Here, we use an approach for characterizing physiological abilities of closely related methylotrophic bacteria *in situ*, through transcriptomic profiling, in an attempt to reveal functions specifically expressed in the environment that may remain silent in the laboratory. We will present data on gene expression in the three *Methylophilaceae* species in guasi-*in situ* conditions, as recorded by RNAseq analysis, compared to gene expression in cultures grown in defined media. We will demonstrate that, while expressing some of the conserved genes enabling methylotrophy, the three organisms also express sets of species-specific genes and pathways that likely enable functions of importance for these microbes as members of a functional community. Expression patterns of some of these genes change in the *in situ* conditions relative to the patterns observed during growth in defined media, potentially

highlighting involvement of these genes in specific adaptations to the changing environmental conditions and possibly into species-species communication.

#### **CloudBWA: A Cloud Computing Based Aligner**

Mingyi Wang \* (mwang@noble.org), Yinbing Ge, and Ji He<sup>†</sup>

Scientific Computing Department, The Samuel Roberts Noble Foundation, Inc., Ardmore, Oklahoma

<sup>†</sup>Corresponding author (jhe@noble.org)

Next-generation sequencing (NGS) technologies can generate multi-billion base-pairs of reads per machine day. It became a challenging task for large sequence data analysis when only using desktop computers with limited computational capacities. For small or mid scale academic laboratories which lack large central computing pool and data storage capacity, one potential solution to the NGS data analysis is the use of cloud computing technologies. Microsoft Azure platform provides a promising computation platform to handle such tasks.

Sequence alignment is the essential first step to NGS analysis. Multiple algorithms have been used for NGS alignment. Among them, BWA is one of NGS alignment software which has best accuracy. Based on the BWA source code, we developed a parallelized CloudBWA alignment software which has been successfully migrated to Microsoft cloud computation platform. Using this software, users will be able to carry out sequence data alignment using remote computational resource; thus, users will take the full advantage of parallel worker nodes in Azure platform and significantly reduce the calculation time accordingly. For biologists' convenient access to BWA functions, we also developed a user-friendly interface for data exchange, parameter tune-up, alignment and result visualization. To test the performance of CloudBWA using multiple worker roles, we also performed tests and compared the running time with different worker role numbers.

### Separation of Terminal Restriction Fragment DNA on a Twodimensional Gel (T-RFs-2D) – A Novel Method for Profiling Complex Microbial Communities

Shanquan Wang\* and Jianzhong He (jianzhong.he@nus.edu.sg)

Department of Civil and Environmental Engineering, National University of Singapore, Singapore

Current fingerprinting techniques such as terminal restriction fragment length polymorphism (T-RFLP) and denaturing gradient gel electrophoresis (DGGE) has limitations on characterizing complex microbial communities, such as different populations possibly sharing the same terminal restriction fragments (T-RFs), or DNA fragments unable to be separated well on DGGE gels. To overcome these limitations, an efficient approach – separation of terminal restriction fragments of 16S rRNA genes on a two-dimensional gel (T-RFs-2D) – was developed. T-RFs-2D involves restriction digestion of terminal fluorescence-labelled PCR amplified 16S rRNA gene products and their separation via a two-dimensional (2D) gel electrophoresis. High-resolution DNA separation maps can be generated according to both T-RFs' size (on the first dimension) and their sequence composition (on the second dimension). The sequence information of interested T-RFs on 2D gels can be obtained through serial poly(A) tailing reaction, PCR amplification, and subsequent DNA sequencing. T-RFs-2D has been demonstrated with

Posters alphabetical by first author. \*Presenting author

improved genotyping capabilities and is a powerful tool to obtain sequence information of specific T-RFs or characterize complex microbial communities, e.g., by employing T-RFs-2D method, 63 operational taxonomic units (OTUs) were identified in a complex river-sediment microbial community, while only 41 OTUs were identified by traditional DGGE in the same sample. T-RFs-2D method sheds light on better understanding of complex microbial ecologies in the environmental samples.

# **Bacterial Community Structure Predicts Function in Full-Scale Bioenergy Systems**

**Jeffrey J. Werner**,<sup>1</sup> Dan Knights,<sup>2</sup> Marcelo L. Garcia,<sup>3</sup> Nicholas B. Scalfone,<sup>1</sup> Samual Smith,<sup>4</sup> Kevin Yarasheski,<sup>4</sup> Theresa A. Cummings,<sup>5</sup> Allen R. Beers,<sup>5</sup> Rob Knight,<sup>2</sup> and Largus T. Angenent<sup>1</sup>\* (la249@cornell.edu)

<sup>1</sup>Cornell University, Ithaca, New York; <sup>2</sup>University of Colorado at Boulder, Colorado; <sup>3</sup>Washington University in Saint Louis, Missouri; <sup>4</sup>Washington University School of Medicine, St. Louis, Missouri; and <sup>5</sup>Anheuser-Busch, Inbev Inc., St. Louis, Missouri

Microbial communities of mixed species can handle complex and varying mixtures of substrates, which is a characteristic of organic waste streams. Therefore, these communities are used in, for example, bioenergy systems to convert wastes into methane gas. However, researchers have not been successful in producing liquid transportation fuels with undefined mixed communities. For this we would need to intentionally shape mixed communities to perform specific metabolic tasks while maintaining their inherent ability to withstand variable conditions. We do not know enough about the structure-function relationships of these communities to start this process of engineering new functions. Therefore, we performed a deep analysis of the structure and dynamics of microbial communities as a function of performance and operating conditions of the most successful bioenergy system in existence – anaerobic digesters. The combination of genome-enabled surveys with highly-parallel multiplex sequencing, the vast amount of measurements that are performed at these full-scale bioenergy systems, and powerful ordination algorithms and machine learning made this analysis possible for the first time (http://www.pnas.org/content/early/2011/02/15/1015676108.abstract).

Specifically, we investigated the relationships of bacterial community structure (>400,000 16S rRNA gene sequences for 112 samples) with function (i.e., bioreactor performance) and environment (i.e., operating conditions) in a yearlong monthly time series of nine fullscale anaerobic digesters treating brewery wastewater (>20,000 measurements). Each of the nine facilities had a unique community structure with an unprecedented level of stability due to resilience to perturbations. Using machine learning, we identified a small subset of operational taxonomic units (OTUs; 145 out of 4,962), which predicted the location of the facility of origin for almost every sample (96.4% accuracy). Of these 145 OTUs, syntrophic bacteria were systematically overrepresented, demonstrating that syntrophs rebounded following disturbances. This indicates that resilience, rather than dynamic competition, played an important role in maintaining the necessary syntrophic keystone populations. In addition, we explained the observed phylogenetic differences between all samples based on a subset of environmental gradients (with constrained ordination algorithms), and found stronger relationships between community structure and its function rather than its environment. These relationships were strongest for two performance variables - methanogenic activity and substrate removal efficiency - both of which were also affected by ecology, evidenced by the fact that these variables were correlated with community evenness (at any given time) and variability in phylogenetic structure (over time), respectively. It is, therefore, not only important to foster species that
perform the important desired metabolic pathways, but also to make sure that the microbial community is even and variable. Our work quantified the relationships between community structure and function, which opens the door to intentionally shape microbial communities.

