Seventh Annual
DOE Joint Genome Institute
User Meeting

Sponsored By

U.S. Department of Energy
Office of Science

March 20-22, 2012
Walnut Creek Marriott
Walnut Creek, California

Contents

Speaker Presentations 1
Poster Presentations 13
Attendees 83
Author Index 89

Speaker Presentations

Abstracts alphabetical by speaker

DOE Systems Biology Knowledgebase (Kbase)

Adam Arkin (APArkin@lbl.gov)

Lawrence Berkeley National Laboratory, Berkeley, California

The Genome of Selaginella, a Remnant of an Ancient Vascular Plant Lineage

Jody Banks (banksj@purdue.edu)

Botany and Plant Pathology, Purdue University, West Lafayette, Indiana

Plants with lignified vascular tissues first appeared on earth about 400MY ago and subsequently diverged into several lineages. Only two of them remain: the euphyllophytes, which includes the ferns, gymnosperms and angiosperms, and the lycophytes. The genome sequence of the lycophyte *Selaginella moellendorffii* described here is the first lycophyte genome sequenced. Its compact genome, about two-thirds the size of Arabidopsis, has fewer genes with small intergenic regions and introns and no evidence of polyploidy. By comparing the Selaginella proteome with those of earlier diverging plants (Chlamydomonas and the moss Physcomitrella) and later diverging angiosperms, we were able to identify genes that coincide with the evolution of traits specific to land plants. Among these traits are vascular tissues consisting of special lignified cell types. Surprisingly, Selaginella produces an angiosperm-specific lignin. Recent studies indicate that the Selaginella lignin biosynthetic genes may be useful in modifying lignin in angiosperms.

Genomics of Energy and the Environment

Steven A. Benner (sbenner@ffame.org)

Foundation for Applied Molecular Evolution, The Westheimer Institute of Science and Technology, Gainesville, Florida

The Earth and its biosphere co-evolve in tandem, each having influenced the other over the 4.5 billion year history of the planet. Genomic sequence data provide an important resource to explore and understand this co-evolution , especially if experimental methods are used to supplement theoretical modeling based on comparative sequence analysis. Paleogenetics provides experimental tools to address historical models based on genomic sequence analysis. Paleogenetics infers the sequences of ancestral genes and proteins from now-extinct organisms by analysis of the sequences of their descendents. Then, paleogenetics exploits recombinant DNA technology to bring these ancient biomolecules back to life, where they can be studied in the laboratory. This talk will describe the use of

experimental paleogenetics to correlate the genomic, paleontological, and geological records of life on Earth. By resurrecting ancestral proteins from extinct organisms that lived long in the past, we can make broad statements about the chemistry behind adaptation, the nature of ancient environments, and the interaction between species in the ecosystem. We will start in the present day, and take steps back in time, first by 40 million years to the start of the most recent global climate deterioration, then back 100 million years to the age of the dinosaurs, and then back over 2 billion years, to the establishment of eubacteria on the planet. We will also discuss how this process if hindered by errors in modern genome sequence databases. Further, we will discuss the construction of "naturally organized" genome sequence databases, such as the MasterCatalog, which allow efficient organization, search, error correction and analysis, especially when compared with standard public genome sequence database.

Getting to the Root of Things: Root Spatiotemporal Regulatory Networks

Siobhan Brady (sbrady@ucdavis.edu)

Genome Center, Department of Biology, College of Biological Sciences, University of California, Davis, Davis, California

Plant root development provides a remarkably tractable system to delineate developmental gene regulatory networks and to study their functionality in a complex multicellular model system over developmental time. We present a gene regulatory network that regulates distinct transcriptional events in developmental time. Distinct regulatory modules were identified that temporally drive the expression of genes involved in xylem specification and in the subsequent synthesis of secondary cell wall metabolites associated with xylem differentiation.

Reprogramming Bacteria to Seek and Destroy Small Molecules

Justin P. Gallivan (justin.gallivan@emory.edu)

Department of Chemistry and Center for Fundamental and Applied Molecular Evolution, Emory University, Atlanta, Georgia

Simple organisms, such as the bacterium *E. coli*., carry out a wide variety of complex functions. *E. coli* cells synthesize complex molecules, communicate with one another, move in response to changing conditions, and replicate themselves every 20 minutes. The programs that control these behaviors are stored in a genome that encodes just over a megabyte of digital information.

In this talk, I will present our recent efforts to reprogram *E. coli* to sense new small molecules and to respond to them with predictable behaviors. Specifically, I will describe our efforts to create synthetic riboswitches, which are designer RNA sequences that control gene expression in a ligand-dependent fashion without the need for proteins. I will show how synthetic riboswitches can be used to engineer bacteria to have a variety of functions, including the ability to seek and destroy small molecules, such as the herbicide atrazine.

The Evolution of Streamlined Genomes in Ocean Bacteria

Stephen J. Giovannoni* (Steve.giovannoni@oregonstate.edu) and J. Cameron Thrash

Department of Microbiology, Oregon State University, Corvallis, Oregon

The smallest genomes known from free-living cells are found in marine bacterioplankton. Genome sequences from cyanobacteria in the genus *Prochlorococcus* range in size from 1.6 - 2.7 Mbp. Genomes from heterotrophic cells in the SAR11 clade of *Alphaproteobacteria* are 1.3 - 1.5 Mbp. Genome sequences from obligate methylotrophs of the OM43 clade of *Betaproteobacteria* are 1.3 Mbp. Thus, in three metabolic categories, marine bacterioplankton hold the record for the smallest free-living genomes. Genome streamlining theory has been invoked to explain the small genomes of these cells. The essence of this theory is that selection is most efficient in microbial populations with large effective population sizes, and favors minimalism in the genomes and cell architecture of bacterioplankton because of selection for the efficient use of nutrient resources. Genome reduction also occurs in bacterial symbionts, where it has been attributed to genetic drift, and produces very different genomic signatures, including the expansion of non-coding genetic material, loss of anapleurotic pathways, and elevated rates of non-synonymous substitution. Comparative study of SAR11 genomes and experimental studies with cultures have revealed the metabolic consequences of genome streamlining. Most strains are deficient in assimilatory sulphate reduction and in normal pathways of glycine biosynthesis, making them dependent on organosulphur compounds and glycine, or glycine precursors, for growth. Many common regulatory systems are absent, and are replaced by simpler systems for maintaining cellular homeostasis, often involving riboswitches. Studies of ultrastructure and the metaproteomics of cells from oligotrophic oceans show that SAR11 have high surface-to-volume ratios and very high ratios of transport proteins, an apparent adaptation to enable efficient replication in ocean “deserts”. These observations support the broad conclusion that metabolic versatility has been sacrificed for simplicity and genome reduction in some bacterioplankton, rendering them able to use ambient nutrient resources efficiently but reducing their versatility. The question remains, how does the evolutionary history and ecology of these organisms differ from microbial plankton with genomes of average size?

Systems Biology Approaches to Dissecting Plant Cell Wall Deconstruction in a Model Filamentous Fungus

N. Louise Glass* (lglass@berkeley.edu), Sam Coradetti, Elizabeth Znameroski, Jianping Sun, James Craig, and Yi Yiong

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

Neurospora crassa colonizes burnt grasslands in the wild and metabolizes both cellulose and hemicellulose from plant cell walls. When switched from a favored carbon source such as sucrose to cellulose, *N. crassa* dramatically upregulates expression and secretion of a wide variety of genes encoding lignocellulolytic enzymes. However, the means by which *N. crassa* and other filamentous fungi sense the presence of cellulose in the environment remains unclear. In *N. crassa*, cellobiose efficiently induces cellulase gene expression in the absence of intra and

extracellular β-glucosidase activity. Using next generation RNA sequencing, we characterized the transcriptional response of *N. crassa* to cellobiose and cellulosic biomass. To identify transcription factors required for plant cell wall deconstruction, we screened a set of 269 deletion mutants in *N. crassa* for predicted transcription factors. From this screen, we identified two uncharacterized zinc binuclear cluster transcription factors (*cdr-1* and *cdr-2*) that are required for utilization of Avicel and cellulase enzyme production, one transcription factor (*xlr-1*) required for utilization of hemicellulose and one transcription factor involved in carbon catabolite repression (*cre-1*). Further manipulation of this control system in industrial production strains may significantly improve production of cellulases for cellulosic biofuel production.

Applications of Genome-based Science in Shaping the Future of the World's Citrus Industries

Fred G. Gmitter Jr. (fgmitter@ufl.edu)

University of Florida Citrus Research and Education Center, Lake Alfred, Florida

Citrus is among the most widely grown and economically significant tree fruit crops produced in the world. Many kinds of fruit, including oranges, mandarins, pummelos, grapefruit, lemons, limes, and others are known and consumed throughout much of the world as delicious and health-promoting fresh fruit. In addition, processed citrus products such as orange juice are, in fact, globally traded commodities. From their areas of origin, principally in southern and eastern Asia, citrus species and cultivars have been disbursed widely throughout the world over at least the past two millennia. In more recent times, however, the more rapid movement of human populations and cultures, along with increased globalization and travel, has seen the accelerated spread of a multitude of citrus pests and diseases, vectored in many cases by unsuspecting human travelers. Within the past 100 years, some of the most devastating diseases of citrus have found their way to nearly all the world's growing regions. The most serious of these is a disease called Huanglongbing (HLB), presumably caused by a psyllid-transmitted, phloem-limited bacterium; there are no known sources of genetic resistance to this disease. HLB, sometimes known as "greening", is widespread through much of Asia, some regions within Africa, and has been spreading very rapidly throughout the Americas in the past decade, threatening in an unprecedented fashion the viability of many of the world's most significant citrus industries. Diseases and other production-based issues aside, simple market competition for citrus products with nearly year-round available fresh and processed fruit products, as well as other food and beverage options, presents a daunting challenge to citrus industries throughout the world.

The International Citrus Genome Consortium was organized to develop globally available genomic resources, as tools for citrus scientists to address many of the current and future challenges facing these important fruit crop industries. Several citrus genomes now have been sequenced, and the information contained therein is beginning to be utilized as solutions are sought for the urgent disease threats, and the multitude of other issues to be addressed, so that citrus production can remain

economically feasible and environmentally sustainable in the coming future. An overview of the current status of citrus genomic resources will be provided, and the challenges of HLB and other diseases will be described along with some genomic-based efforts aimed at developing plants resistant or tolerant to this scourge. Finally, projects aimed at improving not only the quality of citrus fruit and processed products, but their health-promoting attributes as well, will be highlighted.

Using Genomics to Dissect Seed Development

Robert Goldberg (bobg@ucla.edu)

University of California, Los Angeles, Los Angeles, California

During the next 50 years, we will need to produce more food than in the entire history of humankind on a decreasing amount of land for agriculture. A major challenge for the 21st century, therefore, is to increase the yields of major crops, such as soybean, using state-of-the-art genetic engineering and genomic technologies. One way to accomplish this task is to understand all of the genes required to "make a seed" in order to engineer plants for yield traits such as more seeds, bigger seeds, and seeds with improved nutritional composition. My laboratory has been investigating gene activity during seed development in order to identify the genes and regulatory networks required to program seed development. In this lecture, I will discuss transcriptome profiling experiments with mRNAs captured using laser capture microdissection (LCM) from *every* soybean seed compartment (e.g., embryo, endosperm, seed coat), region (embryo proper, suspensor), and tissue (e.g., inner integument, endothelium, seed coat epidermis) throughout seed development – from fertilization through dormancy. In addition, I will discuss complementary experiments that profile seed methylomes, and provide new insight into the epigenetic changes that occur at specific seed developmental stages.

Evolutionary Perspectives on Diversity of Lignocellulose Decay Mechanisms in Basidiomycetes

David Hibbett (DHibbett@clarku.edu)

Clark University, Worcester, Massachusetts

Mushroom forming Fungi (Agaricomycetes) have huge impacts on the carbon cycle of forest ecosystems, through their activities as wood and leaf litter decayers, timber pathogens, and ectomycorrhizal (ECM) symbionts with trees such as pines, oaks, and dipterocarps. Major shifts in carbon nutrition in Agaricomycetes include transitions between saprotrophs, which degrade complex lignocellulosic substrates, and ECM species, which obtain carbon principally as plant-derived sugars. Within saprotrophs, white rot species degrade all components of plant cell walls, including the realcitrant lignin fraction, whereas brown rot species have evolved mechanisms to degrade cellulose while leaving lignin largely intact. Recent large-scale genome sequencing projects supported by the JGI are making it possible to understand not

only the mechanisms but also the evolutionary history of carbon nutrition strategies in Agaricomycetes. An analysis of twenty phylogenetically diverse saprotrophs suggested that the ancestor of the Agrciomycetes was a white rot fungus with multiple lignin-degrading class II fungal peroxidases (PODs). Molecular clock analyses suggested that the origin of white rot roughly coincided with the sharp decrease in the rate of organic carbon burial at the end of the Permo-Carboniferous period. Inclusion of one ECM Agaricomycete in that study suggested that transitions to the symbiotic habit were associated with the loss of many genes encoding decay-related enzymes. Currently, we are analyzing a set of additional ECM genomes to develop a more complete understanding of the pattern, mechanisms and timing of evolution of the ECM habit and its relationship to plant evolution.

Genomic Analysis of Natural Variation for Seed and Plant Size in Maize

Shawn Kaeppler (smkaeppl@wisc.edu)

Department of Agronomy, University of Wisconsin-Madison, Madison, Wisconsin

Crop productivity is a function of basic component traits. Grain yield in maize is determined by the product of the number of ears per hectare, the number of seeds per ear, and seed weight. Stover yield is a function of components including node number, internode length, stalk diameter, and leaf shape and number. We are using sequence-based expression and genotyping in structured populations, collections of diverse lines, and long-term selection populations to characterize genes and alleles underlying natural variation for productivity traits in maize used for food, feed, fiber, and raw materials such as for biofuel. As an example of the approaches that we are using, and as a basis to discusses synergies and challenges of various technologies, I will describe interpretations based on phenotypic and genetic analysis of seed size. Our analyses to date are consistent with 1) a significant pollinator effect on seed size, 2) an important role for the maternal plant in determining seed weight and synchronizing components of development, and 3) pleiotropic effects of some genes on overall plant and seed size.

The Sunflower Genome and Its Evolution

Loren Rieseberg (loren.rieseberg@botany.ubc.ca)

University of British Columbia, Vancouver, British Columbia, Canada

Sunflower (*Helianthus annuus L.*) is a globally important oilseed crop with considerable potential for cellulosic biomass production. Sunflowers have also emerged as an excellent experimental system for studying the ecological genetics of speciation, species boundaries, and hybridization. Here I will describe (1) a draft sequence of the 3.6 Gb sunflower genome, (2) genetic analyses of wood chemistry and biomass related traits, and (3) patterns of genomic divergence in several pairs of wild sunflower species. Results indicate that ultra-high density sequence-based genetic and physical maps offer an effective means for scaffolding contigs obtained

from whole genome shotgun sequencing. Analyses of cellulosic biomass traits provide evidence suggesting that it will be feasible to breed a dual-use sunflower with highly favourable wood composition. Lastly, our evolutionary studies show that plant speciation is surprisingly repeatable at the genome level.

Tapping the Molecular Potential of Microalgae to Produce Biomass

Richard Sayre (rsayre@newmexicoconsortium.org)

Los Alamos National Labs/New Mexico Consortium, Los Alamos, New Mexico

One of the more environmentally sustainable ways to produce energy is the conversion of solar energy into biomass. Plants and algae use solar energy to reduce carbon dioxide to carbohydrates and oils. Biomass conversion to fuels has undergone substantial improvements in the last 20 years. The first-generation biofuels (alcohol and diesel) were/are produced from only a few crop systems. Typically, only a fraction of the solar energy captured and converted into chemical energy (biomass) is harvestable. Inefficiencies in feedstock harvesting and processing further reduce the recoverable energy and reduce net carbon capture. Second-generation biofuel systems including cellulosics are now being developed. Conversion of cellulosics to sugars using advanced enzyme catalysts promises to increase the available reduced carbon resources for fuel production and reduce the land area required for biofuel production. Many second generation biofuel systems do not directly compete with food production, require fewer inputs, and potentially have lower environmental impacts than first-generation biofuels. The third generation of biofuel production systems will be expected to have even lower impact on the environment, greater productivity, greater energy return on investment, and will be directly compatible with the existing energy infra-structure. One of the more attractive third generation biofuel systems under development is algae. Algae grow rapidly, have high oil content (up to 55% oil), and are capable of producing 2-10 times more biomass per unit land area than any terrestrial crop system. In addition, algae can potentially capture CO_2 as bicarbonate in ponds as well as utilize nutrient-rich waste water. Significantly, the single celled algae are also one of the more evolutionary diverse groups of organisms whose biodiversity represents a rich resource for bioprospecting. We will report on progress to optimize biomass productivity from algae using transgenic strategies informed from “omics” and biodiversity surveys.