## Metatranscriptomics of H<sub>2</sub>-evolving Cyanobacterial Mats

**Dagmar Woebken**,<sup>1,2</sup> **Luke C. Burow**,<sup>1,2</sup> Ian P. Marshall,<sup>2</sup> Leslie Prufert-Bebout,<sup>1</sup> Brad M. Bebout,<sup>1</sup> Tori M. Hoehler,<sup>1</sup> Jennifer Pett-Ridge<sup>3</sup>\* (pettridge2@llnl.gov), Peter K. Weber,<sup>3</sup> Alfred M. Spormann,<sup>2</sup> and Steven W. Singer<sup>4</sup>

<sup>1</sup>NASA Ames Research Center, Mountain View, California; <sup>2</sup>Stanford University, Palo Alto, California; <sup>3</sup>Lawrence Livermore National Laboratory, Livermore, California; and Lawrence Berkeley National Laboratory, Berkeley, California

Photosynthetic microbial mats found in coastal environments are complex, stratified microbial communities that fix CO<sub>2</sub> under aerobic conditions during the day and are hypothesized to ferment the fixed carbon under anaerobic conditions at night, generating large amounts of  $H_2$  and organic acids. Fermentation of accumulated photosynthate may be required to provide energy for anaerobic  $N_2$  fixation. To understand nutrient flux through these mats and relate this flux to the observed rates of  $H_2$  evolution and  $N_2$  fixation, we have analyzed the upper 2 mm of a cyanobacterial mat collected from Elkhorn Slough, CA. Biogeochemical experiments have shown that the upper layer of these mats possess high rates of fermentation and  $N_2$  fixation, however rDNA analysis demonstrated that these microbial communities were too complex to characterize effectively by metagenomic analysis. Therefore, we chose to focus our efforts on characterizing the active community using metatranscriptomic sequencing. Clone library and 454 pyrosequencing of the rRNA expressed by this community demonstrated that the active community was dominated by filamentous cyanobacteria, of which the most significant populations were closely related to *Microcoelus chthonoplastes* and a novel cyanobacterial clade, UD3, that has been shown to be an important diazotrophic population in these mats. Approximately 388,000 transcripts (average length ~390 bp) were obtained by sequencing cDNA from two mats samples collected under dark, anoxic conditions. Of these transcripts, ~199,00 transcripts (51%) were classified as protein coding sequences by comparison to the SEED database  $(1e^{-5} \text{ cutoff})$ . Transcripts related to *Cyanobacteria* and *Chloroflexi* represent >70% of the total classified sequences in the transcriptome, suggesting that processes these phyla are important contributors to the nutrient flux in the mats. Metabolic reconstruction of the genome of *Microcoleus chthonoplastes* and fragment recruitment of the transcripts to this genome demonstrated that a complete pathway for fermentation of photosynthate by M. chtonoplastes was expressed in these mats. Genomic sequencing of enrichments of UD3 have been undertaken to correlate the transcripts observed in the mat metatranscriptome with metabolic reconstructions of this unique cyanobacterial clade.

## Lignocellulolytic System of Shiitake Mushroom Lentinula edodes

**M.C. Wong**, C.H. Au, I.S.W. Kwok, X. Luo, K.S. Wong, L. Xing, and H.S. Kwan\* (hoishankwan@cuhk.edu.hk)

The Chinese University of Hong Kong, Hong Kong SAR, People's Republic of China

*Lentinula edodes* is a white-rot basidiomycete which can efficiently degrade lignin and hydrolysis polysaccharides in wood. In order to understand the molecular mechanism of lignocellulolytic in basidiomycete comprehensively, we searched for the genes which are

Posters alphabetical by first author. \*Presenting author

coding for lignocellulolytic enzymes of *L. edodes* by using the whole-genome approach and carried out comparative analysis with other basidiomycetes.

Whole-genome sequencing of *L. edodes* monokaryon L54A was performed. Over 12,000 gene models were predicted from the genome sequence. The gene models were annotated by BLAST and categorized according to Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). In addition, the gene models were subjected to BLASTP against FOLy (Fungal Oxidative Lignin enzymes) and CAZy (Carbohydrate-Active enZymes) databases to identify all the putative lignocellulolytic enzymes. Those having e-value  $<10^{-5}$  were further examined and classified into different FOLy and CAZy groups. Comparison of the sets of FOLymes and CAZymes among basidiomycetes revealed the unique lignocellulolytic system of *L. edodes*. A number of putative lignocellulolytic enzyme genes were selected for quantitative real-time RT-PCR analysis. Differential expression of the genes at various developmental stages and different media (lignocellulosic and non-lignocellulosic) suggested that these genes might play different roles during the development of *L. edodes*. This study helps us to have better understanding about the lignocellulolytic system of *L. edodes*.

# Plant and Genomic Resources to Enhance K-12 and Undergraduate Education in Genetics

#### Scott T. Woody\* (swoody@wisc.edu) and Rick Amasino

University of Wisconsin-Madison, Madison, Wisconsin

We have developed a self-compatible and extensively inbred variety of rapid-cycling Brassica rapa (analogous to the iconic Wisconsin Fast Plants variety) to augment K-12 and undergraduate education in genetics. We have used chemical (EMS) mutagenesis to generate a large collection of mutant derivatives whose phenotypes are useful for exploring Mendelian inheritance patterns and, with the assistance of JGI, we are developing a complementary suite of molecular resources that permit inquiry-driven investigations that make use of the tools of modern genomics. Specifically, the JGI is working to determine the WGS of the reference wild type strain, B3, and a polymorphic B. rapa variety, R500, in the context of a comparative genomics initiative (CGI) targeting selected species that represent the primary phylogenetic clades within the order Brassicales. The Brassicales includes the model plant Arabidopsis and several agronomically important Brassica species; this CGI builds on the solid foundation of Arabidopsis research to advance plant genomic sciences, generally, as the research community moves outward to explore increasingly distant relatives. For our purposes, the WGS will facilitate development of robust and straightforward (PCR-generated) genetic markers that can be used in genetic mapping experiments to enhance students' understanding of fundamental genetic principles such as linkage and to more readily enable students to comprehend the connection between evident phenotype and underlying genotype. We expect that an integrated approach to genetics education that combines hands-on experiments with living organisms and insights increasingly driven by genomics will help to better prepare our students for understanding and practicing 21<sup>st</sup> century biological sciences.

# **Microbial Sentinels of Environmental Change**

**Jody J. Wright**<sup>1</sup>\* (jjwright@interchange.ubc.ca), Alyse K. Hawley, <sup>1</sup> Capser Shyr, <sup>1</sup> Young C. Song, <sup>1</sup> Susannah Tringe, <sup>2</sup> Philippe Tortell, <sup>3</sup> and Steven J. Hallam<sup>1</sup>

<sup>1</sup>Department of Microbiology & Immunology, University of British Columbia, Vancouver, British Columbia, Canada; <sup>2</sup>DOE Joint Genome Institute, Walnut Creek, California; <sup>3</sup>Department of Earth and Ocean Science, University of British Columbia, Vancouver, British Columbia, Canada

Regions of low dissolved oxygen known as oxygen minimum zones (OMZs) are widespread oceanographic features currently expanding due to global warming. Although inhospitable to metazoan life, OMZs support a thriving but cryptic microbiota whose combined metabolic activity is intimately connected to nutrient and trace gas cycling within the global ocean. Therefore, OMZ expansion and intensification represents an emerging ecological phenomenon with potentially harmful effects on ocean health and climate balance. In order to understand, respond to, or mitigate these transitions, studies monitoring and modeling dynamics and systems metabolism of OMZ microbiota in relation to physical and chemical oceanographic parameters are imperative. To this end we are using environmental genomic approaches to chart microbial community responses to changing levels of water column oxygen-deficiency in the northeastern subarctic Pacific Ocean (NESAP). The NESAP is one of the world's most extensive OMZs and provides an exceptional model system for long-term observation and process-oriented studies of OMZ phenotypes.