Entering the Era of Mega-genomics

Michael C. Schatz (mschatz@cshl.edu)

Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

The continuing revolution in DNA sequencing and biological sensor technologies is driving a digital transformation to our approaches for observation, experimentation, and interpretation that form the foundation of modern biology and genomics. Whereas classical experiments were limited to thousands of hand-

collected observations, today's improved sensors allow billions of digital observations and are improving at an exponential rate that exceeds Moore's law. These improvements have made it possible to sequence new genomes and monitor the dynamics of biological processes on an unprecedented "mega-scale," but have brought proportionally greater quantitative and computational requirements.

The growing digital demands have motivated extensive research into computational algorithms and parallel systems for analysis. Recently a great deal of research has been focused on applying emerging scalable computing systems to genomic research. One of the most promising is the Hadoop open-source implementation of MapReduce: it is specifically designed to scale to very large datasets, its intuitive design supports rich parallel algorithms, and is naturally applied to analysis of many biological assays. During my presentation, I will describe some recent innovations using these and other technologies for large-scale genome assembly, variation detection, and transcription analysis. These are promising early results but continued research is essential in the coming years, especially as we hope to model and mine these data to uncover genotype-to-phenotype relations that can only be detected across very large populations.

Designing Synthetic Regulatory RNAs: New Languages for Programming Biological Systems

Christina D. Smolke (christina.smolke@stanford.edu)

Stanford University, Stanford, California

Advances in synthetic biology are transforming our ability to design and build synthetic biological systems. While progress has been made in the design of complex genetic circuits, capabilities for constructing large genetic systems currently surpass our ability to design such systems. This growing 'design gap' has highlighted the need to develop methods that support the generation of new functional biological components and scalable design strategies for complex genetic circuits that will lay the foundation for integrated biological devices and systems.

The vast majority of genetic systems engineered to-date utilize protein-based transcriptional control strategies. However, as the examples of functional RNA molecules playing key roles in the behavior of natural biological systems have grown over the past decade, there has been growing interest in the design and implementation of synthetic counterparts. Researchers are taking advantage of the relative ease with which RNA molecules can be modeled and designed to engineer functional RNA molecules that act as diverse components including sensors, regulators, controllers (ligand-responsive RNA regulators), and scaffolds. These synthetic regulatory RNAs are providing new tools for temporal and spatial control in biological systems.

I will describe recent work in the design of RNA controllers and advances in addressing challenges faced in their broad implementation as user-programmed control systems in living cells. In particular, I will describe the development of high-throughput cell-based screens for rapidly generating synthetic regulatory RNAs with specified quantitative properties. In addition, I will discuss how such

RNA controllers can be implemented as sensitive biosensors providing a noninvasive readout of target metabolites supporting the development of high-throughput screening strategies for enzyme and metabolic pathway engineering. Finally, I will address how the application of synthetic regulatory RNAs as controllers in different biological pathways are leading to the elucidation of integrated systems design strategies and new languages for programming genetic systems.

What if Every Cell Is Different? An Uncharted Territory for Genomics

Ramunas Stepanauskas (rstepanauskas@bigelow.org)

Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, Maine

Single cell genomics is a new, transformative research technology with a rapidly expanding use in areas as diverse as microbial ecology and cancer studies. Robust, high-throughput protocols are being established, resulting in the recovery of genomic DNA from hundreds of thousands of individual microbial cells obtained directly from their natural environment, without the need for cultivation. This game-changing development is starting to fill the huge knowledge gap on the genomic composition, evolutionary histories, metabolic potential, and ecological roles of the microbial “uncultured majority”. One example of the power of single cell genomics is the discovery of chemoautotrophy pathways in abundant, yet uncultured bacteria inhabiting the vast expanse of the dark ocean. In another example, we used single cell genomics to decipher in situ ecological interactions of uncultured protists, discovering their grazing preferences and viral infections. Single cell genomics also offers new opportunities to improve human health and to mine the nearly infinite genetic resources of the uncultured microorganisms for biotechnology products. It also helps us to fully appreciate the extent of biological diversity on our planet. We may find out that almost every cell is unique, calling for a redefinition of what is “known, unknown and unknowable” in genomics.

Ocean Viruses: Tiny Entities with Global Impacts

Matthew B. Sullivan (mbsulli@email.arizona.edu)

Ecology and Evolutionary Biology Department and and Molecular and Cellular Biology Department, University of Arizona, Tucson, Arizona

Ocean cyanobacteria are responsible for about a quarter of global carbon fixation, and ocean viruses are responsible for the largest flux of carbon in the global oceans (150 Gt per year). Cyanobacterial viruses (cyanophages) harbor core photosynthesis genes that are expressed during infection, provide a fitness advantage to the cyanophage, and impact the evolutionary trajectory of these global-distributed host photosystems through recombination. Here I will first discuss brute-force and novel efforts to use these relatively well-studied, ecologically important cyanophage-host systems to ask a simple question, "how many viruses infect a single host in a single seawater sample?" along the coastal-to-open-ocean Line67 oceanographic transect off Monterey California. From there,

I will describe the "core" and "flexible" gene sets displayed across vast ocean viral communities in the first, all-sequence-used, large-scale comparative viral metagenomic analyses to date.

Understanding Historical Human Migration Patterns and Interbreeding Using the Ancient Genomes of a Palaeo-Eskimo and an Aboriginal Australian

Eske Willerslev (ewillerslev@snm.ku.dk)

Centre for GeoGenetics, University of Copenhagen, Copenhagen, Denmark

Understanding of the earlier and more recent histories of modern human populations suffers from the historical extinction or heavy recent admixture of many indigenous populations worldwide. Although museum samples of numerous human populations exist that predate these events, the genetic work on such specimens has largely been limited to short fragments of mitochondrial DNA that cannot be distinguished from more recent contaminant DNA. By using ancient human hair coupled with next generation sequencing we can now bypass these problems, allowing for complete genomes of individuals of extinct populations or individuals predating recent admixture events to be sequenced. Such genomes are currently chancing our understanding of how humans populated our globe.

Omics in the Arctic: Genome-Enabled Contributions to Carbon Cycle Research in High-Latitude Ecosystems

Stan D. Wullschleger (wullschlegsd@ornl.gov)

Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee

A decade or more into the 21st Century and society continues to look to the biological sciences for creative solutions to several pressing concerns; energy security, climate change, and the sustainable use of otherwise limited supplies of Earth resources. Of these, the role of plants and microbial communities in climate change and the technologies required to underpin this science, is particularly compelling. It is one that illustrates not only the highly-coupled nature of basic biological processes that will determine the trajectory of climate in our changing world, but also the molecular-scale approaches that must be employed if we are to merge processes-level knowledge with predictive models.

Gaining a quantitative and predictive understanding of the Earth's climatic system will require that biologists characterize complex interactions between terrestrial ecosystems and the fundamental processes that underlie the global carbon cycle. It is already known that vast amounts of carbon fixed initially by plants and then degraded by microbial communities, enter and leave plant-soil systems. The magnitude of these pools and a first-order understanding of exchange rates are well-known for most ecosystems, but the kinetics and controls on the processes that drive carbon flux into leaves through photosynthesis, the incorporation of that carbon into complex biological compounds, and the turnover and transfer of organic matter into soil carbon pools remain highly uncertain. How best to quantify

and represent these processes in models is also uncertain and will require an ambitious research agenda that links atmospheric science, ecology, biogeochemistry, and knowledge not only of structural, but functional activities encoded in the genomes of plant and microbes communities. This is a formidable challenge, but one that can be met by emerging systems biology research that targets the behavior of plants and microbial communities and their collective contribution to climate at multiple scales.

Many questions need to be addressed as we characterize critical ecosystem-climate feedbacks in poorly understood yet globally-important ecosystems. Among the biomes of the world this is especially true in nitrogen-limited Arctic ecosystems where carbon stored in frozen soil or permafrost is at risk of release to the atmosphere due to warmer temperatures. As permafrost thaws throughout an annual cycle, giving rise to increasing rates of organic matter decomposition and liberating nitrogen as part of the mineralization process, there may also be shifts in plant growth and community composition. The past 50 years have seen a steady increase in plant productivity for the Arctic as shrubs (e.g., *Salix*, *Betula*, *Alnus* spp.) invade the tundra. The mechanisms responsible for this rapid migration of plants remain unresolved.

The Department of Energy through their Office of Science, Biological and Environmental Research (BER) program has recently launched the Next-Generation Ecosystem Experiments (NGEE Arctic) project. The goals of the NGEE Arctic project will be described, as will opportunities that await input from the plant and microbial biologist. In particular, molecular biologists working closely with climate scientists can help bridge the knowledge gap between genome- and global-scale phenomena. Several questions relevant to ecosystem-climate feedbacks in the Arctic as identified. Special attention is given not only to microbial-mediated processes, but also to less appreciated questions for the plant biologist where the emphasis lies on understanding mechanisms of nitrogen acquisition in temperature-limited environments; plant adaptation to extreme climates; and population structure of plant communities along transition zones (i.e., ectones) in boreal and tundra ecosystems. Incorporation of insights derived from these studies into models will also be discussed.

The Genome Beat

Carl Zimmer (carl@carlzimmer.com)

New York Times, New York, New York

Genomes have had a special place in the media long before the first genomes were even sequenced. Later, the simple act of completely sequencing all the DNA in an organism was enough to earn headlines. Today, however, the press's reactions to advances in genomics have become more complex, arguably because the practical results many were expecting from genome sequencing have been slow in coming or difficult to appreciate. This talk will offer one journalist's personal observations of the genome's changing place in the changing world of journalism.

Poster Presentations

Posters alphabetical by first author. *Presenting author

Bacterial Profiling of White Plague Disease in a Comparative Coral Species Framework

Chatchanit Arif, Cornelia Roder, Manuel Aranda, and Christian R. Voolstra* (christian.voolstra@kaust.edu.sa)

Red Sea Research Center, King Abdullah University of Science and Technology (KAUST), Saudi Arabia

Coral disease can heavily impact reef health and has reduced live coral cover by up to 80% in some places. With the advent of novel molecular techniques such as next generation sequencing, we are well aware of the high diversity and specificity of coral-associated bacteria. However, there are only few studies that have looked at changes within the bacterial community of corals as a consequence of disease. Furthermore, most studies have focused on microbial profile shifts of a specific coral species that suffers from a specific disease. Examining coral diseases in a comparative coral species framework might provide insights into a conserved coral microbiome and how it changes in response to disease. Here we analyzed the microbial profiles of two coral species, *Porites lutea* and *Pavona duerdeni*, from Sairee Reef, Koh Tao, in the Gulf of Thailand that displayed typical characteristics of White Plague Disease via 16S PhyloChip assays. PhyloChip assay (SecondGenome) are a microarray-based method that identifies and measures the relative abundance of more than 65,000 individual microbial taxa. With this approach we hope to gain a deeper understanding of the etiology and bacterial footprint of white plague disease by 1) comparing bacterial community changes of healthy and diseased samples within and between coral species, and 2) by identifying key bacterial species that change as a consequence of disease.

KBase: An Integrated Knowledgebase for Predictive Biology and Environmental Research

Adam Arkin,[1] Robert Cottingham,[2] Sergei Maslov,[3] Rick Stevens,[4] Dylan Chivian,[1] Parmavir Dehal,[1] Christopher Henry,[4] Folker Meyer,[4] Jennifer Salazar,[4] Doreen Ware,[5] David Weston,[2] Brian Davison* (davisonbh@ornl.gov),[2] and Elizabeth M. Glass[4]

[1]Lawrence Berkeley National Laboratory, Berkeley, California; [2]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]Brookhaven National Laboratory, Upton, New York; [4]Argonne National Laboratory, Argonne, Illinois; [5]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

The Systems Biology Knowledgebase (KBase) has two central goals. The scientific goal is to produce predictive models, reference datasets and analytical tools and demonstrate their utility in DOE biological research relating to bioenergy, carbon cycle, and the study of subsurface microbial communities. The operational goal is to create the integrated software and hardware infrastructure needed to support the creation, maintenance and use of predictive models and methods in the study of microbes, microbial communities and plants.

KBase is a collaborative effort designed to accelerate our understanding of microbes, microbial communities, and plants. It will be a community-driven, extensible and scalable open-source software framework and application system. KBase will offer free and open access to data models and simulations, enabling scientists and researchers to build new knowledge, test hypotheses, design experiments, and share their findings to accelerate the use of predictive biology. Our immediate 18-month goal is to have a beta-version completed by February 2013.

The KBase microbial science domain will enable the reconciliation of metabolic models with experimental data with the ultimate aim of manipulating microbial function for applications in energy production and remediation The plants science domain will initially target linking genetic variation, phenotypes, molecular profiles, and molecular networks, enabling model-driven phenotype predictions. We will also map plant variability onto metabolic models to create model-driven predictions of phenotypic traits. Our microbial communities team will build the computational infrastructure to research community behavior and build predictive models of community roles in the carbon cycle, other biogeochemical cycles, bioremediation, energy production, and the discovery of useful enzymes.

KBase will be composed of a series of core biological analysis and modeling functions, including an application-programming interface (API) that can be used to connect different software programs within the community. These capabilities will be constructed from the popular analysis systems at each of the KBase sites. Their integration into KBase will combine individual functions to create the next generation of biological models and analysis tools. The KBase API will also enable third-party researchers from our diverse community of users to design new functions. KBase will be supported by a computing infrastructure based on the OpenStack cloud system software, distributed across the core sites.

The success of the KBase project depends not only on producing a large-scale, open computational capability for systems biology research data management and analysis but also on positioning these tools to be used by the community. Our outreach program is designed to target different user groups: data providers, tool builders, and users of both data and tools. …

Genome and Transcriptome Analysis of the Ascomycete *Morchella conica*

Petr Baldrian* (baldrian@biomed.cas.cz), Hana Hršelová, Michaela Urbanová, and Milan Gryndler

Institute of Microbiology of the ASCR, Videnska 1083, 14220, Praha 4, Czech Republic

Genome and in vitro transcriptome of the ascomycete fungus *Morchella conica* (*Morchellaceae*, *Pezizales*) was analysed on a Roche GS Junior sequencer. Assembly was performed using a Roche assembler and Blast2Go and MG-RAST were used for transcript annotation. The predicted size of the genome is 36.2 MB and the sequencing coverage was 4.8-fold. The assembly resulted in a construction of >2917 contigs longer than 2000 bases, out of which 35 were >10000 b long. Transcription of the fungus was analysed after 14 days of growth on a 2% liquid

malt extract medium. In total, 1186 contigs > 1000 bases were identified in the transcriptome. The annotation found most closest hits among the *Tuber melanosporum* proteins (78.5%) and fewer in *Arthrobotrys oligospora* (4.1%) and *Botryotinia fuckeliana* (1.8%). The most expressed sequences were these encoding for the elongation factor 3, polyubiquitin, multiprotein-bridging factor, acyl-desaturase, plasma membrane ATPase, ATP-dependent RNA helicase, translation initiation factor, and actin. Among enzymes, adenosine-triphosphatase, phosphatase, nucleoside-triphosphatase, protein kinase, H+-exporting ATPase, stearoyl-CoA 9-desaturase, chitin synthase, protease, nitrite reductase, and phosphoenolpyruvate carboxykinase were the most expressed. The redox enzymes catalase, laccase, cytochrome c and peroxidase were also transcribed under the conditions of the experiment. The current data show that the *Morechella conica* genome is most similar to this of the *Tuber melanosporum* (*Tuberaceae*, *Pezizales*), although the latter one is much larger (125 MB). The analysis of the *Morchella conica* genome is now being continued by pair-end sequencing on a SOLID platform.