# Production and Partial Characterization of a Novel Thermostable Xylanase by Newly Isolated *Kluyvera georgiana* OM3

Fengxue Xin\* (esexf@nus.edu.sg) and Jianzhong He

Department of Civil and Environmental Engineering, National University of Singapore, Singapore

Hemecellulose is the most abundant heteropolymer and second most abundant renewable biomass in nature after cellulose, which accounts for 25-35% of lignocellulosic biomass. The most important enzyme in hemecellulose degradation is endo-1,4-xylanase. In this study, A novel *Kluyvera Georgiana* OM3 was isolated from mushroom residue wastes and showed high xylanolytic activity under mesophilic, anaerobic conditions. Different xylancontaining agricultural byproducts such as palm oil fiber, wood chips, saw dust and birchwood xylan, beechwood xylan were tested as the carbon source for strain OM3 hydrolyzing stidies. The highest xylanase activity 2500 mU/ml was obtained by using a medium containing 1% birchwood xylan, 0.2% yeast extract and peptone, at 30°C and pH 9.0 under anaerobic conditions. The partially purified enzyme recovered after ammonium sulfate fractionation showed maximum activity at 70°C and pH 8.0. The enzyme was stable over a broad pH range and showed good thermal stability when incubated at 60°C and pH 8.0. Kinetic parameters such as  $K_m$  and  $V_{max}$  for xylan were found to be 6.42 mg/ml and 4.44 IU/mg. Thus far, this enzyme could be considered as a thermotolerant biocatalyst being interesting for biotechnological applications.

# Quantitative Karyotyping and Cytogenetic Mapping of Switchgrass Chromosomes

**Hugh A. Young\*** (hugh.young@ars.usda.gov), Christina L. Lanzatella, and Christian M. Tobias

USDA, ARS, Western Regional Research Center, Genomics and Gene Discovery Unit, Albany, California

Switchgrass (*Panicum virgatum* L.), a warm season C4 perennial grass native to North America, is rapidly advancing as a dedicated feedstock for renewable bioenergy. Despite prolific interest in switchgrass as an energy crop, little is known about genome structure and chromosome architecture in this polyploid species. We have identified two dihaploid (2n=2x=18) individuals from among the progeny of a controlled cross between the tetraploid cultivars Alamo and Kanlow (2n=4x=36). Mitotic chromosome spreads made from these dihaploid individuals have facilitated karyotyping of the base 9 switchgrass chromosomes. We present cytological analyses of physical lengths, centromere locations, and arm ratios as a way to distinguish individual chromosome swere quantitatively analyzed using CHIAS IV software. Fluorescence in situ hybridization (FISH) has also allowed chromosomal assignment of 45S rDNA, CentC, and a centromere-specific repetitive switchgrass sequence. Quantitative karyotyping in combination with cytogenetic mapping has allowed the identification of switchgrass chromosomes, creating a foundation for future genome analyses based on chromosome structure and identity.

## **Development of Resources for Switchgrass Functional Genomics**

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

<sup>1</sup>Samuel Roberts Noble Foundation, Ardmore, Oklahoma; <sup>2</sup>DOE Joint Genome Institute, Walnut Creek, California; <sup>3</sup>University of Georgia, Athens, Georgia; <sup>4</sup>HudsonAlpha Genome Sequencing Center, Huntsville, Alabama; <sup>5</sup>University of California, Davis, California; and <sup>6</sup>DOE BioEnergy Science Center

Switchgrass (P. virgatum L.) is a perennial C4 grass native to North America. It has been used as forage and for soil conservation and has the potential to become a major source of biomass for biofuel production. To realize this potential, breeding and genetic engineering efforts are underway to improve existing germplasm. As the first step towards developing a set of functional genomics resources that are essential for gene discovery, basic biology research, and molecular breeding efforts, large numbers of expressed sequence tags (EST) have been generated for two tetraploid switchgrass genotypes, AP13 a lowland "Alamo" genotype and VS16 a genotype of upland "Summer". In addition to over 11.5 million high quality ESTs generated by 454/Roche pyrosequencing technology, three full-length enriched cDNA libraries were constructed with RNA from multiple AP13 tissues grown under optimal and stress conditions. About 100,000 clones were sequenced from both ends with the Sanger method and over 69,000 high quality reads were produced. To optimize sequence assembly strategies, different programs including the classical CAP3 were tested, and a two-stage approach was finally selected to assemble AP13 unitanscripts. First, 454 ESTs were assembled into 102,000 isotig/contigs using the Newbler program with stringent parameters (overlap 100 bp and identity at 99%). PAVE was then used to assemble Sanger reads and the processed 454 isotig/contigs into ~80,000 unique transcript

sequences. Separately, the VS16 454 ESTs were assembled into ~34,000 isotig/contigs using Newbler with the same parameters. To create a switchgrass gene index for gene annotation and Affymetrix cDNA chip design, a total of 545,000 Sanger ESTs of other genotypes in the public domain were downloaded, grouped, and assembled using the PAVE program. A final 132,000 unigene set (PviUT1.2) was generated from existing ESTs with priority order of AP13, Alamo, Kanlow, VS16, and other sequences including about 1502 virtual transcripts predicted from AP13 BAC sequences. The Affymetrix cDNA microarray chip (Pvi cDNAa520831) based on PviUT1.2 contains ~122,400 probe sets. This chip has an 11µm feature size, with 11 probes for each transcript without mismatch probes. These represent 104,871 switchgrass unitranscript sequences with one or two probe sets. The chip is available to the public through Affymetrix Inc. A switchgrass gene expression atlas is being generated with this platform. The sequence resources will be used for gene annotation, prediction of transcription factor and other gene families of interest, and SNP identification. All switchgrass ESTs generated by this project and the assembled unigene set PviUT1.2 have been deposited to the Switchgrass Genomics database hosted by the Noble Foundation and accessible through the this web link: http://switchgrassgenomics.noble.org/.

# Assessment of Single Molecule Sequencing for Microbial Genome Assembly

**Zhiying Jean Zhao,** Joel Martin, Rob Egan, Matt Blow, Zhong Wang, Kanwar Singh\* (KSingh@lbl.gov), Cindi Hoover, Tao Zhang, Len A Pennacchio, and Feng Chen

DOE Joint Genome Institute, Walnut Creek, California

The assembly and analysis of microbial species on earth remains a largely unexplored area of life. This is partially due to their inability to be cultured but also based on the large historic cost of drafting and finishing individual microbial species genomes. Single molecule DNA sequencing provides the potential for a quick turnaround and cost-effective solution to further scale the ability to provide reference genomes for this important sector of biology. Here, we present our exploration of the recently developed single molecule real time (SMRT<sup>TM</sup>) technology as a future approach to exponentially scale our ability to sequence microbial isolates at reduced cost. Preliminary data supports the validity of the approach and ongoing studies will be presented to discuss the status of the technology and its trajectory moving forward for microbial genome assembly and beyond.

# **Ancient Nature of Alternative Splicing and Function of Introns**

Kemin Zhou\* (KZhou@lbl.gov), Asaf Salamov, Alan Kuo, Andrea Aerts, and Igor Grigoriev

DOE Joint Genome Institute, Walnut Creek, California

Equipped with a new algorithm COMBEST that generates gene models from EST and genomic sequences, we analyzed alternative splicing and function of introns in four genomes: *Chamydomonas reinhardtii*, *Agaricus bisporus*, *Aspergillus carbonarius*, and *Sporotricum thermophile* with EST coverage of 2.9x, 8.9x, 29.5x, and 46.3x respectively.

We found the percentage of alternative splicing (AS) 15, 35, 52, and 63% in multiple exon genes respectively. Alternative spliced genes are more ancient, and AS is conserved at phylum level. Intron containing genes are over expressed compared to intronless genes

consistent with intron boosting gene expression. Linear regression analysis demonstrated that the number of alternative spliced isoforms correlate with number of exons, expression level, and maximum intron length of the gene. Intron retention (RI) is the dominant forms of AS in all four genomes. Shorter introns are more likely to be retained. Stop codons are suppressed in introns. Introns in coding regions uses either stop codon or length of not multiples of three (3n+1, 3n+2) to trigger NMD; a small set of introns that are retained in major RI isoforms (major RI, 0.2-6% of all introns) favor 3n length presumably generate protein diversity. Introns of 3n length favor phase 1 that can be explained by more flexible and hydrophilic amino acids in both ends of phase 1 introns when retained which would favor insertion of short peptides into protein structures. We proposed a model in which minor RI intron could evolve into major RI. This model predicts a new intron loss mechanism through abolishing the non-RI alternative isoform.

#### Current as of March 1, 2011

Andrea Aerts DOE Joint Genome Institute alaerts@lbl.gov

Oleg Alexandrov DOE Joint Genome Institute oalexandrov@lbl.gov

Gul Ali University of Floroda gul.ali@colostate.edu

Ed Allen DOE Joint Genome Institute eaallen@lbl.gov

Iain Anderson DOE Joint Genome Institute IJAnderson@lbl.gov

Largus Angenent Cornell University la249@cornell.edu

Chun Hang Au The Chinese University of Hong Kong tommyau@cuhk.edu.hk

Frank Aylward University of Wisconsin, Madison faylward@wisc.edu

Jacob Baelum Lawrence Berkeley National Lab jbaelum@lbl.gov

Cheryl Bailey University of Nebraska, Lincoln cbailey7@unl.edu

Massie Ballon DOE Joint Genome Institute mlballon@lbl.gov

Kerrie Barry DOE Joint Genome Institute kwbarry@lbl.gov

John Battista Louisiana State University jbattis@lsu.edu

Randy Berka Novozymes, Inc. ramb@novozymes.com

Karan Bhatia DOE Joint Genome Institute kbhatia@lbl.gov

Srijak Bhatnagar University of California, Davis sbhatnagar@ucdavis.edu Robert Blazej Allopartis Biotechnologies, Inc.robert@allopartis.com

Matthew Blow DOE Joint Genome Institute mjblow@lbl.gov

Annette Bollmann Miami University bollmaa@muohio.edu

Jeffrey Boore Genome Project Solutions, UC Berkeley jlboore@genomeprojectsolutions.com

Jennifer Bragg USDA, ARS, WRRC jennifer.bragg@ars.usda.gov

Jim Bristow DOE Joint Genome Institute JBristow@lbl.gov

Eoin Brodie Lawrence Berkeley National Lab elbrodie@lbl.gov

David Bruce DOE Joint Genome Institute, LANL dbruce@lanl.gov

Edward Buckler USDA-ARS, Cornell University esb33@cornell.edu

Yury Bukhman Great Lakes Bioenergy Res. Center ybukhman@glbrc.wisc.edu

Emmanuel Buschiazzo University of California, Merced ebuschiazzo@ucmerced.edu

Mike Cantor DOE Joint Genome Institute mncantor@lbl.gov

Manu Capoor Rockefeller University Mcapoor@Rockefeller.edu

Charlotte Carlstrom University of California, Berkeley ccarlst@berkeley.edu

Romy Chakraborty Lawrence Berkeley National Lab RChakraborty@lbl.gov

Patricia Chan University of California, Santa Cruz pchan@soe.ucsc.edu Feng Chen DOE Joint Genome Institute fchen@lbl.gov

Amy Chen Lawrence Berkeley National Lab, JGI IMAChen@lbl.gov

Jan-Fang Cheng DOE Joint Genome Institute jfcheng@lbl.gov

Swapnil Chhabra Lawrence Berkeley National Labsrchhabra@lbl.gov

**Tony Chiang** University of Washington chiang@ocean.washington.edu

Stephen Chisholm Colorado State University chisholm@colostate.edu

**Dylan Chivian** Lawrence Berkeley National Lab DCChivian@lbl.gov

Cindy Choi DOE Joint Genome Institute cchoi@lbl.gov

Julianna Chow DOE Joint Genome Institute jchow@lbl.gov

Ulla Christensen Joint BioEnergy Institute UChristensen@lbl.gov

Ken Chu Lawrence Berkeley National Lab, JGI klchu@lbl.gov

Scott Clingenpeel DOE Joint Genome Institute srclingenpeel@lbl.gov

Alicia Clum DOE Joint Genome Institute aclum@lbl.gov

Frank Collart Argonne National Lab fcollart@anl.gov

Tudor Constantin Eureka Genomics tudor@eurekagenomics.com

Robert Cottingham Oak Ridge National Laboratory cottinghamrw@ornl.gov

Aaron Cozen University of California, Santa Cruz acozen@soe.ucsc.edu

Kelly Craven The Samuel Roberts Noble Foundation kdcraven@noble.org

Daniel Cullen University of Wisconsin, Madison dcullen@facstaff.wisc.edu

David Culley Pacific Northwest National Lab david.culley@pnl.gov

John Cumbers NASA Ames, Brown University john.cumbers@nasa.gov

John Curry Investigen, Inc curry@investigen.com

Jeff Dangl University of North Carolina dangl@email.unc.edu

Chris Daum DOE Joint Genome Institute daum1@llnl.gov

Karen Davenport Los Alamos National Laboratory kwdavenport@lanl.gov

Jennifer Davis Brown University jennifer\_reeve@brown.edu

Ronald De Vries CBS-KNAW Fungal Biodiversity Centre r.devries@cbs.knaw.nl

Kristen DeAngelis Lawrence Berkeley National Lab kdeangelis@lbl.gov

Janine Detter Los Alamos National Lab janined@lanl.gov

Patrik D'haeseleer Lawrence Livermore National Lab, JBEI patrikd@gmail.com

Michael Dougherty Joint BioEnergy Institute mjdougherty@lbl.gov

Persis Drell SLAC National Accelerator Laboratory persis@slac.stanford.edu

Changbin Du DOE Joint Genome Institute cdu@lbl.gov

Inna Dubchak DOE Joint Genome Institute, LBNL ildubchak@lbl.gov Kecia Duffy DOE Joint Genome Institute kmduffy-wei@lbl.gov

Kathleen Duncan University of Oklahoma kathleen.e.duncan-1@ou.edu

Erin Dunwell DOE Joint Genome Institute emdunwell@lbl.gov

Sébastien Duplessis INRA (French National Institute for Agricultural Research) duplessi@nancy.inra.fr

Paul Dyer University of Nottingham paul.dyer@nottingham.ac.uk

Elizabeth Edwards University of Toronto elizabeth.edwards@utoronto.ca

Rob Egan DOE Joint Genome Institute rsegan@lbl.gov

Jonathan Eisen UC Davis Genome Center jaeisen@ucdavis.edu

Jaana Ekojärvi University of Helsinki jaana.ekojarvi@helsinki.fi

Hamza El Dorry The American University in Cairo dorry@aucegypt.edu

Anna Engelbrektson University of California, Berkeley aengelbrektson@berkeley.edu

Aren Ewing DOE Joint Genome Institute aewing@lbl.gov

Shui-zhang Fei Iowa State University sfei@iastate.edu

Marsha Fenner DOE Joint Genome Institute MWFenner@lbl.gov

Jen Fisher Desert Research Institute (DRI) jen.fisher@dri.edu

Susan Fuerstenberg Genome Project Solutions sifuerst@genomeprojectsolutions.com

Sita Ghimire The Samuel Roberts Noble Foundation srghimire@noble.org

David Gilbert DOE Joint Genome Institute degilbert@lbl.gov Peter Girguis Harvard University pgirguis@oeb.harvard.edu