Mining Halophilic and Halotolerant Microbial Communities for the Genetic Correlates of Ionic Liquid Tolerant Lignocellulolytic Activities

Nicholas R. Ballor* (nrballor@lbl.gov),[1] Hannah Woo,[1] Thomas Ruegg,[1] Terry Hazen,[4] Blake Simmons,[1] Steven Singer,[1] and Janet Jansson[1,2,3]

[1]Joint BioEnergy Institute, Berkeley, California; [2]Lawrence Berkeley National Laboratory, Berkeley, California; [3]DOE Joint Genome Institute, Walnut Creek, California; [4]University of Tennessee, Department of Earth and Planetary Sciences, Knoxville, Tennessee

Lignocellulose is the most abundant plant material on earth. It is the objective of next generation biofuel refineries to convert this plant matter into liquid fuels. Lignocellulose presents a formidable challenge to the biofuels industry as a complex composite material recalcitrant to microbial degradation. To date, the most successful means of reducing the recalcitrance of lignocellulose to enzymatic saccharification is through the use of an ionic liquid-based pretreatment process. This presents a challenge for downstream processes because many enzymes and microbes are inhibited by the presence of residual ionic liquids. To reduce operating costs for ionic liquid removal we are seeking out nature's solutions to ionic liquid tolerance. Saline environments may hold the solutions to ionic liquid tolerance coveted by the biofuels industry. We collected samples from salt ponds in the San Francisco South Bay and Cabo Rojo, Puerto Rico and from the highly productive turtle grass (*Thalassia testudinum*) beds on the floor of the Bioluminescent Bay in Puerto Rico. In collaboration with the DOE Joint Genome Institute we are sequencing metagenomes from each of these environments and screening them for candidate ionic liquid tolerant lignocellulolytic activities. In addition, we have established enrichments for halotolerant or halophilic microbial communities utilizing candidate biorefinery feedstocks: miscanthus, eucalyptus, and pine. Analysis of the metagenomes and transcriptomes of these microbial communities that are selected on specific feedstocks relevant to the biofuels industry should yield discoveries of novel ionic liquid tolerant enzymes and

microbes. Moreover, the comparatively low diversity of these communities should facilitate metagenome sequence assemblies. Discoveries made in this work will further empower synthetic biology to engineer solutions paving the way toward the next generation of biofuels.

Identification of Genes Required for Metabolite Utilization in Bacteria Using Mutant Libraries and Metabolomics

Richard Baran* (RBaran@lbl.gov),[1] Benjamin P. Bowen,[1] Morgan N. Price,[2] Adam P. Arkin,[2] Adam M. Deutschbauer,[2] and Trent R. Northen[1]

[1]Life Sciences Division, [2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California

Microorganisms exhibit complex metabolism and metabolic interactions with their environment, large parts of which remain unknown. Deficiencies in functional annotations of microbial genomes as well as incomplete knowledge of small molecule repertoires (metabolomes) of microorganisms limit the understanding of their metabolism. We have used untargeted mass spectrometry-based metabolomics to identify unexpected metabolites in microorganisms as well as to profile uptake and release of a wide array of compounds. *Escherichia coli* K12 and *Shewanella oneidensis* MR-1 were cultured in different complex media and metabolite profiles of spent media were compared to metabolite profile of uncultured control media to identify metabolites utilized by these bacteria. Ten of these metabolites were selected and added to minimal media and over 8000 mutant strains of these bacteria were cultured in this supplemented minimal media in 96-well format. Lower complexity of spent media extracts from these screens (compared to complex media extracts) facilitated high-throughput analysis by mass spectrometry. Presence of one of the tested metabolites in the spent media of a specific mutant indicated a defect in the utilization of the metabolite and pointed to the corresponding gene. This screening led to the identification genes of known enzymes and transport proteins as well as genes with ambiguous annotations to be required for the utilization of tested metabolites. This generally applicable approach can be used to discover novel metabolic capabilities of microorganisms and identify the corresponding genes.

Description of the WRRC *Brachypodium distachyon* T-DNA Collection and the Generation of Homozygous Lines

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

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

During the initial phase of our T-DNA tagging project we generated 8,491 fertile T-DNA lines in the model grass *Brachypodium distachyon* (Brachypodium). Data from 17,637 sequencing reactions for 7,145 of the lines were compared to the Brachypodium genome assembly, and 7,389 sequences (average 195 bases) matched to the Brachypodium genome. The top scoring match from each sequence

was assigned as a flanking sequence tag (FST). FSTs were assigned to insertions sites (ISs) in 4,402 of the lines. These represent 5,285 unique loci. The distribution of ISs across the chromosomes was analyzed by plotting the number of insertions within 500 kb windows. Insertions span the length of all five chromosomes and are generally proportional to chromosome length (average 19-21 IS/Mb), except for Bd4 (16 IS/Mb). The number of ISs is positively correlated with regions of higher gene density and 28% of the ISs reside in genic regions (exons, introns, UTRs) and 21% are within 1 kb of genes. The WRRC T-DNA population contains >2,400 tagged genes and can be viewed at http://wheat.pw.usda.gov/bEST/ and http://www.brachypodium.org/browse. Lines can be ordered through the USDA site.

Metagenome-enabled Investigations of Carbon and Hydrogen Fluxes within the Serpentinite-hosted Subsurface Biosphere

William J. Brazelton* (wbrazelton@gmail.com)[1,2] and Matthew O. Schrenk[1]

[1]NASA Astrobiology Institute, Department of Biology, East Carolina University, Greenville, North Carolina; [2]School of Oceanography, University of Washington, Seattle, Washington

Ultramafic rocks in the Earth's mantle represent a tremendous reservoir of carbon and reducing power. Upon tectonic uplift and exposure to fluid flow, serpentinization of these materials generates copious energy, sustains abiogenic synthesis of organic molecules, and releases hydrogen gas (H_2). Microbial communities hosted within serpentinites may be important mediators of carbon and energy exchange between the deep Earth and the surface biosphere. Actively serpentinizing rocks are present on all of the world's continents, comprise significant portions of the deep seafloor, generate large quantities of geochemical energy, and yet are some of the most poorly understood portions of the biosphere. Our team is involved in a series of ongoing interdisciplinary investigations aimed at defining the global serpentinite microbiome in the context of detailed chemical and physical data.

Our JGI community sequencing project includes complete genome sequencing of cultivated isolates and metagenome and metatranscriptome sequencing of subsurface fluid samples collected from several sites of active serpentinization. The intent of this project is to improve our understanding of both the taxonomic and functional diversity of microbial communities in the serpentinite subsurface. Specifically, we intend to target catabolic processes and microbial interactions with carbon pools (autotrophy, fermentation, methanogenesis, respiration). By studying community genomes in the context of detailed environmental data, we will also be able to resolve physiological adaptations to the serpentinite microbiome, with implications for both culturing approaches and practical applications (*e.g.* carbon capture and storage, extremozymes, alternative energy). By targeting the metagenome and metatranscriptome of such samples in parallel, we will assess both the functional potential and the activities of serpentinite-hosted subsurface microbial communities.

Our previous results have revealed genes involved in lithotrophy, fermentation, and hydrogen oxidation, suggesting that the dominant organisms are supported by

serpentinization-related processes (Brazelton *et al*., 2012). More specifically, our results point to H_2-utilizing Betaproteobacteria thriving in shallow, oxic-anoxic transition zones and anaerobic Clostridia thriving in anoxic, deep subsurface habitats. These data demonstrate the feasibility of metagenomic investigations into novel subsurface habitats via surface-exposed seeps and indicate the potential for H_2-powered primary production in serpentinite-hosted habitats.

Brazelton, W.J., B. Nelson and M.O. Schrenk (2012) Metagenomic evidence for H_2 oxidation and H_2 production by serpentinite-hosted subsurface microbial communities. *Frontiers in Extreme Microbiology* 2:268. doi: 10.3389/fmicb.2011.00268.

The KBase Architecture and Infrastructure Design

Tom Brettin* (brettints@ornl.gov),[1] Bob Olson,[2] Ross Overbeek,[2] Terry Disz,[2] Bruce Parello,[2] Shiran Pasternak,[5] Folker Meyer,[2] Michael Galloway,[1] Steve Moulton,[1] Dan Olson,[2] Shane Canon,[3] Dantong Yu,[4] Pavel Novichkov,[3] Daniel Quest,[1] Narayan Desai,[2] Jared Wilkening,[2] Miriam Land,[1] Scott Deviod,[2] Adam Arkin,[3] Robert Cottingham,[1] Sergei Maslov,[4] and Rick Stevens[2]

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]Argonne National Laboratory, Argonne, Illinois; [3]Lawrence Berkeley National Laboratory, Berkeley, California; [4]Brookhaven National Laboratory, Upton, New York; [5]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

The Systems Biology Knowledgebase (KBase) has two central goals. The scientific goal is to produce predictive models, reference datasets and analytical tools and demonstrate their utility in DOE biological research relating to bioenergy, carbon cycle, and the study of subsurface microbial communities. The operational goal is to create the integrated software and hardware infrastructure needed to support the creation, maintenance and use of predictive models and methods in the study of microbes, microbial communities and plants. The KBase architecture and infrastructure design focus is on creating an unprecedented user experience. The integrated software and hardware infrastructure comprises a continuously expanding collection of software and services. These are hosted on a physical infrastructure consisting of high-speed wide area networking, cloud computing resources, and state of the art cluster computing resources.

Achieving our goal of an integrative architecture to support predictive modeling requires constructing a user experience that covers a range of users. These include biologists determined to understand and create biological models, to bioinformaticists chaining together complex workflows to generate, summarize and integrate data that feed into the biological models. These user activities are enabled at various levels of abstraction, including a) knowledge creation, reproduction and sharing; b) rich web applications; c) programmatic Application Programming Interface (API) libraries and scripts; and d) wire level communication access.

The development of the KBase Unified API is based on a service-oriented approach to deliver both functionality and data to the community. Behind this lies a set of services backed by servers. Initially, these services will be developed by the KBase infrastructure team and will support a long term goal of community

developed and contributed services. Our initial set of services will be backed by the following servers:

Genomic Servers that provide access to a rapidly growing set of genomes, features of those genomes, and annotations of both genomes and features.

Expression Data Servers creating access to a growing body of expression data, plus underlying encoding of metadata needed to support interpretation.

Protein Family Servers supporting access to a variety of the existing collections of protein families.

Polymorphism Servers capturing various genetic polymorphisms such as single nucleotide polymorphisms, tandem repeats, and copy number variations.

Phenotype Servers enabling an understanding the relationships between genotype and phenotype.

Compound and Reaction Data Servers supporting a unified and maintained representation of reaction networks.

Metabolic Modeling Servers that support the construction and maintenance of metabolic models.

Regulatory Models Servers that support the construction and maintenance of regulatory models.

…

The Enigmatic Mummified Seals of the Mcmurdo Dry Valleys: Biological Consequences with Implications to Climate Change Predictions

S.C. Cary* (caryc@waikato.ac.nz; caryc@udel.edu),[1,2] C.K. Lee,[1] G. Tiao,[1] I.R. McDonald,[1] and D.A. Cowan[3]

[1]University of Waikato, New Zealand; [2]University of Delaware, Newark, Delaware; [3]University of the Western Cape, Cape Town, South Africa

Since Scott first observed the presence of their weather-worn carcasses on his 1901-1904 *Discovery* expedition to Antarctica, mummified seals were thought to be rare, isolated natural artifacts peculiar to the ice-free areas of the continent. Our recent discovery of over 100 mummified crabeater seal carcasses concentrated on a precipitous portion of a small remote ventifact cliff over 10 km inland in the McMurdo Dry Valleys, deepens the mystery of their provenance and raises unexpected questions concerning the significance of the phenomena and its impact on a severely oligotrophic ecosystem. Dry Valley organisms typically endure extreme temperature and light regimes, high soil salt concentrations, and near-negligible levels of organic C and N; but a single carcass, by providing protection, substantially alters this inhospitable microenvironment, stabilizing temperatures, raising relative humidity by up to 230%, reducing UV exposure, and increasing organic C levels by over 500% in underlying soil. To investigate the biological consequences and timing of these disturbances and as part of the New Zealand Terrestrial Antarctic Biocomplexity Survey (NZTABS), we transplanted a seal

carcass to a pristine patch of valley floor and monitored the microbial communities under the old and new site for 5 years. Remarkably, we found that within only two Antarctic summer seasons, genetic and stable isotopic analyses of bacterial communities taken from the new seal-augmented site resemble those of samples taken from the original site, occupied for >250 years. Moreover, using deep read pyrosequencing the composition and structure of these communities is not unique to the amended environment but a response from the rare members. However, the resulting community is significantly less phylogenetically diverse than those of control sites. Differences in communities failed to correlate with the chemical properties of their respective soils. It seems that abiotic physical modifications (relative humidity, reduction in temperature fluctuations) associated with the presence of a seal carcass, rather than nutrient or biological inputs, led to the observed changes in microbial communities. Furthermore, microbial biomass is an order of magnitude greater and CO_2 respiration 4-5x higher, indicating rates of potential metabolic response and C turnover in the Dry Valleys are far greater than widely believed. This paradigm shift indicates that increases in water availability and C as a result of warmer predicted regional temperatures could radically reduce biodiversity on the continent's ice-free areas on a near-immediate time scale.

Bacterial Xylan-utilization Regulons: Systems for Coupling Depolymerization of Glucuronoxylans with Assimilation and Metabolism

Virginia Chow,[1] Guang Nong,[1] Franz St. John,[2] John D. Rice,[1] and James F. Preston[*] (jpreston@ufl.edu)[1]

[1]Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida; [2]U.S. Forest Service, Forest Products Laboratory, Madison, Wisconsin

Bioconversion of lignocellulosic resources to fuels and chemicals offers an economically promising path to renewable energy. Among the technological challenges to travelling this path is the development of cost-effective processes that render the cellulose and hemicellulose components of these resources to fermentable hexoses and pentoses. Natural bioprocessing of the hemicellulose fraction of lignocellulosic biomass requires depolymerization of methylglucuronoxylans. This depends upon the secretion of endoxylanases that release xylooligosaccharides and aldouronates. Physiological, biochemical and genetic studies with *Paenibacillus* sp. JDR2 (Pjdr2) support a process in which a cell-anchored multimodular glycoside hydrolase family 10 (GH10) endoxylanase catalyzes the depolymerization of methylglucuronoxylans to release the saccharides, aldotetrauronate (methylglucuronoxylotriose), xylotriose, and xylobiose, that are directly assimilated and metabolized (St. John et al., 2006, PubMed). Gene clusters encoding intracellular enzymes, including α-glucuronidase (GH67), xylanase (GH10), β-xylosidase (GH43), ABC transporter proteins, and transcriptional regulators, as well as a separate gene encoding the secreted multimodular GH10 xylanase, are coordinately responsive to substrate induction or repression, supporting their collective role as a xylan-utilization regulon (Chow et al, 2007, PubMed). The rapid rates of growth with polymeric methylglucuronoxylan (in contrast to simple sugars), along with the absence of

detectable products of depolymerization in the medium, indicate that assimilation and depolymerization are coupled processes (Nong et al., 2009, PubMed. The Pjdr2 genome (7,184,930 bp), sequenced and annotated by JGI (http://www.ncbi.nlm.nih.gov/nuccore/NC_012914.1), contains a single copy of the genes comprising the xylan-utilization regulon. Comparisons with other sequenced genomes provide evidence that such systems occur in xylanolytic species in other genera, including *Clostridium*, *Paenibacillus*, and *Thermotoga*. These systems may be used, either in their native configurations or through gene transfer to other organisms, to develop biocatalysts for efficient production of fuels and chemicals from the hemicellulose fractions of lignocellulosic resources.

The Complete Genomes of *Cellulomonas fimi* and *Cellvibrio gilvus* ATCC 13127 Reveal Two Cellulolytic Bacteria and the Reclassification of *Cellulomonas gilvus* comb. nov.