**Tijana Glavina del Rio** DOE Joint Genome Institute tglavina@lbl.gov

Barry Goldman Monsanto bsgold@monsanto.com

Yunchen Gong University of Toronto yunchen.gong@utoronto.ca

Brad Goodner Hiram College GoodnerBW@hiram.edu

Lynne Goodwin Los Alamos National Laboratory lynneg@lanl.gov

Sean Gordon USDA, Albany seangordon@gmail.com

Andrey Grigoriev Rutgers University agrigoriev@camden.rutgers.edu

Igor Grigoriev DOE Joint Genome Institute ivgrigoriev@lbl.gov

Stephen Gross DOE Joint Genome Institute smgross@lbl.gov

Christopher Hack DOE Joint Genome Institute cahack@lbl.gov

Megan Hall University of California, Berkeley mhall45@berkeley.edu

Shunsheng Han Los Alamos National Laboratory han\_cliff@lanl.gov

Sara Hansen Joint BioEnergy Institute sfhansen@lbl.gov

Loren Hauser Oak Ridge National Laboratory hauserlj@ornl.gov

Margo Haygood Oregon Health & Science Institute haygoodm@ebs.ogi.edu

Shaomei He DOE Joint Genome Institute she@lbl.gov

Sabine Heinhorst The University of Southern Mississippi sabine.heinhorst@usm.edu

Uffe Hellsten DOE Joint Genome Institute uhellsten@lbl.gov

Joshua Herr The Pennsylvania State University jrh408@psu.edu

Sur Herrera Paredes University of North Carolina sur00mx@gmail.com

Matthias Hess Washington State University matthias.hess@tricity.wsu.edu

Jaqueline Hess Harvard University jaqueline.hess@gmail.com

David Hibbett Clark University dhibbett@clarku.edu

Uwe Hilgert Dolan DNA Learning Center, CSHL hilgert@cshl.edu

Nils Hoegberg Swedish Univ. of Agricultural Sciences nils.Hogberg@mykopat.slu.se

Susan Hua DOE Joint Genome Institute shua@lbl.gov

Jenni Hultman Lawrence Berkeley National Lab jjhultman@lbl.gov

William Inskeep Montana State University binskeep@montana.edu

Javier Izquierdo Dartmouth College javier.izquierdo@dartmouth.edu

Janet Jansson Lawrence Berkeley National Lab jrjansson@lbl.gov

Jerry Jenkins DOE JGI, HudsonAlpha jjenkins@hudsonalpha.org

Tomas Johansson Lund University tomas.johansson@mbioekol.lu.se

**Tom Juenger** University of Texas at Austin tjuenger@austin.utexas.edu

Ulas Karaoz Lawrence Berkeley National Lab UKaraoz@lbl.gov

Lisa Kegg DOE Joint Genome Institute lrkegg@lbl.gov **Richard Kerrigan** Sylvan Research rwk@sylvaninc.com

Rob Knight University of Colorado, HHMI rob.knight@colorado.edu

Annegret Kohler INRA (French National Institute for Agricultural Research) kohler@nancy.inra.fr

Wayne Kontur University of Wisconsin, Madison wskontur@wisc.edu

Heather Koshinsky Eureka Genomics heather@eurekagenomics.com

Lauralynn Kourtz C5-6 Technologies, Inc. lkourtz@c56technologies.com

Cheryl Kuske Los Alamos National Laboratory kuske@lanl.gov

Christopher Kvaal St. Cloud State University cakvaal@stcloudstate.edu

Hoi Shan Kwan Chinese Univ of Hong Kong hoishankwan@cuhk.edu.hk

Nikos Kyrpides DOE Joint Genome Institute NCKyrpides@lbl.gov

Miriam Land Oak Ridge National Laboratory landml@ornl.gov

**Debbie Laudencia-Chingcuanc** US Department of Agriculture debbie.laudencia@ars.usda.gov

Janey Lee DOE Joint Genome Institute jlee2@lbl.gov

Ruth Ley Cornell University rel222@cornell.edu

Mingkun Li DOE Joint Genome Institute mli@lbl.gov

Dawei Lin UC Davis Genome Center lhslin@ucdavis.edu

Anna Lipzen DOE Joint Genome Institute alipzen@lbl.gov

Wen-Tso Liu University of Illinois wtliu@illinois.edu Todd Lowe University of California, Santa Cruz lowe@soe.ucsc.edu

Susan Lucas DOE Joint Genome Institute SMLucas@lbl.gov

Derek Lundberg Univ. of North Carolina, Chapel Hill derek.lundberg@gmail.com

Taina Lundell University of Helsinki taina.lundell@helsinki.fi

Jacqueline MacDonald University of Toronto jacqueline.macdonald@utoronto.ca

Eugene Madsen Cornell University elm3@cornell.edu

Miia Mäkelä University of Helsinki miia.r.makela@helsinki.fi

**Stephanie Malfatti** DOE Joint Genome Institute SAMalfatti@lbl.gov

Yuzuki Manabe Joint BioEnergy Institute ymanabe@lbl.gov

Kristen Marhaver University of California, Merced kmarhaver@ucmerced.edu

Victor Markowitz DOE Joint Genome Institute vmmarkowitz@lbl.gov

Francis Martin INRA (French National Institute for Agricultural Research) fmartin@nancy.inra.fr

Jeffrey Martin DOE Joint Genome Institute jamartin@lbl.gov

Joel Martin DOE Joint Genome Institute j\_martin@lbl.gov

Angel T. Martinez CIB, CSIC atmartinez@cib.csic.es

Emma Master University of Toronto emma.master@utoronto.ca

Konstantinos Mavrommatis DOE Joint Genome Institute KMavrommatis@lbl.gov

Jack McFarland US Geological Survey jmcfarland@usgs.gov

Michael Melnick CMEA Capital michael@cmea.com

Ryan Melnyk University of California, Berkeley melnyk@berkeley.edu

Xiandong Meng DOE Joint Genome Institute xiandongmeng@lbl.gov

Keith Mewis University of British Columbia keithmewis@gmail.com

Folker Meyer Argonne National Laboratory folker@mcs.anl.gov

Bonnie Millenbaugh DOE Joint Genome Institute millenbaugh1@llnl.gov

John Miller University of Maryland jjmiller@umd.edu

Mike Millikin Green Car Congress mmillikin@bioagemedia.com

Maria Monteros Noble Foundation mjmonteros@noble.org

Mary Ann Moran University of Georgia mmoran@uga.edu

Geoffrey Morris University of Chicago gmorris@uchicago.edu

Mustafa Morsy The Samuel Roberts Noble Foundation mrmorsy@noble.org

**Olaf Mueller** Duke University olaf.mueller@duke.edu

**Biswarup Mukhopadhyay** Virginia Tech biswarup@vt.edu

Senthil Murugapiran University of Nevada, Las Vegas senthil.murugapiran@unlv.edu

Gerard Muyzer Delft University of Technology g.muijzer@tudelft.nl

Pejman Naraghi-Arani Lawrence Livermore National Lab pnaraghi@llnl.gov

**Donald Natvig** University of New Mexico dnatvig@gmail.com Ali Navid Lawrence Livermore National Lab navid1@llnl.gov

Saraswoti Neupane Swedish Univ. of Agricultural Sciences Saraswoti.Neupane@slu.se

Jessica Newburn Desert Research Institute (DRI) jessica.newburn@dri.edu

Chew Yee Ngan DOE Joint Genome Institute cyngan@lbl.gov

Jessica Nguyen Eureka Genomics jessica@eurekagenomics.com

Matt Nolan DOE Joint Genome Institute mpnolan@lbl.gov

Magnus Nordborg Gregor Mendel Institute magnus.nordborg@gmi.oeaw.ac.at

Jeanette Norton Utah State University jeanette.norton@usu.edu

Mari Nyyssonen Lawrence Berkeley National Lab mjnyyssonen@lbl.gov

**Kylea Odenbach** Sandia National Laboratories kjodenb@sandia.gov

Ai Oikawa Joint BioEnergy InstituteAOikawa@lbl.gov

Ilona Oksanen University of Helsinki ilona.oksanen@helsinki.fi

Peter Olsen Novozymes pbo@novozymes.com

Åke Olson Swedish Univ. of Agricultural Sciences ake.olson@slu.se

Krishnaveni Palaniappan Lawrence Berkeley National Lab, JGI kpalaniappan@lbl.gov

Katherine Pappas University of Athens kmpappas@biol.uoa.gr

Len Pennacchio DOE Joint Genome Institute lapennacchio@lbl.gov

Christa Pennacchio DOE Joint Genome Institute cppennacchio@lbl.gov Rene Perrier DOE Joint Genome Institute raperrier@lbl.gov

Lin Peters DOE Joint Genome Institute, LBNL lgpeters@lbl.gov

Jennifer Pett-Ridge Lawrence Livermore National Lab pettridge2@llnl.gov

Antonio Pisabarro University of Navarre gpisabarro@unavarra.es

Samuel Pitluck DOE Joint Genome Institute s\_pitluck@lbl.gov

Juergen Polle Brooklyn College of CUNY jpolle@brooklyn.cuny.edu

Paraskevi Polymenakou Hellenic Centre for Marine Research olymen@her.hcmr.gr

Andrea Porras-Alfaro Western Illinois University a-porras-alfaro@wiu.edu

Amy Powell Sandia National Laboratories ajpowel@sandia.gov

Abhishek Pratap DOE Joint Genome Institute apratap@lbl.gov

Simon Prochnik DOE Joint Genome Institute seprochnik@lbl.gov

Theodore Raab Stanford University tkraab@stanford.edu

Lara Rajeev Lawrence Berkeley National Lab Irajeev@lbl.gov

Lucía Ramírez University of Navarre LRamirez@unavarra.es

Jayashree Ray Lawrence Berkeley National Lab jray@lbl.gov

Joanna Redfern University of New Mexico jredfern@unm.edu

Kelynne Reed Austin College kreed@austincollege.edu

Kathryn Richmond Great Lakes Bioenergy Res. Center kerichmond@wisc.edu

**Robert Riley** DOE Joint Genome Institute rwriley@lbl.gov

Christian Rinke DOE Joint Genome Institute crinke@lbl.gov

Frank Roberto Idaho National Laboratory francisco.roberto@inl.gov

Simon Roberts DOE Joint Genome Institute srroberts@lbl.gov

Gene Robinson University of Illinois generobi@illinois.edu

Jorge Rodrigues University of Texas at Arlington jorge@uta.edu

Dan Rokhsar DOE Joint Genome Institute DSRokhsar@gmail.com

Rafael Rosengarten Joint BioEnergy Institute rdrosengarten@lbl.gov

Jennifer Saito Universiti Sains Malaysia jsaito@gmail.com

Asaf Salamov DOE Joint Genome Institute aasalamov@lbl.gov

Antonio Sanchez-Amat University of Murcia antonio@um.es

Erin Sanders University of California, Los Angeles erinsl@microbio.ucla.edu

**Outi Savolainen** University of Oulu outi.savolainen@oulu.fi

Wendy Schackwitz DOE Joint Genome Institute wsschackwitz@lbl.gov

Henrik Scheller Lawrence Berkeley National Lab hscheller@lbl.gov

Monika Schmoll Vienna University of Technology monika.schmoll@tuwien.ac.at

Jeremy Schmutz DOE JGI, HudsonAlpha jschmutz@hudsonalpha.org

Christopher Scholin Monterey Bay Aquarium Res. Institute scholin@mbari.org Lori Scott Augustana College loriscott@augustana.edu

Kathleen Scott University of South Florida kscott@cas.usf.edu

Erin Scully Pennsylvania State University eds14@psu.edu

Alexander Sczyrba DOE Joint Genome Institute asczyrba@lbl.gov

Joao Setubal Virginia Bioinformatics Institute setubal@vbi.vt.edu

Maria Shin Eureka Genomics maria@eurekagenomics.com

Bryan Siepert Joint Genome Institute bsiepert@lbl.gov

Stefan Sievert Woods Hole Oceanographic Institution ssievert@whoi.edu

Pamela Silver Harvard Medical School pamela\_silver@hms.harvard.edu

Shaneka Simmons Jackson State University angelique.c.lee@jsums.edu shaneka.s.simmons@gmail.com

Kanwar Singh DOE Joint Genome Institute ksingh@lbl.gov

Steven Slater Great Lakes Bioenergy Res. Center scslater@glbrc.wisc.edu

Tatyana Smirnova DOE Joint Genome Institute tsmirnova@lbl.gov

Mariya Smit Oregon Health and Science University smitm@ebs.ogi.edu

Katherine Smith USDA Forest Service smithk@ufl.edu

Charlotte Smith University of California, Berkeley charlottesmith@berkeley.edu

Ayme Spor Cornell University ayspor@gmail.com

Shawn Stricklin Monsanto shawn.lynn.stricklin@monsanto.com Garret Suen University of Wisconsin, Madison gsuen@wisc.edu

Ernest Szeto Lawrence Berkeley National Lab, JGI eszeto@lbl.gov

Eric Tang DOE Joint Genome Institute etang@lbl.gov

Neslihan Tas Lawrence Berkeley National Lab ntas@lbl.gov

Steven Theg University of California, Davis smtheg@plb.ucdavis.edu

Michael Thomashow Michigan State University thomash6@msu.edu

Hope Tice DOE Joint Genome Institute HNTice@lbl.gov

James Tiedje Michigan State University tiedjej@msu.edu

Emilie Tisserant INRA (French National Institute for Agricultural Research) tisseran@nancy.inra.fr

Lars Tomanek California Polytechnic State University Itomanek@calpoly.edu

Tamas Torok Lawrence Berkeley National Lab ttorok@lbl.gov

Trudy Torto-Alalibo Virginia Bioinformatics Institute trutot@gmail.com

Susannah Tringe DOE Joint Genome Institute sgtringe@lbl.gov

Wendy Trzyna Marshall University trzyna@marshall.edu

Adrian Tsang Concordia University tsang@gene.concordia.ca

Jerry Tuskan Oak Ridge National Laboratory tuskanga@ornl.gov

Ludmila Tyler University of California, Berkeley ltyler@berkeley.edu

Michael Udvardi The Samuel Roberts Noble Foundation mudvardi@noble.org

George Vacek Convey Computer gvacek@conveycomputer.com

George VanDegrift Convey Computer gvandegrift@conveycomputer.com

Tracy Vence Genome Technology magazine tvence@genomeweb.com

John Vogel USDA-ARS Western Reg. Res.Center john.vogel@ars.usda.gov

Alexey Vorobev University of Washington voralex@u.washington.edu

Mark Waldrop US Geological Survey mwaldrop@usgs.gov

Mingyi Wang The Samuel Roberts Noble Foundation mwang@noble.org

Shanquan Wang National University of Singapore wangshanquan1982@gmail.com

Zhong Wang DOE Joint Genome Institute zhongwang@lbl.gov

Sarah Watkinson University of Oxford sarah.watkinson@plants.ox.ac.uk Chia-Lin Wei DOE Joint Genome Institute cwei@lbl.gov

Elizabeth Wilbanks University of California, Davis egwilbanks@ucdavis.edu

Man Chun Wong The Chinese University of Hong Kong manchun3@yahoo.com.hk

**Derek Wood** Seattle Pacific University woodd1@spu.edu

Scott Woody University of Wisconsin-Madison swoody@wisc.edu

Tanja Woyke DOE Joint Genome Institute twoyke@lbl.gov

Crystal Wright Joint Genome Institute cawright@lbl.gov

Jody Wright University of British Columbia jjwright49@me.com

Cindy Wu Lawrence Berkeley National Lab chwu@lbl.gov Guohong Wu DOE Joint Genome Institute gwu@lbl.gov

Dongying Wu UC Davis Genome Center dongyingwu1@gmail.com

Hugh Young USDA, WRRC, GGD hugh.young@ars.usda.gov

Scott Yourstone University of North Carolina islandhopper81@msn.com

Pablo Zamora University of California, Davis, PIPRA pazamora@ucdavis.edu

Matthew Zane DOE Joint Genome Institute mczane@lbl.gov

Xueling Zhao DOE Joint Genome Institute xzhao@lbl.gov

Kemin Zhou DOE Joint Genome Institute kzhou@lbl.gov

Natasha Zvenigorodsky DOE Joint Genome Institute nzvenigorodsky@lbl.gov

# Author Index

Au, C.H9
Aanen, Duur62
Ackerman, Eric51
Adams, Aaron S62
Adams, Michael W.W23
Adams, Paul16, 22
Adams, Sandra M9, 62
Aerts, Andrea73
Allen, Michael S59
Alström, Sadhna46
Amasino, Rick70
Anderson, Claire60
Anderson, Olin11
Andersson, Björn46
Angenent, Largus T68
Aranda, M12
Arkin, Adam P52
Arp, Daniel J 10, 47
Au, C.H
Auer, Manfred53
Auvinen, Petri18
Aylawrd, Frank O9, 57, 61, 62
Baek, Kyung-Hwa63
Balasch, Monica Moya33
Barry, Kerrie W 57, 62
Bartley, Laura E72
Bayer, T
Beam, J
Bebout, Brad M 69
Beck, D.A.C
Beers, Allen R68
Benke, Peter13
Benner, J2
Berka, Randy51
Bernstein, H28
Berry, Kerrie W9
Bhatia, Karan10
Biedermann, Peter H.W62
Blow, Matt73