Melissa R. Christopherson* (melissachristopherson@gmail.com),[1] Garret Suen,[1] Shanti Bramhacharya,[1] Kelsea A. Jewell,[1] Frank O. Aylward,[1,2] Joseph A. Moeller,[1,2] A. Christine Munk,[3] Lynne A. Goodwin,[3,4] Cameron R. Currie,[1,2] David Mead,[2,5] and Phillip J. Brumm[2,6]

[1]Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin; [2]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison Madison, Wisconsin; [3]DOE Joint Genome Institute, Walnut Creek, California; [4]Los Alamos National Laboratory, Biosciences Division, Los Alamos, New Mexico; [5]Lucigen, Middleton, Wisconsin; [6]C5-6 Technologies, Middleton, Wisconsin

A better understanding of how microbes deconstruct cellulose could inform efforts to develop next generation biofuels like cellulosic ethanol. To better understand different cellulolytic strategies, we sequenced the genomes of two soil-dwelling cellulolytic bacteria, *Cellvibrio gilvus* and *Cellulomonas fimi*. *Cellvibrio* and *Cellulomonas* species are robust in their ability to degrade a diverse set of complex carbohydrates. *Cellvibrio* species belong to the Gammaproteobacteria, a group in which cellulose degradation is rare. *Cellulomonas* species, members of the Actinobacteria, are the only known facultative anaerobes capable of cellulose degradation. To gain a better understanding of their biology, the type strains for *Cellulomonas fimi* (ATCC 484 ^{T}) and *Cellvibrio gilvus* (ATCC 19169^{T}) were sequenced at the JGI. An initial analysis of these genomes revealed that *Cellvibrio gilvus* is likely not a member of the Gammaproteobacteria but belongs to the genus *Cellulomonas*. Specific analysis using phylogenetics and shared orthology support this hypothesis therefore we propose that *Cellvibrio gilvus* be reclassified as *Cellulomonas gilvus* comb. nov. We performed a comparative genomics analysis between these two *Cellulomonas* genome sequences and the recently completed *Cellulomonas flavigena* genome sequence. An orthology analysis revealed a shared core set of 1,998 genes. Proteins unique to each bacteria consisted mostly of uncharacterized proteins and carbohydrases. KEGG pathway reconstruction analysis suggested limited biosynthetic capabilities for *C. gilvus* compared to the other cellulomonads. In contrast, these cellulomonads had similar cellulolytic capabilities, and an analysis using the Carbohydrate Active Enzymes (CAZyme) database revealed surprisingly few enzymes involved complex carbohydrate degradation. Specifically, each *Cellulomonas* species encodes for a mere 126-173

CAZymes and with 50% of these predicted to be secreted. Despite the minimal number of CAZymes we found that the *Cellulomonas* species were proficient at degrading and utilizing a diverse set of carbohydrates, including crystalline cellulose *in vitro*. Previous studies have identified cellulosome-like protuberances on the cell surface of *Cellulomonas* species containing carbohydrase activity. We found little evidence supporting these observations, as homology searches did no identify any scaffoldins, dockerins, or cohesions typically associated with cellulosomes. Previous work also found physiological differences between *C. fimi* and *C. flavigena* suggesting that cellulomonads employ different cellulolytic strategies. However, we could not identify any major differences in predicted CAZymes that would indicate a difference between these three sequenced cellulomonads. Thus, these genome sequences revealed cellulolytic strategies that do not closely match the current models for cellulose degradation indicating that further work must be conducted in order to understand their highly-cellulolytic nature.

Using the *Corngrass1* Gene to Enhance the Biofuel Properties of Crop Plants

George Chuck* (georgechuck@berkeley.edu),[1] Amy Anderton,[2] Rita Nieu,[2] Jennifer Bragg,[2] Lan Sun,[3] Chenlin Li,[3] Peter Rubinelli,[4] Christian Tobias,[2] John Vogel,[2] Dean Dibble,[3] Seema Singh,[3] Blake Simmons,[3] Rick Meilan,[4] and Sarah Hake[1]

[1]Plant Gene Expression Center /UC Berkeley, Albany California; [2]Western Regional Research Center / USDA, Albany, California; [3]DOE Joint Bioenergy Research Institute, Emeryville, California; [4]Purdue University, West Lafayette, Indiana

All land plants undergo dramatic developmental changes during the juvenile to adult phase transition. In general, juvenile plant material is less lignified and display differences in biomass character and accumulation. Thus, by identifying and controlling the genes that specify this transition, it may be possible to enhance the biomass properties of any crop plant of choice.

Our analysis of the dominant *Corngrass1* (*Cg1*) mutant in maize has directed us to a group of plant specific transcription factors that controls this phase transition. *Cg1* mutant plants are fixed in the juvenile phase of development and increase biomass of vegetative shoots by continuously initiating axillary meristems and juvenile leaves. Furthermore, *Cg1* leaves contain decreased amounts of lignin and increased levels of glucose and other sugars. Thus, the *Cg1* gene keeps the maize plant in a prolonged juvenile state, causing increased biomass and providing an improved substrate for fermentation.

We cloned *Cg1* and showed that it is an unusual grass-specific tandem microRNA gene that is overexpressed in the mutant. Since the target genes of this microRNA are highly conserved in many plant species, we hypothesized that it should be possible to transfer the biofuel properties of the maize *Cg1* mutant into any crop of choice simply by overexpressing the *Cg1* cDNA and downregulating its targets. This was tested in the model dicot *Arabidopsis*, the model tree *Populus*, the model grass system *Brachypodium*, and the biofuel crop plant *Panicum virgatum* (switchgrass).

Field trials of transgenic switchgrass plants overexpressing the maize *Cg1* gene were completed last summer. Similar to maize *Cg1* mutants, these plants displayed increased vegetative biomass and dramatic alterations in flowering time. Moreover, composition analysis using 2-D NMR and FTIR microscopy showed that overall lignin content was reduced, the ratio of glucans to xylans was increased, and surprisingly, that starch levels were greatly increased. Saccharification assays using starch-degrading enzymes in addition to standard cell wall degrading enzymes resulted in 3-4 fold higher sugar release in *Cg1* biomass compared to normal plants. These results point to the utility of this approach for designing new biofuel crop plants.

This research was supported by DOE physical biosciences grant DE-A102-08ER15962.

Single Cell Genomics on *Arabidopsis thaliana* Root Endophyte Communities

Scott Clingenpeel* (srclingenpeel@lbl.gov),[1] Derek Lundberg,[2] Tanja Woyke,[1] Susannah Tringe,[1] and Jeff Dangl[2]

[1]DOE Joint Genome Institute, Walnut Creek, California; [2]University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

Plants grow in close association with microorganisms in the soil. The plant's interactions with its microbial community span the range from pathogenic to commensial (symbiotic or mutualistic). These interactions occur both outside the root in the rhizosphere and with the microbes as endophytes inside the root tissue. Outside of the well-studied examples of mycorrhizal fungi and nitrogen fixing bacteria, little is known about these plant-microbe interactions.

Although metagenomics and metatranscriptomics can provide valuable insight into the capabilities of a microbial community, they have limited utility in studying soil communities due to the very high diversity found in most soil environments. In order to make the best use of such data, it is necessary to have reference genomes from microbes that are closely related to those found in the sample. Historically, this has required the culturing and isolation of particular microbial strains from the environment which is typically only successful for a limited selection of the total diversity present. A recently developed technology, single cell genomics, allows one to obtain genomic sequence from single microbial cells from a sample. This allows the linkage of cellular identity with functional genes and in principle would allow the sampling of greater diversity than can be obtained from culturing. Unlike reference genomes that come from other environments, the cells used to produce single cell genomic data come from the same sample as used for metagenomics or metatranscriptomics and thus are directly relevant to interpreting the metagenomic/metatranscriptomic data.

Single cell genomics technology was used to help explore plant-microbe interactions in *Arabidopsis thaliana*. We decided to focus on the endophyte microbial community inside the plant roots. This community is simpler than that found in the rhizosphere, and microbes living inside the plant could have a clearer impact on the plant than those outside in the rhizosphere. Extensive community

profiling was done using 16S pyrotagging on several natural accessions of *A. thaliana*; each grown in two soil types and sampled at two times during their lifespan. This produced a list of target OTUs that were enriched in the endophyte samples versus bulk soil. Three *A. thaliana* accessions had their endophyte communities sampled for single cell genomics: the reference strain Col-0, and two with endophyte communities the most different from the Col-0, CVI and Sha. Fluorescently activated cell sorting (FACS) was used to obtain single microbial cells from the samples. The cells were lysed and their genomes underwent the multiple displacement amplification (MDA) process to produce enough DNA for sequencing. The amplicons were used as template for 16S PCR and these sequences were matched to the target OTUs from the pyrotag data. To date 409 single cell amplified genomes (SAGs) have been obtained from 98 OTUs. These SAGs represent 9 of 16 bacterial phyla present in the pyrotag data. Forty eight of the SAGs have been shotgun sequenced with the largest assembly being 2 Mbp.

Temporal Patterns of Bacterial Community Composition in Trout Bog Lake, Wisconsin

Benjamin Crary* (Benjamin.crary@gmail.com) and Katherine D. McMahon

Departments of Civil and Environmental Engineering, and Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin

We have investigated intra- and interannual patterns of bacterial community composition in a dystrophic bog lake located in northern Wisconsin. Bog lakes are important features in temperate and boreal landscapes because they are foci for biogeochemial processing of terrestrial organic matter. Trout Bog Lake is part of the North Temperate Lakes Long Term Ecological Research program, and has been routinely sampled for a decade. Studies of change in bacterial communities in Trout Bog have previously been performed using automated ribosomal integrated special analysis (ARISA). However, 16S rRNA gene tag sequencing has recently been performed on a time-series of archived samples. Integrated samples from the epilimnion and hypolimnion spanning from 2007-2009 were chosen for analysis. 16S rRNA amplicons were sequenced using the Illumina platform, and analysis was carried out with the standard operating procedure in mothur, with slight modifications. In short, each sample was rarified to contain 5000 reads after quality filtering, the reads were aligned and an uncorrected pairwise distance matrix was calculated between the reads, and OTUs were then clustered using a average neighbor algorithm. Taxonomies were then assigned to the OTUs using a Bayesian approach. OTUs were first classified against a curated custom database of exclusively freshwater 16S sequences using a 70 percent confidence threshold. Any OTUs that were not confidently classified to a lineage level were classified against the Greengenes 16S trainingset with a 60 percent confidence threshold using the same Bayesian approach. Multivariate ordinations were created using non metric multidimensional scaling and correspondence analysis to infer temporal trajectories of each layer throughout the summer and fall months. Results were compared with the previous studies carried out with ARISA. The previous results showed regular phenologies that were repeated across years. In addition, seasonal events, such as water column mixing and water temperature trends, strongly

correlated to the bacterial community as measured by ARISA. Furthermore, interesting freshwater taxa were tracked through time and compared intra-annually to reveal trends. Such taxa included the acI-B clade of the phylum Actinobacteria and the PnecC clade of the class Betaproteobactera. Future work will incorporate more physical and chemical data to elucidate community drivers.

Mass Genotyping by Sequencing Technology

John D. Curry* (curry@eurekagenomics.com),[1] Amanda K. Lindholm-Perry,[2] Maria Shin,[1] JingTao Liu,[3] Nadeem Bulsara,[3] Paul Dier,[1] R. Mark Thallman,[2] Viacheslav Y. Fofanov,[3] and Heather Koshinsky[1]

[1]Eureka Genomics, Inc., Hercules, California; [2]U.S. Meat Animal Research Center, USDA-ARS, Clay Center, Nebraska; [3]Eureka Genomics, Inc., Houston, Texas

Large scale genotyping of a moderate number of loci is cost prohibitive with current chip-based technologies. We demonstrate the ability to use next generation sequencing (NGS) technologies to genotype thousands of DNA samples for a moderate number (~100's) of loci – a mass genotyping by sequencing technology (**MGST**). Our MGST is a highly multiplexed ligation-dependent PCR that uses barcodes contained within the ligation probes as well as barcodes added by PCR to prepare a library suitable for sequencing on the Illumina and other NGS platforms. A single sample reaction requires as little as 20ng of genomic DNA. Probe hybridization, ligation and the sample indexing PCR can be carried out in a single well of a 96 or 384-well PCR plate. Library construction containing thousands of samples can be completed within 24 hours and requires a single lane on the Illumina flowcell. The cost per sample is highly competitive and will permit the routine genotyping of a moderate number of loci on a large number of samples. We have currently multiplexed between 24 and 137 different loci on 24 to 1536 bovine samples and achieved between 96% to 100% genotype concordances with data established using the Illumina BovineSNP50 BeadChip. Our statistical modeling suggests that a single lane of a GAIIx instrument can accommodate between 5000 to 10000 animals at a density of 100 loci, making this a truly inexpensive assay.

USDA and Eureka Genomics are equal opportunity providers and employers.

The DOE Systems Biology Knowledgebase: Microbial Science Domain

Paramvir S. Dehal* (PSDehal@lbl.gov),[1] Chris S. Henry,[2] Ben Bowen,[1] Steven Brenner,[1] Ross Overbeek,[2] John-Marc Chandonia,[1] Dylan Chivian,[1] Pavel S. Novichkov,[1] Keith Keller,[1] Adam P. Arkin,[2] Robert Cottingham,[3] Sergei Maslov,[4] and Rick Stevens[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, California; [2]Argonne National Laboratory, Argonne, Illinois; [3]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [4]Brookhaven National Laboratory, Upton, New York

The Systems Biology Knowledgebase (KBase) has two central goals. The scientific goal is to produce predictive models, reference datasets and analytical tools and demonstrate their utility in DOE biological research relating to bioenergy, carbon cycle, and the study of subsurface microbial communities. The operational goal is

to create the integrated software and hardware infrastructure needed to support the creation, maintenance and use of predictive models and methods in the study of microbes, microbial communities and plants. The microbes component will be centered on an analysis pipeline that will include annotation of genome sequences, metabolic and regulon reconstruction, generation of metabolic and regulatory models, and reconciliation of models with existing 'omics datasets and datasets uploaded by a user.

The microbes component of the KBase project aims to unify existing 'omics datasets and modeling toolsets within a single integrated framework that will enable users to move seamlessly from the genome annotation process through to a reconciled metabolic and regulatory model that is linked to all existing experimental data for a particular organism. More importantly, we will embody tools for applying these models and datasets to drive the advancement of biological understanding and microbial engineering.

In order to drive the development of the microbes area and enable new science, we will focus on accomplishing prototype science workflows rather than general tasks. Work will be bootstrapped by leveraging data sets and tools developed and maintained by the MicrobesOnline, SEED, RegPrecise and ModelSEED resources. The initial microbes efforts will integrate prototype workflows for: (1) genome annotation and metabolic reconstruction, (2) regulon reconstruction, (3) metabolic and regulatory model reconstruction, and (4) reconciliation with experimental phenotype and expression data.

*1. Evidence Based Genome Annotation and Metabolic Reconstructi*on*:* High quality gene annotations with confidence measures are a critical to all genome scale modeling. Efforts to create genome scale regulatory and metabolic models are limited by the poor quality of existing gene models. To help resolve this, we are proposing a workflow that takes genome sequence, RNASeq and/or high density tiling array data, and functional 'omics datasets.

*2. Regulon Reconstructio*n*:* Given accurate gene models, the KBase framework will provide integrated pipelines for building and refinement of higher level, regulatory and metabolic models. Reconstruction of genome-wide transcriptional regulatory network (TRN) is a necessary step toward the ultimate goal, building a *predictive* model of microbial organism. …

Genome-Scale Discovery of Cell Wall Biosynthesis Genes in *Populus*

S.P. DiFazio* (spdifazio@mail.wvu.edu),[1] Wellington Muchero,[2] Gancho T. Slavov,[1] Joel Martin,[3] Wendy Schackwitz,[3] Eli Rodgers-Melnick,[1] Christa P. Pennacchio,[3] Uffe Hellsten,[3] Len Pennacchio,[3] Lee E. Gunter,[2] Priya Ranjan,[2] Dan Rokhsar,[3] and Gerald.A. Tuskan[2]

[1]Department of Biology, West Virginia University, Morgantown, West Virginia; [2]BioSciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]DOE Joint Genome Institute, Walnut Creek, California

The primary goal of the *Populus* Activity in BESC1 has been the identification of genes controlling cell wall formation which ultimately positively impact sugar

release, i.e., overcoming recalcitrance. The quantitative genomics portion of the project has focused on Quantitative Trait Locus analysis in two large interspecific families, and association genetics to mine natural variation in *Populus trichocarpa*. Specifically, we established 2 QTL populations, 1 in West Virginia and 1 in eastern Oregon and have created a saturated genetic map containing >6000 genetic markers. This map was used to identify regions of the *Populus* genome that control sugar release. Six such regions were found and, in combination with the transcript profiling and association genetics results, six genes within these regions have been verified as improving sugar release. In addition, in the association mapping study we collected 1,100 genotypes from across the native range of *Populus trichocarpa*, established clonal replicates of each genotype in three common gardens in the Pacific Northwest and subjected two-year-old samples from the Corvallis, OR site to the high throughput phenotyping pipeline established at NREL. Simultaneously, we created a 36,000 single nucleotide polymorphisms (SNP) genotyping Infinium chip based on resequencing data generated by JGI from 15 alternate *P. trichocarpa* genotypes. This SNP array was used to interrogate all 1,100 genotypes found in the common gardens. Association genetics statistical approaches were used to identify specific SNP within specific genetic loci that controlled sugar release and other relevant cell wall phenotypes. From this analysis we identified 46 genes and their amino acid substitutions that are controlling the phenotypes measured in this population. The average increase in sugar yield associated with each SNP is approximately 26% above the wild type control. These genes have been nominated to and accepted within the Transformation pipeline managed by our corporate partner ArborGen. We are continuing to phenotype the QTL and association population for wood chemistry and sugar release, as well as a wide array of traits that will impact productivity in addition to recalcitrance, thereby paving the way for follow-on studies and commercialization during the next phases of the project.