Blumer-Schuette, Sara	23
Bohannan, Brendan J	. 63
Bollmann, Annette 10	, 47
Boonmee, Atcha	. 29
Borevitz, Justin	45
Borglin, Sharon	. 15
Bork, Peer	1
Boyum, Julie33	, 61
Bragg, Jennifer	. 11
Brenner, Steven E	. 18
Brettin, Thomas	. 11
Brewer, Tony	. 59
Brodie, Eoin L	42
Brown, C. Titus	7
Brown, R	28
Brumm, Phillip J	, 61
Buckler, Edward	1
Bulsara, N	20
Burns, C	9
Burnum, Kristin E	9
Burow, Luke C	. 69
Buschiazzo, E	. 12
Byrne-Bailey, Kathryne G	. 18
Canlas, Patrick	13
Canon, Shane	59
Cantarel, Brandi	. 35
Carlson, Hans K.	42
Carlson, John	. 58
Carlson, R	28
Cass, Cynthia	66
Chain, Patrick7	, 40
Chair, Antinea H.	. 18
Chang, Jan-Fang	21
Chavarría, Krystle	27
Chen, Feng24	, 73
Cheng, C.K.	9
Cheng, Jan-Fang35	, 53
Chertkov, Olga14	, 61
Cheung, M.K.	. 35

Chisholm, Stephen 26
Chistoserdova, L 66
Choi, Cindy 39
Chokhawala, Harshal 24
Chou, Wen-Chi72
Christensen, Ulla13
Chum, W.W.Y
Clark, Douglas S 24
Clingenpeel, Scott 35
Coates, John D 18, 42
Cohly, Hari H.P 59
Collins, Kirby 59
Collins, Scott
Constantin, T 19, 20
Copeland, Alex 13, 36, 59
Cottingham, Robert 11, 23
Coutinho, Pedro M 16
Cummings, Theresa A 68
Cunic, Craig 11
Currie, Cameron R.9, 57, 62, 61
Cushman, J.C 51
D'haeseleer, Patrik 16
D'Jamoos, Mike 59
da C. Jesus, Ederson 63
da S. Paula, Fabiana 63
Daligault, Hajnalka 14
Dangl, Jeffrey
Darling, A 65
Darzi, Youseff35
Daum, Chris 13
Davenport, Karen 14, 30
Davis, Jennifer R14
Davis, John M 60
de Vries, Ronald P 16
Dean, Ralph A 51
DeAngelis, Kristen M 15
Dehal, Paramvir S 52
del Rio, Tijana Glavina 7, 27
Delwiche, Charles 45

#### Authors

Deneke, Jan33, 61
DeSalvo, M.K12
Detter, Chris14, 51
Distel, Dan L2
Dixon, Richard72
Dodsworth, Jeremy A46
Dominguez-Bello, Maria Gloria 20
Donohue, Timothy J 32
Dougherty, Michael J16
Drell, Persis2
Drinkwater, Colleen 33, 61
Du, Changbin17
Dubinsky, Eric A40
Easterday, Ray11
Edwards, Elizabeth A27
Egan, Rob24, 73
Eichorst, Stephanie A
Eisen, Jonathan A53, 65
Ekojärvi, Jaana18, 49
Engelbrektson, Anna L18
Erickson, Alison35
Evans, R. David33
Facciotti, M65
Feigl, Brigitte J63
Fernández, Elena55
Finlay, Roger46
Fisher-Ramos, C51
Fofanov, V.Y19, 20
Fofanov, Y20
Fortney, Julian15, 40
Fraser-Liggett, Claire35
Fredrickson, J
Frock, Andrew23
Froula, Jeff17
Fulbright, Scott26
Fung, J2
Gallegos-Graves, La Verne 33
Garcia, Marcelo L68
Garcia-Amado, Maria A20
Ge, Yinbing67

Geib, Scott58
Ghiban, Cornel25
Gille, Sascha
Godiska, Ronald33
Godoy-Vitorino, Filipa20
Goodwin, Lynne14
Goodwin, Lynne A9, 10, 30, 47, 61, 62
Gowda, Krishne33
Grabowski, Paul45
Grigoriev, Igor73
Grimwood, Jane72
Gross, Stephen21, 39
Gu, Yong11
Guy, Jessica46
Hack, Christopher A21
Hadi, Masood 16, 22, 48
Halfvarson, Jonas
Hallam, Steven 43, 53, 71
Hambright, W. Sealy 29
Han, Cliff14
Han, James 13, 36, 40, 59
Han, Shunsheng30
Hansen, Sara Fasmer
Harholt, Jesper13
Hatakka, Annele37
Hauser, Loren J23
Hawley, Alyse K71
Hazen, Terry3, 15, 40
He, Ji
He, Jianzhong 67, 71
He, Shaomei25
Heazlewood, Joshua L48
Hedlund, Brian P 46, 53
Henrissat, Bernard2, 16
Hess, Matthias 17, 24
Hettich, Robert35
Hiibel, S.R
Hilgert, Uwe
Hillson, Nathan J55
Hodges, Scott

Hoehler, Tori M 69
Hoeprich, Paul 42
Högberg, Nils 25, 46
Holmes, Susan27
Hoover, Cindi73
Hoover, Kelli 58
Howard, J 19
Howe, Adina7
Htwe, Soe M 38
Huang, Q.L
Huang, Shi26
Hug, Laura A27
Hugenholtz, Philip 20, 53
Hultman, Jenni 27, 40
Hungate, Bruce A
Huntemann, M 41
Hutchinson, Miriam 51
Imam, Saheed
Inskeep, William P 28, 53, 54
Isanapong, Jantiya 29
Isokpehi, Raphael D 59
Ivanova, Natalia N 41, 61
Izquierdo, Javier A 30
Jackson, Robert B 33
Jansson, Janet 7, 27, 35, 40
Jay, Z28
Jennings, R 28
Jeong, Eun-Sook
Joergensen, Bodil 13
Joseph, F 51
Joshi, Hiren J 48
Juenger, Tom3
Kalyuzhnaya, M.G 66
Karp, Peter D31
Kataeva, Ira 23
Keasling, Jay D 55
Keller, Kimberly L 53
Kelly, Robert M 23
Kerfeld, Cheryl A 17
Khalfan, Mohammed 25
Kim, Tae-Wan24

Kirton, Edward51
Klotz, Martin G 10, 47
Knierim, Bernhard53
Knight, Rob4, 68
Knights, Dan68
Knox, J. Paul38
Kohler, A25
Kontur, Wayne S
Koshinsky, H 19, 20
Kourtz, LauraLynn33
Kozubal, M28
Kuo, Alan73
Kuske, Cheryl R33
Kwan, H.S9, 34, 69
Kwok, I.S.W
Kyrpides, N.C7, 41, 49
Laanbroek, Hendrikus J10, 47
Laine, Pia K 18, 49
Lamendella, Regina 7, 27, 34, 40
Lamkin, E2
Lammers, Peter26
Lanzatella, Christina L72
Larsen, D65
Latendresse, Mario31
Lazo, Gerard11
Lebeis, Sarah
Lee, J2
Lee, Janey
Lee, Yi-Ching72
Lemos, M.S51
Ley, Ruth4
Li, L
Li, Mingkun 13, 36
Lindquist, Erika
Liolios, K49
Lipton, Mary S9, 28
Lipzen, Anna40
Liu, J19, 20
Liu, Kuan-Liang
Liu, Wen-Tso53

Loque, Dominique
Lucas, Susan13
Lundberg, Derek S 36
Lundell, Taina 18, 37, 49
Luning, Eric G 52
Luo, Shujun24
Luo, X69
Lynd, Lee R30
Mabery, Shalini
MacDonald, Jacqueline
Mackelprang, Rachel7, 27, 40
Mackie, Roderick
Magnuson, Jon51
Mäkelä, Miia 18, 37, 49
Malfatti, Stephanie 7, 20, 35, 53, 62
Malmstrom, Rex2, 27, 35
Manabe, Yuzuki
Marshall, Ian P69
Martin, F25
Martin, Jeffrey 39, 43
Martin, Joel40, 73
Martínez, Ángel T 55
Martinez, Diego A51
Martínez, María Jesús55
Martinez, Sarah 59
Mason, Olivia U 27, 40
Master, Emma
Mavromatis, K41
Mayali, Xavier42
McAndrew, Ryan22
McInerney, Peter 22, 48, 55
McMurdie, Paul J27
Mead, David 33, 61
Medina, M
Megonigal, J. Patrick33
Meincke, Linda14
Melnyk, Ryan A 42
Meng, Xiandong43
Mewis, Keith 43
Micklos, David25

Mikhailova, Natalia61
Miller, John 44
Mirza, Babur 63
Moeller, Joseph A 57, 62
Moon-Lim, Eunice 18
Moran, Mary Ann 4
Morris, Geoffrey P 45
Morrison, Stephanie N 48
Morsy, Mustafa R 45
Mukhopadhyay, Aindrila 52, 53
Mukhopadhyay, Biswarup 64
Muller, Rebecca 63
Murali, T.M64
Murugapiran, Senthil Kumar 46
Nafisi, Majse 38
Natvig, Don51
Nelson, C. Dana 60
Neupane, Saraswoti 46
Nicora, Carrie D9
Nielsen, Agnieszka Zygadlo . 13
Nielsen, Susanne 10
Noguera, Daniel R 32
Nong, W.Y
Nordborg, Magnus 4
Norton, Jeanette M 10, 47
Nüsslein, Klaus63
O'Neil, C2
Odenbach, Kylea51
Oikawa, Ai 48
Oksanen, Ilona 18, 37, 49
Orfila, Caroline
Ozdemir, Inci23
Pagani, I 49
Paredes, Sur, Herrera
Pati, A 41
Paulin, Lars 18, 49
Pauly, Markus 38
Peacock, Joseph 46
Pellizari, Vivian H63
Pennacchio, Christa72

#### Authors

Pennacchio, Len24, 35, 40, 73
Pérez, Gúmer 50, 56
Perry, Leslie M59
Pett-Ridge, Jennifer42, 69
Pisabarro, Antonio G 50, 56
Polle, J.E.W51
Poole, Farris23
Porras-Alfaro, Andrea 33, 51
Poulsen, Michael62
Powell, Amy Jo51
Pratap, Abhishek59
Price, Morgan N52
Prufert-Bebout, Leslie69
Pukkila, P.J9
Purvine, Samuel O9
Qin, J34
Quest, Daniel 11, 23
Raes, Jeroen
Raffa, Kenneth G62
Rajeev, Lara52
Ramírez, Lucía
Ray, Jayashree53
Redfern, Joanna51
Rennie, Emilie A
Repo, Suzanna18
Rinke, Christian35, 53
Robert, Vincent16
Roberto, Frank F54
Robinson, Gene5
Rodrigues, Jorge L.M63, 29
Romine, M28
Ronald, Pamela13, 72
Roossinck, Marilyn J45
Rosengarten, Rafael D55
Rubin, Edward M 25, 24, 40
Ruiz-Dueñas, Francisco Javier 55
Rusch, D27
Saha, Malay72
Sakuragi, Yumiko38, 48
Salamov, Asaf73

Salmeen, Annette17
Samburova, V51
Santoyo, Francisco 50, 56
Scalfone, Nicholas B 68
Schackwitz, Wendy40
Schadt, Christopher W
Scheller, Henrik Vibe 13, 22, 38, 48
Schmutz, Jeremy72
Scholin, Christopher5
Schroth, Gary24
Schuster, Stephen5
Scott, Jarrod J 9, 57, 62
Scully, Erin58
Sczyrba, Alex 24, 35, 53, 59
Sedbrook, John 66
Sello, Jason K 15
Setubal, Joao C64
Shah, Manesh34
Sharp, K2
Shen, Hui72
Shin, M 19, 20
Shyr, Capser71
Sievert, Stefan M53
Silver, Pamela A6
Silver, Whendee15
Simmons, Blake 16, 51
Simmons, Shaneka S 59
Singer, Steven W69
Singh, Kanwar73
Sinsabaugh, Robert51
Smith, Jason A 60
Smith, Katherine E 60
Smith, Richard D9
Smith, Samual 68
Søgaard, Casper Nicholas 48
Song, Young C71
Soper, David59
Spero, Melanie A32
Spormann, Alfred M 27, 69
Starrett, Gabriel J9

Stein, Lisa Y 10, 47
Steinmetz, Eric 33
Steinwand, Michael A 66
Stenlid, J25
Stepanauskas, Ramunas 35, 53
Stevenson, David M 61
Suen, Garret 9, 57, 61, 62
Sunagawa, S 12
Suwa, Yuichi 10, 47
Tarver, Angela 13
Taupp, Marcus 43
Thomashow, Michael F6
Ticknor, Lawrence O 33
Tiedje, James M 7, 63
Tien, Ming 58
Tighe, Damon35
Tobias, Christian M72
Tom, Lauren M 40
Tomanek, Lars 64
Torres-Jerez, Ivone72
Tortell, Philippe71
Torto-Alalibo, Trudy64
Tran, Huu M 55
Tringe, Susannah 2, 7, 9, 20, 23, 24, 28, 40, 46, 51, 57, 62, 71
Tritt, A
Tsai, Siu M63
Tsiamis, George 53
Tuskan, Jerry
Tyler, Brett M 64
Tyler, Ludmila 66
Udvardi, Michael 72
Ulvskov, Peter 13
Vacek, George 59
VerBerkmoes, Nathan 35
Verhertbruggen, Yves
Vilgalys, Rytas 34
Visel, Axel 21, 24
Vogel, John 11, 66
Voolstra, C.R12
Vorobev, A

### Authors

Waldrop, Mark23, 27
Wall, Judy D53
Waller, Alison S27
Wang, Mingyi67, 72
Wang, Shanquan67
Wang, Zhong 10, 17, 24, 39, 43, 59, 73
Wardle, Greg26
Weber, Carolyn F34
Weber, Peter K42, 69
Wei, Chia-Lin
Weimer, Paul J61
Werner, Jeffrey J68
Wiebenga, A16
Wilke, S.K9
Williams, Jason25
Williams, P41

Windham-Myers, Lisamarie23
Woebken, Dagmar69
Wong, K.S
Wong, M.C35, 69
Woody, Scott T70
Woyke, Tanja 2, 14, 23, 24, 27, 30, 35, 40, 46, 53
Wright, Jody J71
Wrighton, Kelly C42
Wu, Jiajie11
Xie, Gary
Xin, Fengxue71
Xing, L69
Xu, Jian26
Xu, Ying72
Yang, Sung-Jae23
Yarasheski, Kevin68

Yates, Tracy 26
Yelick, Kathy8
Yin, Yanbin72
Yip, P.Y35
Young, Hugh A 72
Yuhong, Tang72
Zak, Donald R 34
Zane, Matt 13
Zeng, Xiaowei 26
Zhang, Ji-Yi 72
Zhang, Tao 18, 24, 39, 73
Zhao, Zhiying Jean73
Zhou, Kemin73
Ziegelhoffer, Eva C 32
Zielinska, B 51
Zolan, M.E9

Notes