Ecological Genomics of the Opportunistic Fungus *Trichoderma*

Irina S. Druzhinina* (irina.druzhinina@tuwien.ac.at)

Area Gene Technology and Applied Biochemistry, Institute of Chemical Engineering, Vienna University of Technology, Vienna, Austria

Trichoderma is a mycoparasitic filamentous fungus with a broad spectrum of applications, such as industrial production of enzymes and secondary metabolites and use in agriculture for plant defense from pests (*bio*fungicide) and stimulation of plant immune system and growth. The recent molecular ecological and metagenomic surveys show that the majority of *Trichoderma* species inhabit dead wood and/or fruiting bodies of other fungi. Only a minor portion of species also occur in soil, rhizosphere or in association with plants (endophytes) and animals (opportunistic pathogens of immunocompromised humans). In this report I will present the advances in molecular ecology and genomics of *Trichoderma* and demonstrate that mycotrophy, including saprotrophy on fungal biomass and various forms of mycoparasitism, is the innate property of the genus. This trait - combined with broad environmental opportunism - enables establishment of *Trichoderma* in various ecological niches and its association with innumerable substrata including plants and animals.

Read Error Filtering for Illumina Marker Gene-sequencing

Robert C. Edgar* (robert@drive5.com),[1] Viacheslav Y. Fofanov,[2] and Heather Koshinsky[3]

[1]Independent scientist, Tiburon, California; [2]Eureka Genomics, Houston, Texas; [3]Eureka Genomics, Hercules, California

High-throughput sequencing technologies are increasingly being employed to enable cost-effective characterization of populations by analysis of sequence data from marker genes, such as 16S rRNA (bacteria), 18S (eukaryotes) and ITS (fungi). In these types of studies, sequence variation in the reads arises from three sources: (i) natural biological variations (variations between individuals, between strains, species and higher taxonomic groups), (ii) PCR artifacts (copy errors and chimeric amplicons), and (iii) sequencer error. In genome re-sequencing applications (if there is sufficient depth of sequence data), single nucleotide natural variation between individuals and between strains (SNPs) can be successfully distinguished from experimental errors by mapping the reads onto the known genome reference sequence. However, when sequence data from marker genes is used to characterize a population, the exact combination, relative abundance, and even the reference sequence of the organisms in the population are often unknown. Sequence data differences due to natural biological variation (the portion that provides the biological information), PCR artifacts, and sequencer error cannot be readily distinguished.

A central bioinformatics challenge in marker gene-based studies is therefore to filter errors due to amplification and sequencing, without a known reference gene. We have explored using overlapping paired-end reads, where the two sides of the read pair cover the two sides of the double stranded DNA. In this approach, biological variations are reflected in both subsequences of the read pair, while sequencer errors may be identified by mismatches in the overlap. Furthermore, we have explored two approaches for filtering and auto-correcting red errors: (1) utilizing Phred quality scores reported by the sequencing platform, and (2) using spatial location of the mismatch within the read to determine the consensus nucleotide in the mismatched location.

Using overlapping paired end reads from a deep sequenced sample of a bacteria with a well established in silico reference genome - *Pseudomonas aeruginosa* (NC_008463), as well deep sequenced 16S rRNA from a microbial population mixture, we show that even in the absence of a reference genome, sequencing errors can be successfully filtered.

Transcriptomic Analysis of a Mid-winter Algal Bloom in Lake Erie Provides Insight on Adaptations to Psychrophilic and Low-light Environments and the Role of Pathogenic Oomycetes

Robyn Edgar, Paul Morris* (pmorris@bgsu.edu), Vipa Phuntumart, George Bullerjahn, and Robert Michael McKay

Department of Biological Sciences, Bowling Green State University, Bowling Green, Ohio

In 2009, samples from a mid-winter algal bloom were collected with bias to filamentous diatoms using a 153µ vertical net tow to a depth of 15m. The algal bloom population was predominately Aulacoseira sp. with notable presence of Stephanodiscus sp. and Fragilaria sp. cDNA was produced from the samples and sequenced using 454. After assembly using MIRA 3.0, the resulting metagenomic library was 11,576 contigs. BLAST analysis against the NCBI non-redundant nucleotide and protein databases, and NCBI's environmental protein and nucleotide databases revealed that ~56% of the assembled contigs had relevant hits (E-value < $1xE^{-10}$) and ~60% of those had best hits to algal genomes. The assembled library was also blasted against the *Phytophthora sojae* version 3.0 (predicted proteins and genes from JGI), *Phytophthora infestans (*proteins from the BROAD), AphanoDB, *Saprolegnia parasitica* (proteins from the BROAD), and *Pythium ultimum (*proteins). This analysis identified an additional ~300 sequences with stronger homology to oomycetes than anything in NCBI's databases and ~10% of these hits were secreted proteins. The surprising abundance of organisms that are known to be major pathogens in terrestrial ecosystems suggests that they may be playing a similar role in the annual decline of this bloom.

Kmernator—An HPC Application and C++ Toolkit for Analyzing, Assembling, and Manipulating Terabase and Larger Metagenomic Datasets

Rob Egan* (RSEgan@lbl.gov) and Zhong Wang

DOE Joint Genome Institute, Walnut Creek, California

Few assemblers to date can execute with the massive amount of deeply sequenced metagenomic data that is now possible with next generation sequencers. We present Kmernator as a tool to reduce and cluster such datasets in a scalable and useful way. Kmernator can remove and trim low abundance reads and errors from a metagenomic dataset. These problematic reads, which are thought to be either sequencing errors or reads from rare genomes, have very low coverage which both prevents and unnecessarily complicates the assembly process. The reads with low abundant kmers can constitute over 90% of the required memory by various assemblers and because they do not contribute productively to the assembly waste CPU cycles as well. Because Kmernator is built in C++ using MPI, it can scale to enormous data sizes by using a cluster of commodity nodes and will use all the collective memory of those machines in an efficient manner. Kmernator was used as a filtering step in the landmark cow rumen paper published in Science Jan 2011, and we hope it can be released under an open source license in the near future.

Airborne Bacterial Populations Structure in Urban, Industrial, and Newly Urbanized Areas in Cairo by 16S Rdna Pyrotag Analysis Reveals Location Specific Diversity

Sawsan Elgogary* (s.elgogary@aucegypt.edu),[1] Narguess Marei,[1] Maha Kadry,[1] Ihab Osman,[1] Hebatulla Morgan,[1] and Ari Ferreira[1, 2]

[1]YJ- Science and Technology Research Centre and [2]Department of Biology, School of Sciences and Engineering, The American University in Cairo, New Cairo, Egypt

The city of Cairo offers an interesting model to study the microbial populations dynamic in natural outdoor air environments. Cairo ranks 16th amongst the most populated cities in the world, its population being estimated around 17 million people. It is located on the eastern bank of the River Nile, and has a desert climate characterized by very dry heat and specific wind pattern. It receives natural dust carried out by wind from its surrounding desert and hills that greatly influence its outdoor air environments. Cairo's air quality profile has been extensively studied at the level of chemical pollutants, but before this report, its microbial populations composition and structure had not been described in details. To assess the microbial assembly of outdoor air environments in Cairo and to indicate its specific location profiles, we targeted three strategic areas: 1) Tahrir Square, in downtown Cairo and one of its most urbanized locations, 2) Shoubra El-Kheima, north of Tahrir Square and the most industrialized area in Cairo, and 3) New Cairo, a new satellite city of Cairo east to Tahrir Square. These three locations are strategic, because the northern and southern winds blow dust carrying pollutants and microorganisms to downtown Cairo. This makes air particles pollution in downtown Cairo higher than 10 to 100 times of acceptable world standards.

Bacterial 16S rDNA pyrotag analysis (V4-V6 regions) revealed that a specific class is predominant in the aerial assembly of each location: Gammaproteobacteria at Tahrir Square, Bacilli at Shoubra El-Kheima and Betaproteobacteria at New Cairo. Furthermore, each local aerial microbial community is dominated by a specific genus. The genus *Escherichia*, which is among the microbial inhabitants of the gastrointestinal tract and has many pathogenic species, prevails at Tahrir Square The genus *Planomicrobium*, which contains some hydrocarbon-degrading species, abounds in Shoubra El-Kheima, whereas New Cairo is dominated by the genus *Ralstonia*. Diversity analysis at the genus level indicated that the aerial microbial communities of Tahrir Square and Shoubra El-Kheima are more related to each other than each to New Cairo. Besides, microbial diversity varies among the three locations; Tahrir Square shows the highest and New Cairo, the lowest diversity.

The goal of this project is to identify the bacterial population structures in all neighbourhoods of Cairo, and to relate these structures to Egypt's climate and wind pattern, an approach that will provide a deep understanding of the dynamic of microbial populations in the natural outdoor environment, using Cairo as a model case.

The TrophinOak Project and the Genome Draft of the Mycorrhizal Fungus *Piloderma croceum*

Lasse Feldhahn* (lasse.feldhahn@ufz.de),[1] Florence Kurth,[1] Sabine Recht,[1] Ivo Grosse,[2] Tesfaye Wubet,[1] Igor Grigoriev,[3] Annegret Kohler,[3] Francis Martin,[3] Sylvie Herrmann,[1] Mika T. Tarkka,[1] and Francois Buscot[1]

[1]Helmholtz Centre for Environmental Research – UFZ, Halle, Germany; [2]Martin-Luther University, Halle, Germany; [3]UMR INRA/UHP 1136, Centre INRA de Nancy, Champenoux, France ; [4]DOE Joint Genome Institute, Walnut Creek, California

The TrophinOak project "Multitrophic Interactions in Oaks" offers an experimental system to study plant-microbe interactions. Pedunculate oak *Quercus robur* is an important experimental model for forest tree research, characterized by its wide geographic distribution and by the multitude of interacting organisms. These features make *Q. robur* a valuable model for the analysis of plant responses to environmental and biotic stimuli. The project consortium TrophinOak involves seven German Research teams, which investigate how *Q. robur* coordinates its physiological and molecular responses during interactions with beneficial and detrimental organisms. Since oak genome has not yet been sequenced, we used a *de novo* hybrid assembly approach to produce a reference contig library of our oak clone DF159. For this, 16 normalized cDNA libraries were sequenced with 454 and two non-normalized cDNA pools were paired ends sequenced with Illumina. The data was combined and assembled with a short read transcriptome assembler. The future oak RNA-Seq studies will be based on mapping on this assembly.

Micro-organisms of central importance in nutrient acquisition and health of the pedunculate oak are mycorrhizal fungi such as *Piloderma croceum*. The draft genome of *P. croceum* has been sequenced by JGI in the frame of the CSP "Exploring the Genome Diversity of Mycorrhizal Fungi to Understand the Evolution and Functioning of Symbiosis in Woody Shrubs and Trees". Gene annotation of *P. croceum* is in progress, as well as a RNA-Seq study to reveal the mycorrhiza related transcriptome of the fungus.

In our contribution, we will present first data from these ongoing genomics projects. *LINK:* www.**trophinoak**.ufz.de/

Class-II Peroxidases in the *Pleurotus ostreatus* genome: A Study of Their Catalytic Properties, Thermal/pH Stability, and Lignin-depolymerizing Ability

Elena Fernández-Fueyo,[1] Francisco J. Ruiz-Dueñas,[1] María Jesús Martínez,[1] Kenneth E. Hammel,[2] and Angel T. Martínez* (ATMartinez@cib.csic.es)[1]

[1]CIB, CSIC, Madrid, Spain; [2]Forest Products Laboratory, U.S. Department of Agriculture, Madison, Wisconsin

Two *Pleurotus ostreatus* monokaryons were sequenced at JGI (representing 12206 and 12330 gene models) in a project coordinated by Gerardo Pisabarro (Pamplona, Spain), due to the interest as a model ligninolytic organism and as a mushroom consumed worldwide. After manual curation of all the predicted heme peroxidase genes, their deduced amino-acid sequences were converted into structural models

at the Swiss-Model server (using crystal structures as templates) and a first inventory of the different peroxidase types was obtained. They consisted of one Class-I and nine Class-II peroxidases, all in the superfamily of plant-fungal-bacterial peroxidases, together with seven members of the heme-thiolate peroxidase and dye-decolorizing peroxidase superfamilies. Because of their biotechnological interest for lignin removal, we focused first on Class-II peroxidases. Taking into account the predicted Mn-oxidation and aromatic-oxidation sites (the latter at an exposed tryptophan), they were classified as five manganese peroxidases (MnP) and four versatile peroxidases (VP). Using *Escherichia coli* expression with in vitro activation, these nine Class-II peroxidases were produced and purified. Their catalytic properties on five selected substrates were then evaluated, together with their stability properties. Two of the putative VPs oxidized Mn2+ but were unable to oxidize the high redox-potential substrates veratryl alcohol or Reactive Black 5, so they were reclassified as MnPs. The structural reasons for this inability, despite the presence of a tryptophan residue homologous to that of typical LiPs/VPs, remain to be investigated. All the MnPs oxidized Mn2+, and also the low redox-potential substrates 2,6-dimethoxyphenol and ABTS, in agreement with results described for other "short MnPs", whereas the two true VPs were able to oxidize all the above substrates. Comparison of the temperature stabilities (in the 25-70ºC range) and pH stabilities (in the pH 2-9 range) revealed surprising differences for: i) T50-10min (temperature at which 50% of activity is lost after 10 min), which ranged from 34 to 63ºC; and ii) residual activity at both acidic pH (from 0 to 96% after 4 h at pH 3) and alkaline pH (from 0 to 57% after 4 h at pH 9). Differences in the catalytic and stability properties suggest that *P. ostreatus* may differentially express the two VP and seven MnP isoenzymes according to environmental and/or growth conditions, a hypothesis to be probed in further studies. Finally, ligninolysis by *P. ostreatus* VP was demonstrated by depolymerization of 14C-labelled synthetic lignin (14C-DHP) in the presence of veratryl alcohol, whereas the MnP isoenzymes were unable to depolymerize lignin, although they can contribute by oxidizing lignin-derived phenols or other compounds. This is the first time that ligninolysis has been demonstrated for a member of the VP family of peroxidases. This work was supported by the European Union PEROXICATS project (www.peroxicats.org) and the U.S. DOE.

Insights into the Evolution of Microbial Iron Oxidation through Comparative Genomics

Erin K. Field* (efield@bigelow.org), David A. McClellan, Ramunas Stepanauskas, and David Emerson

Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, Maine

After oxygen, iron is the most abundant redox element in the Earth's crust, and Fe(II) is an important electron donor for chemolithoautotrophic growth by microorganisms. Iron-oxidizing bacteria (FeOB) are prevalent in numerous freshwater and marine environments. They are well known for their role in acid mine drainage systems and more recently their role in marine systems including the detrimental process of biocorrosion. A better understanding of the evolution of

FeOB and their strategies for oxidizing iron may provide further insight into the role they play in these environments as well as how to control their growth. We have conducted comparative genomics analysis of cultured freshwater and marine FeOB and plan to expand this to include genomes acquired through single cell analyses, a culture-independent method, of novel marine FeOB. This work presents the first analysis of the complete genomes of two freshwater, obligately lithotrophic, oxygen-dependent FeOB that grow at circumneutral pH. Both *Sideroxydans lithotrophicus* ES-1 and *Gallionella capsiferriformans* ES-2 are adapted to a microaerobic lifestyle and have cbb_3-type cytochrome oxidases and cytochrome *bd* complexes with high affinities for O_2 that are adapted for suboxic environments. The biochemical mechanisms of biological Fe-oxidation coupled to energy conservation at neutral pH are not understood. The genomes of ES-1 and ES-2 offer some clues. Both contain homologs of the pioA/mtrA and pioB/mtrB genes that have been shown to be involved in photoferrotrophy in *Rhodopsuedomonas palustris* TIE-1, and in extracellular electron transfer to Fe minerals in the Fe-reducer *Shewanella oneidensis* MR-1; however neither pioC nor mtrC are present. Both strains have gene clusters for the alternative complex III (ACIII) respiratory complex based around molybdopterin oxidoreductases that could be used to capture electrons from Fe(II) and complete an electron shuttle to the cytoplasmic membrane. Interestingly, this is the only gene cluster that they share in common with the marine Fe-oxidizing bacterium *Mariprofundus ferrooxydans* PV-1. The use of molybdopterin oxidoreductases is potentially a unique mechanism for iron oxidation that is different than that used by acidophilic microorganisms and may represent a model iron oxidation system for other FeOB at circumneutral pH. These results suggest that this could either be an instance of horizontal gene transfer or convergent evolution, as marine and freshwater FeOB belong to different classes of *Proteobacteria*, and, overall, share highly homologous genes in common. Unfortunately, as yet, there are few genomes available to evaluate these evolutionary relationships, partly due to an inability to culture them. Therefore, single cell techniques have also been employed through which whole genome amplification and assembly will significantly improve the number of genomes available for comparative analyses. Specifically, single cells of the *Zetaproteobacteria*, a novel group of marine FeOB, have been obtained from the Loihi Seamount and successfully subjected to whole genome amplification. These single amplified genomes are currently being sequenced. …

High Resolution Characterization of Bacterial Diversity and Geochemistry in a Meromictic Lake (Green Lake, Fayetteville, NY)

Jinnie M. Garrett* (jgarrett@hamilton.edu), Michael L. McCormick, Elizabeth Pendery, Valerie Valant, Andrew Seriachik, Agne Jakubauskaite, and Meghan Griesbach

Department of Biology, Hamilton College, Clinton, New York

Meromictic lakes are composed of non-mixing geochemically distinct strata. Green Lake is perhaps the most studied meromictic lake with regard to its geology and limnology; however, the microbiological characterization of this lake remains limited. Here we present the first high-resolution molecular based survey of

bacterial community composition and geochemistry throughout the water column. A novel multilevel sampler was constructed to aseptically acquire samples at 1M intervals throughout the water column.and at 0.25M or 0.01M intervals through the chemocline (transition zone). Geochemical data (pH, alkalinity, ORP, temp., conductivity, S2-, O2, NH3, and CH4) were recorded at each depth. Cells were collected by filtration and subsequently extracted to isolate genomic DNA. Domain specific PCR primers for bacteria were used to amplify target regions of the 16S rRNA gene. Terminal restriction fragment length polymorphism (TRFLP) was used to assess shifts in community composition with depth with samples. Clone libraries were constructed at 13 depths and sequenced for phylogenetic assignment to nearest related organisms. Finally, TRFLP and geochemical data were statistically analyzed to determine which geochemical parameters best correlated with the abundance of specific taxa and best predicted ecological characteristics such as species dominance and diversity. The oxic oligotrophic surface waters (*mixolimnion*) are dominated by *Synechococcus* spp. while the major bacterial taxa in the sulfidic bottom waters (*monomolimnium*) are Deltaproteobacteria (Syntrophobacterales and Desulfobacterales). TRFLP and clone library analyses both showed peaks in bacterial diversity (total operational taxanomic units) immediately above the chemocline and minima where purple sulfur bacteria (mostly *Thiocystis* like Chromatiales) predominate. Strong correlations exist between the abundance of specific taxa and prevailing geochemistry suggesting functional roles for specific populations that will be tested in future work. Initial statistical analyses indicate ORP is the best predictor of diversity in this system.

Identification of Potential Serine Proteases from the Microbial Community of the Red Sea Atlantis-II Brine Pool Using a Metagenomic Approach

Mohamed Ghazy, **Nadine El-Said*,** Tamer Said* (tsaid@aucegypt.edu), Ayman Yehia, Ahmed Sayed, Ari J. S. Ferreira, Rania Siam, and Hamza El Dorry

Department of Biology, YJ-Science and Technology Center, The American University in Cairo, New Cairo, Egypt

The Red Sea has a multitude of interesting and unexplored environments including deep-sea brine pools with diverse chemical, physical, and geological conditions. The Atlantis II deep is the largest brine pool, 60 km^2, located in the central Red Sea (21°21' N, 38°04' E) at a depth of 2200 meters from the sea surface. The deepest part of this brine, the Lower Convective Layer (LCL), has extreme environmental conditions characterized by high salinity (26%) and temperature (68 °C), low pH (5.3), anoxia, and a high concentration of heavy metals. This environment exerts selective pressure allowing adapted extremophilic microbial communities to survive in the harsh abiotic conditions. Such extremophiles possess proteins and enzymes with unique structural and enzymatic activities that could have biotechnological applications.

Serine proteases are one of the most ubiquitous enzymes found all over the phylogenetic kingdoms. They are an important group of proteolytic enzymes with a broad range of biotechnological applications including but not limited to detergents, food, and pharmaceutical industries. In this work we utilized an

integrated metagenomic approach to identify novel subtilisin-like serine proteases. This was achieved by (a) mining for subtilisin-like serine proteases sequences in an LCL metagenomic dataset generated by pyrosequencing of genomic DNA extracted from the LCL prokaryotic community, (b) functional screening of an LCL metagenomic fosmid library composed of 10,656 clones.

Out of 1,167,000 LCL metagenomic dataset, we identified 276 sequences matching potential subtilisin-like serine proteases that were subsequently assembled into 31 contigs. In parallel functional screening allowed us to identify 9 positive candidates. Further characterization, revealed that the amino acid composition of one of the potential serine protease exhibit a prominent acidophilic signature when compared to reference serine protease. Interestingly it also possessed a binary aspartic acids at the C-terminal of the protein, which may attribute to the halophilic nature of the protein. The unique characteristic of this extremophilic subtilisin-like serine protease, isolated from the deepest layer of the Red Sea and its potential role in biotechnology will be addressed.

Lake Bacteria: Sneaky Thieves, Primary Producers or Just Boring Floaters?

Trevor Ghylin* (ghylin@wisc.edu),[1] Katherine McMahon[2]

[1]Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, Wisconsin; [2]Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, Wisconsin

Freshwater lakes support billion dollar recreational and fishing industries. Lakes also provide invaluable ecosystem services such as carbon sequestration, nutrient cycling and primary production, not to mention valuable shoreline real estate. Unfortunately, nutrient runoff has degraded the quality of many lakes due to eutrophication, jeopardizing the important industries and ecosystem services that our society relies on.

In order improve the quality of lakes we must understand the fundamental biochemical processes in the lakes. Bacteria are critical in mediating the lake biochemical processes that underpin the lake food web, ecosystem services and water quality. In order to improve lakes we must understand the lake bacteria. If we can understand the microbiological processes in the lake we can then create accurate biochemical models of lakes that can predict water quality and aid policy makers and watershed managers in improving lake quality and increasing lake value.

Actinobacteria represent one of the most dominant phyla of bacteria in most freshwater lakes. Specifically, the acI lineage of Actinobacteria can represent over 50% of bacterioplankton cells in a lake's epilimnion. This lineage occurs only in freshwater systems but is found in a wide variety of lakes across the globe. It appears to be well adapted to lake life. The acI lineage are small cells (~1micron) with equally small genomes (~1 Mbp). They are free-floating planktonic organisms that tend to live as single cells, not clusters or flocs. It's not clear how the acI tribes get energy but they are believed to be heterotrophic bacteria. Some research has shown that they may feed on organic substances excreted by other organisms.

Research has also shown that some acI tribes contain rhodopsin genes which may allow them to harvest sunlight for energetic and/or photosensory functions. Research has also shown that some of acI's success may be due to its small size, which makes it a less desirable food source for grazing protists.

Information regarding the characteristics and metabolism of members of the acI lineage has been difficult to obtain because none of these organisms have been cultured. We have utilized culture independent molecular methods to quantify several tribes (i.e. species) of acI in Lake Mendota (Madison, WI) over time. We have correlated these abundances with environmental variables in order to determine environmental variables that impact each tribe.

New Isolates of *Geobacter*, *Desulforegula*, *Desulfovibrio*, and *Pelosinus* and Their Roles in a Low Diversity Consortia During Sustained *In Situ* Reduction of U(VI)

Thomas M. Gihring,[1] Wei-Min Wu,[4] Gengxin Zhang,[1] Craig C. Brandt,[1] Scott C. Brooks,[1] James H. Campbell,[1] Susan Carroll,[1] Craig S. Criddle,[4] Stefan J. Green,[2] Phil Jardine,[3] Joel E. Kostka,[2] Kenneth Lowe,[2] Tonia L. Mehlhorn,[2] Will Overholt,[3] David B. Watson,[1] Zamin Yang,[1] and **Christopher W. Schadt*** (schadtcw@ornl.gov) [1,3]

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]Florida State University, Tallahassee, Florida; [3]University of Tennessee, Knoxville, Tennessee; [4]Stanford University, Stanford, California

Subsurface amendments of slow-release substrates (*e.g.*, emulsified vegetable oil; EVO) are potentially a pragmatic alternative to using short-lived, labile substrates for sustained bioimmobilization within contaminated groundwater systems. The spatial and temporal dynamics of geochemical changes and changes in subsurface microbial communities during EVO amendment are unknown, and will likely differ significantly from populations stimulated by readily utilizable soluble substrates such as ethanol and acetate. In this study we tracked the dynamic changes in geochemistry and microbial communities for 270 days following a one-time EVO injection that resulted in decreased groundwater U concentrations that remained below initial levels for approximately 4 months. Pyrosequencing and quantitative PCR of 16S rRNA from monitoring well samples revealed a rapid decline in groundwater bacterial community richness and evenness after EVO injection, concurrent with increased 16S rRNA copy levels, indicating the selection of a very narrow group consisting of 10-15 dominant OTUs. By association of the known physiology of close relatives it is possible to infer a hypothesized sequence of microbial functions leading the major changes in electron donors and acceptors in the system. Members of the Firmicutes family *Veillonellaceae* dominated after injection and most likely catalyzed the initial oil decomposition and utilized the glycerol associated with the oils and possibly energy sources associated with the small amounts of yeast extract in the EVO product. Sulfate-reducing bacteria from the genus *Desulforegula*, known for LCFA oxidation to acetate, also increased greatly shortly after EVO amendment and are thought to catalyze this process. Acetate and H_2 production during LCFA degradation appeared to also stimulate NO_3-, Fe(III), U(VI), and SO_4^{2-} reduction by members of the *Comamonadaceae*, *Geobacteriaceae*, and *Desulfobacterales*. Methanogenic archaea flourished late in the experiment and at points comprised over 25% of the total microbial

community. Subsequent to the experiment we were able to isolate several of these organisms into pure culture including representatives of phylogenetically distinct isolates and/or species of *Geobacter* and *Desulforegula*, *Pelosinus* and *Desulfovibrio*. A hypothesized model for the functioning of these limited communities is has been developed and is being verified using defined combinations of these isolates where possible. Currently, in collaboration with the JGI, representative strains of the above isolates are undergoing full genome sequencing.

Genome Diversity in *Brachypodium distachyon*: Deep Sequencing of Highly Diverse Natural Accessions

Sean Gordon* (sgordon@lbl.gov),[1] Henry Priest,[2] Wendy Schackwitz,[4] Joel Martin,[4] Anna Lipzen,[4] Kerrie Barry,[4] Ludmila Tyler,[3] Doug Bryant,[2] Wenqin Wang,[5] Antonio Manzaneda,[6] Christopher Schwartz,[7] Richard Amasino,[7] David Garvin,[8] Hikmet Budak,[9] Joachim Messing,[10] Metin Tuna,[11] Thomas Mitchell-Olds,[12] Dan Rokhsar,[4] Len Pennacchio,[4] Ana Caicedo,[13] Samuel Hazen,[13] Thomas Jeunger,[14] Robert Hasterok,[15] John Doonan,[16] Pilar Catalan,[17] Luis Mur,[18] Todd C. Mockler,[2] and John Vogel[1]

[1]USDA, ARS, WRRC, Albany, California; [2]Donald Danforth Plant Science Center, St. Louis, Missouri; [3]University of California, Berkeley, Berkeley California/USDA, Albany, California; [4]DOE Joint Genome Institute, Walnut Creek, California; [5]Waksman Institute, Rutgers University, Piscataway, New Jersey; [6]Universidad de Jaen, Jaen, Andalucia, Spain; [7]University of Wisconsin-Madison, Madison, Wisconsin; [8]USDA-ARS PSRU, St. Paul, Minnesota; [9]Sabanci University, Istanbul, Turkey; [10]Rutgers University, Piscataway, New Jersey; [11]Namik Kemal University, Tekirdag, Turkey; [12]Duke University, Durham, North Carolina; [13]University of Massachusetts, Amherst, Massachusetts; [14]University of Texas at Austin, Austin, Texas; [15]University of Silesia, Katowice, Poland; [16]John Innes Centre, Norwich, United Kingdom; [17]University of Zaragoza, Zaragoza, Spain; [18]Aberystwyth University, Aberystwyth, Wales

Natural variation is a powerful resource for studying the genetic basis of biological traits. Brachypodium distachyon (Brachypodium) is an excellent model grass with a large collection of inbred, diploid lines. These collections contain extensive phenotypic variation. To provide a genomic foundation for future studies, we are deep sequencing 56 diverse inbred natural accessions in collaboration with the US Department of Energy Joint Genome Institute. Analysis of the first six accessions shows tremendous genetic diversity with SNP frequencies ranging from every 200-600 base pairs. We have generated a set of 2,485,097 nonredundant, high confidence SNPs among these six accessions, including 152,920 SNPs in protein-coding regions. The SNP set in this study contains 96.6% (538 of 557) of SNPs previously used to produce a genetic linkage map, indicating a false negative rate of 3.4%. Deep sequencing also revealed numerous indels (61,582-163,776 small indels (1-30bp) and 2,064-8414 large indels and rearrangements (75bp - 20kbp) per accession), which we are confirming by other methods. In addition to comparing the resequenced reads to the reference genome to identify SNPs and structural variation, we are using de-novo assembly and targeted assembly approaches to identify sequences present in the resequenced genomes that are absent from the reference genome. Data and detailed analysis of these results will be made publicly available.

JGI Fungal Genomics Program

Igor Grigoriev* (ivgrigoriev@lbl.gov)

DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, California

Doubling the number of sequenced and annotated genomes every year, the JGI Fungal Program (jgi.doe.gov/fungi) is moving towards the new large scale initiatives aligned with the 2010 Grand Challenges for Biological and Environmental Research: a long term vision. One of the initiatives, the *1000 Fungal Genomes* project, is aimed to explore fungal diversity across the Fungal Tree of Life in order to provide references for research on plant-microbe interactions and environmental metagenomics. Another initiative, the *Genomic Encyclopedia of Fungi*, is focused on diversity among DOE relevant fungi in the areas of plant health and biorefinery parts lists, which will help us to explore the interactions of bioenergy crop species with symbionts and pathogens as well as to catalog industrially relevant genes, pathways, and hosts for biotechnology applications. In addition to broad exploration of fungal diversity, we will also focus on the functional analysis of several *fungal systems* of varying complexity: new model organisms, symbiotic systems such as lichens, and metagenomes of complex communities

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

Agave Transcriptomes and Microbiomes for Bioenergy Research

Stephen Gross* (smgross@lbl.gov),[1,2] Jeffrey Martin,[1,2] June Simpson,[3] Laila Partida-Martinez,[3] Gretchen North,[4] Yuri Peña-Ramirez,[5] Aidee Orozco-Hernandez,[5] Scott Clingenpeel,[1,2] Kristen DeAngelis,[6] Tanja Woyke,[1,2] Susannah Tringe,[1,2] Zhong Wang,[1,2] and Axel Visel[1,2]

[1]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California; [2]DOE Joint Genome Institute, Walnut Creek, California; [3]CINVESTAV, Irapuato, Guanajuato, Mexico; [4]Occidental College, Los Angeles, California; [5]Brown-Forman, Inc., Amatitán, Jalisco, Mexico; [6]University of Massachusetts, Amherst, Massachusetts

Development of new lignocellulosic bioenergy feedstocks minimizing impacts on staple food production and withstanding abiotic stresses anticipated from climate change is key to building a future with sustainable liquid transportation fuels. Because of their exceptional ability to thrive on nutrient-poor soils in arid, hot environments with minimal water and nitrogen requirements, *Agave* species have recently been proposed as an additional bioenergy feedstock. While the physiological mechanisms enabling agaves to survive in their native arid environments are understood, a paucity of sequence information prohibits powerful sequence-based analyses of *Agave* adaptations to abiotic stress. Additionally, as microbes associated with agaves remain unstudied, the role microbe communities may have in augmenting stress resistance is unclear. To address these issues, we are constructing *de novo* reference transcriptomes for *Agave tequilana*, an economically important species cultivated in Mexico for spirit distillation, and *A. deserti*, an extremely thermo- and drought-tolerant species native to the Colorado

Desert. Using the reference transcriptomes, we plan to explore gene regulatory and physiological responses to drought and heat in controlled greenhouse experiments. In parallel, we are investigating the microbiomes of cultivated *A. tequilana*, and wild *A. deserti* and *A. salmiana* using sequence-based microbial community profiling techniques. Select microbiomes will be chosen for deep metagenome sequencing in order to understand microbial genes and pathways conferring additional stress resistance to agaves. Endophytic microbes living within agave tissues will be targeted for single-cell genome sequencing. Taken together, our work builds a robust platform to accelerate discovery of plant adaptations to abiotic stress and further development of *Agave* species as a bioenergy feedstocks.

Discovery of Biomass Degrading Genes from Uncultured Rumen Microbes

Erik Hawley[*1] and **Matthias Hess (**Matthias.Hess@Tricity.WSU.edu)[1,2,3]

[1]Washington State University, Systems Biology & Applied Microbial Genomics, Richland, Washington; [2]Pacific Northwest National Laboratory, Chemical and Biological Process Development Group, Richland, Washington; [3]DOE Joint Genome Institute, Microbial Genomics Group, Walnut Creek, California

Cellulosic plant material provides a renewable alternative to fossil fuels. To overcome the recalcitrance nature of the plant cell wall more efficient biomass-degrading enzymes will be needed. The microbial community of the cow rumen converts grasses that are rich in hemicelluloses, celluloses and other polysaccharides readily into biofuel precursors. Our inability to cultivate the majority of the rumen microbes prevented the large-scale discovery of new biomass-degrading enzymes - but pure cultures might no longer be required for bioprospecting, as direct sequencing of DNA extracted from an environmental sample (known as Metagenomics) has been proven to be a very efficient approach to *i)* identify several thousands of biomass-degrading enzymes and to *ii)* assemble genomes (Hess et al, 2011).

In the project presented here, we analyzed four of the 15 assembled biomass-degrading genomes from the cow rumen microbiome in more detail. We identified 145 genes that have significant sequence identity to know biomass-degrading genes and we cloned some of these putative carbohydrate active enzymes into an expression system. Biochemical assays to determine the physicochemical properties of the heterologously expressed genes are underway.

Microbial Communities in Restored Wetland Sediments

Shaomei He[*] (she@lbl.gov),[1] Mark Waldrop,[2] Lisamarie Windham-Myers,[2] Stephanie Malfatti,[1] Tijana Glavina del Rio,[1] and Susannah G. Tringe[1]

[1]DOE Joint Genome Institute, Walnut Creek, California; [2]U.S. Geological Survey, Menlo Park, California

Wetland restoration on peat islands previously drained for agriculture has a great potential to reverse land subsidence. In addition, the high primary production and slow decomposition rates found in restored wetlands may result in a net

atmospheric CO2 sequestration. However, one major concern is the emission of CH4 that could potentially offset the carbon captured. In this project, we use high-throughput sequencing tools to characterize microbial communities in restored wetland sediments, aiming to understand how the biotic and abiotic environmental factors govern microbial community structure and how microbial communities influence carbon flux, and thus impact long-term biological carbon sequestration. We collected belowground samples from a restored wetland from a U.S. Geological Survey pilot-scale restoration project on Twitchell Island in the Sacramento/San Joaquin Delta, CA. The wetland is continuously fed with river water, and is primarily vegetated with cattails (*Typha* spp.) and tules (*Schoenoplectus acutus*). Samples were collected in both February and August, from three sites that have varied proximity to the inflow, thus exhibiting gradients in physicochemical conditions and peat accretion rates. From each site, samples were collected at two depths (0-12 and 12-25 cm below ground) from three sample types, including the bulk decomposed material, cattail rhizomes and tule rhizomes. The 16S rRNA pyrotag analysis showed that wetland microbial community composition pattern is primarily governed by sample site and sample type. Particularly, the community difference was largely correlated to the physicochemical gradients along these sites. In parallel, mesocosm anaerobic incubation of these samples showed that CO2 flux was significantly higher in the rhizome samples than in the bulk samples. By contrast, statistically significant difference in CH4 flux was observed among sample sites. Low CH4 flux communities were associated with the site closest to the inflow, correlated to higher availabilities of sulfate and nitrate. Some of their more abundant microbial populations, as compared to high CH4 flux communities, are likely reducers of these electron acceptors. High CH4 flux communities were associated with sites further from the wetland inflow, which have shown higher peat accretion rates. These sites harbored more abundant methanogenic archaeal populations, which likely contributed to the higher methane flux observed. Currently, comparative metagenomic analyses are being conducted to reveal differences in community functional profiles.

A Multiplexed, High-throughput Screening Pipeline for Lignocellulosic Enzyme Discovery and Evolution

Richard Heins,[1,4] Xiaoliang Cheng,[1,3] Samuel Deutsch,[2,3] Kai Deng,[1,4] Mary Tran-Gyamfi,[1,4] Suzan Yilmaz,[1,4] Anup Singh,[1,4] Paul Adams,[1,3] Eddy Rubin,[2,3] Blake Simmons,[1,4] Ken Sale,[1,4] and **Trent Northen*** (TRNorthen@lbl.gov)[1,3]

[1]Joint BioEnergy Institute, Emeryville, California; [2]DOE Joint Genome Institute, Walnut Creek, California; [3]Lawrence Berkeley National Laboratory, Berkeley, California; [4] Sandia National Laboratory, Livermore, California

A major challenge to synthetic biology is the disconnect between the rate of clone production vs. specific functional analysis. Typically, pooled assays or selections are required to down select libraries prior to chemically specific analysis using mass spectrometry. Here we present the novel integration and application of a nanoliter-scale acoustic sample deposition and nanostructure-initiator mass spectrometry (NIMS) analysis platform to rapidly detect and characterize glyco-

and lignolytic enzyme activities and substrate specificities from complex environmental samples and crude cell lysates. We are applying this approach for analysis of *in vitro* expression systems and single cell sorting, to quickly screen large libraries of mutant β-glucosidases for enhanced thermostability. This effort will serve as the foundation in the development of this new technology that will have several applications, including enzyme "cocktail" engineering for enhanced performance in industrially relevant biorefining operating environments for the production of sugars from biomass.

Amanita-omics: Towards an Understanding of the Evolution of Ectomycorrhizal Symbiosis, from a Genomic Perspective

Jaqueline Hess* (jhess@fas.harvard.edu),[1] Inger Skrede,[1,2] Benjamin Wolfe,[3] and Anne Pringle[1]

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts; [2]Microbial Evolution Research Group (MERG), Department of Biology, University of Oslo, Oslo, Norway; [3]FAS Center for Systems Biology, Harvard University, Cambridge, Massachusetts

Ectomycorrhizal (ECM) symbiosis is one of the major determinants of forest architecture in temperate and boreal forests. The ubiquitous mutualistic association between fungi and trees is predominantly based on an exchange of nutrients aggregated by the fungus in return for photosynthetically-derived carbon supplied by the plant via a root system colonized by fungal hyphae. ECM symbiosis increases forest primary productivity, provides a below-ground carbon sink and allows the colonization of nutrient-poor areas. Understanding the genetic mechanisms that led to ECM symbiosis and their evolution is thus of central importance for understanding the dynamics of the forest ecosystem and the resulting impact on global carbon cycles in a changing climate. ECM symbiosis is thought to have evolved convergently at least six times and, although genome sequences of ECM fungi are beginning to become available, at the moment relatively little is known about the existence of common genomic underpinnings accompanying the evolution of ECM symbiosis. Earlier studies of the genomes of the Basidiomycete *Laccaria bicolor* and Ascomycete *Tuber melamsporum* indicate that the loss of plant cell wall (PCW) decomposition enzymes is one shared signature of ECM genomes. *L. bicolor* also shows expanded repertoires of signaling proteins, proteases and nutrient transporters.

The genus *Amanita,* containing the ectomycorrhizal fly agaric *Amanita muscaria* and its relatives, spans one of the transitions from a free-living, saprotrophic niche to ECM symbiosis. Using next-generation sequencing technologies, we have generated low-coverage genomic data from five *Amanita* species (three ECM species and two saprotrophs) and the saprotrophic outgroup *Volvariella volvacea.* One lane of paired-end Illumina HiSeq reads per species was used to perform de-novo genome assemblies which we subsequently annotated using a combination of homology- and EST-based approaches. Validation using RNA-seq data suggests that even though assemblies are relatively fragmented, we are achieving a good overview of gene content in the respective species. Here, we will present our *de-novo* assembly and annotation approach and a preliminary comparative analysis of

PCW decomposition enzyme repertoires in ECM and saprotrophic *Amanita* genomes. Our initial analyses indicate the transition within the *Amanita* is also associated with a reduced capability to degrade plant cell walls, emphasizing the importance of this genetic change as a key feature of ECM genomes.

Expression Profile of Biomass-Degrading Fungi Inhabiting the Cow Rumen

Matthias Hess* (Matthias.Hess@Tricity.WSU.edu),[1,2,3] David E. Culley,[2] Kenneth S. Bruno,[2] James Collett,[2] Rod I. Mackie,[4] Igor Grigoriev,[3] Susannah Tringe,[3] Bernard Henrissat,[5] Jon K. Magnuson,[2] and Scott E. Baker[2]

[1]Washington State University, Systems Biology & Applied Microbial Genomics, Richland, Washington; [2]Pacific Northwest National Laboratory, Chemical and Biological Process Development Group, Richland, Washington; [3]DOE Joint Genome Institute, Microbial Genomics Group, Walnut Creek, California; [4]University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois; [5]Architecture et Fonction des Macromolécules Biologiques, Marseille, France

One of the major bottlenecks in the industrial production of biofuels from recalcitrant plant material is the paucity of efficient biomass-degrading enzymes. Natural systems, such as the microbiome (the microbial community) of the cow rumen, have evolved over thousand of years and possess the molecular machinery to efficiently degrade recalcitrant plant cell wall structures. In a recent study (Hess et al Science 2011), we focused on the prokaryotic fraction of the cow rumen microbiome and identified more than 27,000 genes of putative carbohydrate-active genes from prokaryotes that colonized rumen-incubated switchgrass.

Although fungi account only for ~8% of the microbial biomass in the cow rumen, they represent key players in the anaerobic degradation of biomass that occurs in the rumen. Importantly, fungal enzymes appear to have higher specific activities towards plant cell wall polymers than bacterial biomass-degrading enzymes. An extensive catalog of fungal enzymes with biomass-degrading activity would represent a valuable resource for scientists developing more efficient processes for the production of cellulosic biofuels.

The project presented here is part of JGI's Community Sequencing Program 2012 and has the objective to identify and catalogue the biomass-degrading genes that are expressed by rumen fungi during biomass-degradation in the cow rumen.

Hybrid de novo Assembly of Microbial Genomes Utilizing Error Correction of Long Read Sequencing Data

Lawrence Hon* (lhon@pacificbiosciences.com)**,** Lawrence Lee, John Beaulaurier, Jason Chin, Aaron Klammer, Khai Luong, Jonas Korlach

Pacific Biosciences, Menlo Park, California

Current *de novo* assembly tools have generally been designed to target shorter read data. The PacBio® *RS* platform is unique in that it can generate reads much longer than 1 kb and can be run in different ways to optimize for longer reads or lower error depending on the needs of the application. To provide the best performance

on this type of data, assembler tools need to be modified to account for these characteristics. Here, we examine several assemblers, including ALLORA, Celera® Assembler, ALLPATHS-LG, and MIRA, which are able to handle PacBio data. In particular, we focused on the error correction hybrid assembly approach enabled by the PacBioToCA module within Celera Assembler. By analyzing several microbial data sets, we suggest how to incorporate PacBio data to achieve the best possible assemblies.

Rapid and Efficient Methods for Ribosomal RNA Removal from Plant and Metatranscriptome Samples

Cindi A. Hoover* (cahoover@lbl.gov**)**[1] and Cris Kinross[2]

[1] OE Joint Genome Institute, Walnut Creek, California; [2] Epicentre (an Illumina® company), Madison, Wisconsin

Deep sequencing of cDNA prepared from total RNA (RNA-Seq) or mRNA (mRNA-Seq) has become the method of choice for transcript profiling, discovery of novel transcripts, and identification of alternative splicing events. However, standard whole-transcriptome approaches to RNA-Seq face a significant challenge, as the vast majority of reads map to rRNA. One solution—poly(A) enrichment—does not capture several biologically relevant RNA species, such as microRNA and other noncoding RNAs, and is ineffective for prokaryote samples.

To overcome these challenges, Epicentre developed Ribo-Zero™ rRNA removal technology for mammalian, plant, and bacterial total RNA samples. The technology provides excellent removal of rRNA, even from degraded and archived FFPE RNA samples. Here we present preliminary rRNA removal data from two prokaryotic metatranscriptome samples, cow rumen and a sample of mixed prokaryotes. The data show effective rRNA removal and an increase in mapped reads compared to non-depleted control samples. Additionally, we present a comparison of Ribo-Zero kits for removal of rRNA from Plant Leaf or Plant Seeds/Roots on the same rice-stem sample to illustrate the difference in reads mapped to rRNA between these kits. Sequence data were generated using an Illumina HiSeq, but the Ribo-Zero™ technology is compatible with many downstream applications.

Reproductive Genetics and Development in the Fungus *Myceliophthora heterothallica,* a Thermophilic Model Organism for Biomass Degradation

Miriam I. Hutchinson* (miramira@unm.edu),[1] Amy J. Powell,[2] Kylea J. Parchert,[2] Joanna L. Redfern,[1] Randy M. Berka,[3] Eric Ackerman,[2] Blake Simmons,[4,5] Igor V. Grigoriev,[6] and Donald O. Natvig[1]

[1]Department of Biology, University of New Mexico, Albuquerque, New Mexico; [2]Sandia National Laboratories, Albuquerque, New Mexico; [3]Novozymes, Inc., Davis, California; [4]DOE Joint BioEnergy Institute, Emeryville, California; [5]Sandia National Laboratories, Livermore, California; [6]DOE Joint Genome Institute, Walnut Creek, California

Members of the fungal family Chaetomiaceae (order Sordariales) are of interest for their abilities to produce thermostable carbohydrate-active enzymes. The need for new enzymes for efficient biomass deconstruction led to the sequencing by the DOE Joint Genome Institute of genomes from two thermophilic members of the Chaetomiaceae, *Myceliophthora thermophila* and *Thielavia terrestris*. Until now, however, there has been no genetically tractable model either for this family, or more generally, for thermophilic fungi. We have characterized the reproductive biology of the thermophile *Myceliophthora heterothallica* towards the goal of establishing this organism as the model species for the group, and toward developing it as an expression platform.

As its name implies, *M. heterothallica* was reported to be heterothallic based on the fact that matings between two strains resulted in the production of fruiting bodies and ascospores. Prior to the work reported here, however, heterothallism had not been confirmed by showing independent segregation of mating type loci and autosomal genes. We speculate that this lack of confirmation of true heterothallism resulted from a failure to obtain ascospore (sexual spore) germination. We found that ascospores are completely resistant to germination at temperatures below 47-50°C. Above 50°C, germination rates can approach 100%. This discovery allowed us to confirm heterothallism and analyze the segregation of molecular markers in crosses. Sequences obtained from strains possessing different mating types show that the mating regions are conserved relative to other members of the Sordariales. Interestingly, different stages of development have different temperature optima: ascospore germination occurs at 50°C and above, ascocarp formation is optimal at 30°C, and growth is optimal at 45°C. Distinct developmental optima point to a complex life-history evolution from which industrially-useful features, such as high optimal growth temperatures, arose.

To date, we have successfully crossed *M. heterothallica* strains from Indiana, New Mexico and Germany, and we are seeking to expand the number of known strains by surveying across latitudinal and elevation gradients. In addition, we are developing methods for rapid transformation and gene replacement. Our goal is to develop *M. heterothallica* as a flexible, full-service platform for biofuel and other industrial applications, and as a model organism to study fundamental aspects of thermophily and the biology of Chaetomiaceae.

Conserved Peptide Upstream Open Reading Frames (CPuORFs) in Plant and Dipteran mRNAs Are Associated with Regulatory Genes

Richard Jorgensen* (rajorgensen@langebio.cinvestav.mx)

Langebio, Cinvestav-IPN, Irapuato, Guanajuato, Mexico

The amino acid sequences of upstream open reading frame (uORF) are usually not conserved in evolution, but the small class of uORFs whose encoded peptides are conserved seem likely to play roles in translational control of downstream (major) open reading frames (mORFs). By comparing full-length cDNA sequences from Arabidopsis and rice we identified 30 distinct homology groups of conserved uORFs, only three of which had been reported previously. Pairwise Ka/Ks analysis showed that purifying selection had acted to conserve peptide sequences in nearly all CPuORFs. Functions of predicted mORF proteins could be reasonably inferred for 24 homology groups and each of these proteins appears to have a regulatory function, including 6 involved in transcriptional control, 8 involved in signal transduction, 4 involved in small signal molecule pathways, 4 with other known regulatory functions, and two with protein interaction domains. Duplicate copies of genes with a conserved uORF that were created by tetraploidy in an Arabidopsis ancestor are much more likely to have been retained in Arabidopsis than are duplicates of other genes (39% vs. 14% of ancestral genes, p=5x10-3). We propose that the function of most CPuORFs in plants is to act like a 'rheostat' to sensitively modulate ('fine tune') translation in response to a signal molecule that is recognized by the conserved peptide. Analysis of CPuORF genes in Drosophila shows that they too are associated with regulatory genes, though with a different spectrum of functions.

Metabolic Engineering Enhancements to Pathway Tools

Peter D. Karp* (pkarp@ai.sri.com), Michael Travers, and Mario Latendresse

SRI International, Menlo Park, California

This project aims to support computational design of metabolic pathways for metabolic engineering. Users will specify a target metabolite, a feedstock compound, and other constraints on their design problem. A pathway search algorithm will construct alternative pathways by combining reactions from the MetaCyc database. The search algorithm will rank pathways according to multiple criteria. Users will view the output of the pathway search algorithm using a graphical interface that facilitates user comprehension and evaluation of the pathways. The resulting software will be an add-on module to SRI's Pathway Tools software.

We have implemented an initial version of a graph search algorithm to design novel pathways, given a source and target metabolite. This software is capable of integrating a variety of metrics and filters that affect the search. Metrics implemented to date include atom conservation, molecular similarity, avoiding certain molecules, penalties for using a reaction that is outside of the taxonomic range of the target organism, and penalties for generating or consuming certain

kinds of side products/reactants. The search engine is capable of using reactions from any PGDB including MetaCyc.

Atom conservation is determined using atom mappings. A reaction atom mapping describes explicitly the one to one transfer of each atom from the reactants of a reaction to its products. A reaction might have several chemically valid mappings due to the symmetries of reactants and products or for other reasons. These mappings allow the computation of the flow of essential atoms from source to target metabolites in a pathway.

As far as we know, all computational approaches to atom mapping published so far, are combined with a post-processing step involving manual curation. That is, a scientist reviews the computed atom mappings for possible errors and appropriate corrections are applied when necessary. But the necessary corrections are not applied to the computational approach itself to avoid future computed errors.

In the approach we have taken, all the possible mappings of most reactions of MetaCyc are computed. Moreover, we aim to have an accurate computational approach that does not require manual post-processing. Technically, our approach is based on Mixed Integer Linear Programming: for each reaction, a linear program is generated from all possible valid mappings of the reaction where the objective function to minimize is the sum of the costs of the bonds broken and made. We currently use bond cost values that have been determined by a chemist. A linear solver solves the linear program, giving all possible optimal solutions, that is, all possible correct mappings for one reaction. This technique has been applied to almost all the reactions of MetaCyc. It is more computationally efficient than all other known techniques published so far.

Nitrification and Denitrification by Aerobic Methanotrophic Bacteria: Genes, Modules, and Activities

K. Dimitri Kits, Ariel Kangasniemi, Dustin J. Campbell, and Lisa Y. Stein* (lisa.stein@ualberta.ca)

Department of Biological Sciences, University of Alberta, Edmonton, Alberta Canada

Aerobic methanotrophic bacteria share several commonalities with lithotrophic ammonia-oxidizing bacteria. For instance, several, but not all methanotrophs can oxidize ammonia to nitrite via hydroxylamine and can reduce nitrite to nitrous oxide via nitric oxide similarly to ammonia-oxidizers. Genome sequencing and physiological analysis has allowed comparison of nitrifying and denitrifying pathways of aerobic methanotrophs. Although both nitrifying and denitrifying activities are widespread among these bacteria they do not always co-occur in the same strain nor is there conservation of pathways by phylogeny. Thus, the ability of a particular strain to nitrify or denitrify is most likely a consequence of horizontal transfer of genes or modules and environmental pressure. The ammonia-oxidation pathway of methanotrophs involves particulate methane monooxygenase to oxidize ammonia to hydroxylamine. While some strains have homologues to hydroxylamine oxidoreductase that can convert hydroxylamine to nitrite, other methanotrophs use alternative pathways that have yet to be determined. Some strains, like *Methylosinus trichosporium* OB3b, encode a homologue to

hydroxylamine reductase and can convert exogenously added hydroxylamine back to ammonia. Strains of *Methylomonas methanica* are so far the only ones without measureable ammonia- or hydroxylamine-oxidation activities. Denitrification modules likely involve nitrite reductase (*nir*S or *nir*K) and nitric oxide reductase (*nor*CB), although several strains that can reduce nitrite to nitrous oxide lack one or both reductase gene homologues. Therefore, alternative enzymes for nitrite and nitric oxide reduction are yet to be discovered in these strains. Together, genomic and physiological data have revealed the capacity for aerobic methanotrophs to nitrify and/or denitrify. We are currently investigating the function and regulation of these pathways to understand the contribution of methanotrophic bacteria to the global nitrogen cycle.

Metagenome Sequencing of Geothermal Iron-Mat Microbial Communities Reveals Deeply-rooted Archaeal Populations that Vary in Abundance across Changes in pH and Temperature

M. Kozubal,[1] J. Beam,[1] Z. Jay,[1] R. Jennings,[1] H. Bernstein,[2] R. Carlson,[2] D. Rusch,[3] S. Tringe,[4] M. Romine,[5] R. Brown,[5] M. Lipton,[5] H. Kreuzer,[5] J. Moran,[5] and W. Inskeep* (binskeep@montana.edu)[1]

[1] Department of Land Resources and Environmental Sciences and Thermal Biology Institute, Montana State University, Bozeman, Montana; [2]Department of Chemical and Biological Engineering, Montana State University, Bozeman, Montana; [3]Department of Biology, Indiana University, Bloomington, Indiana; [4]DOE Joint Genome Institute, Walnut Creek, California; [5]Department of Energy-Pacific Northwest National Laboratory, Richland Washington

Prior molecular characterization (16S rRNA gene sequence) of numerous Fe(III)-oxide and jarositic microbial mats ranging in temperature from 55 to 85 ^{o}C has revealed several undescribed members belonging to the Sulfolobales, Thermoproteales, Desulfurococcales (orders of the Crenarchaeota) and two novel phylum-level lineages. Moreover, prior metagenome sequencing of two Fe-mats samples (Sanger platform) has provided significant *de novo* assemblies for several of these lineages. However, transects within the primary outflow channels of acidic (pH 2.5-3.5) geothermal springs have revealed significant changes in community composition with decreases in temperature, small changes in pH, and other key geochemical variables (e.g. H_2S and O_2). Consequently, the overall goal of this study was to examine metagenome sequence collected at different temperatures within two well-studied acidic Fe(III)-oxide mats of Norris Geyser Basin (YNP). Microbial Fe-mat samples were collected from three positions (60-75 ^{o}C) in OSP Spring (pH=3.5) and from two positions (60-65 ^{o}C) in Beowulf Spring (pH=3.1), and sequenced using 454 Ti-pyrosequencing, followed by assembly (Newbler), and subsequent functional analysis of gene content of large contigs. Given the total number of different predominant populations (e.g. genera) we are studying across these mats (~12-15), more detailed genome annotation and functional interpretations will follow with subsequent experiments. However, the recent pyrosequencing results on five different samples provide a consistent estimate of the community structure as it changes over small decreases in temperature (e.g. 10-15 ^{o}C) and pH (different springs). Specifically, metagenome assemblies were evaluated using nucleotide word frequency analysis-principal components analysis as a mechanism to screen sequence with similar G+C content and codon usage

bias. Contigs exhibiting similar sequence character were also checked using blast searches to establish phylogenetic identity and consistency. High temperature (75 ^{o}C) samples from OSP Spring contain four predominant archaeal community members, one of which represents a 'novel archaea group 1' (NAG1) population. An Fe(II)-oxidizing member of the Sulfolobales (*M. yellowstonensis*-like) comprised approximately 25% of the archaeal sequence reads. Abundance of NAG1 populations decreases with temperature in OSP Spring, and at lower temperatures (i.e. 60-65 ^{o}C), two additional archaeal populations are observed, which both fall within a different 'novel archaeal group 2' lineage (NAG2). Moreover, metagenome sequence from Beowulf Spring contains considerably less NAG1-like organisms, but higher amounts of both populations within the NAG2 lineage, as well as novel Thermoplasmatales and Thaumarchaeota. *Metallosphaera* populations remain an important fraction of the community across the entire temperature and pH range (e.g. 25-30% of sequence reads), consistent with its role in Fe(II)-oxidation and carbon fixation in acidic high-temperature Fe-mats. …

A Collection of Algal Genomes from JGI

Alan Kuo* (akuo@lbl.gov) and Igor Grigoriev

DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, California

Algae, defined as photosynthetic eukaryotes other than plants, constitute a major component of fundamental eukaryotic diversity. Acquisition of the ability to conduct oxygenic photosynthesis through endosymbiotic events has been a principal driver of eukaryotic evolution, and today algae continue to underpin aquatic food chains as primary producers. Algae play profound roles in the carbon cycle, can impose health and economic costs through toxic blooms, and are candidate sources for bio-fuels; all of these research areas are part of the mission of DOE's Joint Genome Institute (JGI). A collection of algal projects ongoing at JGI contributes to each of these areas and illustrates analyses employed in their genome exploration.

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.

Deep Sequencing of a Plant Transcriptome

Jeffrey Martin* (jamartin@lbl.gov),[1] Stephen Gross,[1] James Schnable,[2] Cindy Choi,[1] Mei Wang,[1] Kanwar Singh,[1] Erika Lindquist,[1] Feng Chen,[1] Chia-Lin Wei,[1] and Zhong Wang[1]

[1]DOE Joint Genome Institute, Walnut Creek, California; [2]University of California, Berkeley, Berkeley, California

De novo assembly of the transcriptome is crucial for functional genomics studies within bioenergy crops, since many of them lack high quality reference genomes. Plant gene annotations are often generated using limited experimental evidence, and largely rely upon the accuracy of gene calling algorithms. Previously, we developed a *de novo* assembly pipeline, Rnnotator, for assembling transcriptomes

in lower eukaryotes using only 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 distinct transcripts from homologous genes. Here, we present our improvements to Rnnotator for large plant transcriptomes and discuss the biological insights gained by analyzing a transcriptome assembled from a deeply sequenced maize sample.

We generated 341Gb of RNA-Seq data from a single maize RNA sample. Rnnotator generated 187,045 transcript isoforms with a median size of 440bp. Using the reference genome and annotated gene models we estimated the accuracy, completeness and contiguity of the *de novo* assembled transcripts to be 93.4%, 78.2% and 63.4%, respectively. There are only 0.7% chimeric transcripts based on the analysis of a small set of 141 homologous gene pairs. In order to assess the accuracy of assembled variants, we independently generated a set of highly confident transcripts using PacBio sequencing. 98.8% of the PacBio long reads validated the assembled transcripts.

In summary we have generated a very accurate and comprehensive maize transcriptome exclusively from short RNA-Seq reads. Current ongoing analysis of this transcriptome will greatly improve the current maize gene annotation, and comparative analysis with rice and sorghum transcriptomes will reveal the set of genes from the maize lineage.

Gene Expression Profiles during Determinate Primary Root Growth in Cardón, *Pachycereus pringlei* (Cactaceae)

Marta Matvienko* (mmatvienko@clcbio.com),[1] Svetlana Shishkova,[2] Alexander Kozik,[3] Mayra Lopez-Valle,[2] Yamel Ugartechea-Chirino,[2] Selene Napsucialy-Mendivil,[2] and Joseph G. Dubrovsky[2]

[1] CLC bio, Davis, California; [2]Instituto de Biotecnología, Universidad Nacional Autónoma de Mexico, Cuernavaca, Mexico; [3]Genome Center, University of California, Davis, Davis, California

Unlike roots of most plant species, the primary root of cardón *Pachycereus pringlei,* a Sonoran Desert Cactaceae, exhibits determinate growth. The root apical meristem of seedling primary root exhausts and all cells in the root tip differentiate. Determinate growth of primary and most lateral roots results in the formation of a compact root system that provides seedlings an advantage for survival in a desert environment. In order to identify and characterize genes involved in root meristem maintenance and determinate root growth in *P.pringlei*, we employed mRNA-seq using IGA II. The 85 nt reads were assembled *de novo* into about 26,000 contigs using the CLC Genomics Workbench. The largest contig of >15 kb represented the longest plant transcript from the *BIG* gene. The transcriptome contigs were annotated using the similarity search against GenBank Ref-seq proteins. Differential gene expression was estimated in the primary root tip. Over 400 and almost 900 transcripts were up-regulated more than 5 times during initial and terminal phases of root growth, respectively. Sixteen putative transcription regulators were up-regulated during the initial phase. Significant conservation

between *P.pringlei* and *Arabidopsis* was revealed for the amino acid sequences and RNA expression patterns for various genes. We also detected differences in expression profiles of some PIN auxin efflux carriers between *P.pringlei* and *Arabidopsis* mutants with determinate primary root growth. The cytokinin synthesis related genes are expressed during *P. pringlei* terminal phase of root development, suggesting that the root tip is functionally active after meristem exhaustion.

The DOE Systems Biology Knowledgebase: Microbial Communities Science Domain

Folker Meyer* (folker@anl.gov),[1] Dylan Chivian,[2] Andreas Wilke,[1] Narayan Desai,[1] Jared Wilkening,[1] Kevin Keegan,[1] William Trimble,[1] Keith Keller,[2] Paramvir Dehal,[2] Robert Cottingham,[3] Sergei Maslov,[4] Rick Stevens,[1] and Adam Arkin[2]

[1]Argonne National Laboratory, Argonne, Illinois; [2]Lawrence Berkeley National Laboratory, Berkeley, California; [3]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [4]Brookhaven National Laboratory, Upton, New York

The Systems Biology Knowledgebase (KBase) has two central goals. The scientific goal is to produce predictive models, reference datasets and analytical tools and demonstrate their utility in DOE biological research relating to bioenergy, carbon cycle, and the study of subsurface microbial communities. The operational goal is to create the integrated software and hardware infrastructure needed to support the creation, maintenance and use of predictive models and methods in the study of microbes, microbial communities and plants. The microbial communities component will be focused on building the computational infrastructure to understand the community function and ecology through study of genomic and functional data and integration of community models with single-organism models. This will allow for researching community behavior and building predictive models of communities in their role in the environmental processes and the discovery of useful enzymes..

The KBase microbial communities team will integrate both existing and new tools and data into a single, unified framework that is accessible programmatically and through web services. This will allow the construction of sophisticated analysis workflows by facilitating the linkages between data and analysis methods. The standardization, integration and harmonization of diverse data types housed within the KBase and data located on servers maintained by the larger scientific community will allow for a single point of access, ensuring consistency, quality assurance, and quality control checks of data.

We have begun by creating KBase data and analysis services that will link our core resources: MG-RAST, metaMicrobesOnline, SEED, IMG/M and ModelSEED. These services will allow clients to access data and analysis methods across these tools without the burden of reconciling identifiers, learning different data access and programmatic access methods, ensuring data quality, and maintaining relevant metadata. New functionality, not currently available in our core tools, is being created within KBase using the programmatic interfaces.

Protoypical applications:

Bioprospecting: Microbial diversity is a key element in the search for new, valuable compounds such as enzymes with novel properties. Integration of metaMicrobesOnline functions with MG-RAST data through the KBase programmatic interface will allow users to elucidate novel proteins from microbial communities. It will allow for deep comparative analysis of protein families, expanding significantly the current functionality in MG-RAST. This includes detailed trees and alignments combining metagenomic sequences and sequences from complete genomes. The initial version will allow in-depth characterization of novel members of existing protein families; future versions will allow characterization of completely novel protein families. This will exercise the communities part of the API and also the microbes set of API calls and provide a useful, missing component to the combined tool suite.

…

In situ Expression of Acidic and Thermophilic Carbohydrate Active Enzymes by Filamentous Fungi

Annika Mosier* (annika.mosier@gmail.com),[1] Christopher Miller,[1] Robin Ohm,[2] Igor Grigoriev,[2] Chongle Pan,[3] Brian Thomas,[1] Robert Hettich,[3] Steven Singer,[4] and Jill Banfield[1]

[1]University of California, Berkeley, Berkeley, California; [2]DOE Joint Genome Institute, Walnut Creek, California; [3]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [4]Lawrence Berkeley National Laboratory, Berkeley, California

Microbial communities are indispensible to the study of carbon cycling and its impacts on the global climate system. We are using a well-characterized model system—acid mine drainage (AMD) biofilms—to develop approaches to systematically examine carbon cycling in communities at the molecular level. We employed genomics, transcriptomics, and proteomics to link functional activities encoded and expressed by fungi with biogeochemical processes within the ecosystem. We reconstructed the near-complete genome (27 Mbp) of the dominant fungal AMD community member, *Acidomyces richmondensis*. The *A. richmondensis* genome contains over 275 putative carbohydrate-active enzymes, including approximately 175 glycoside hydrolases (GHs) and 35 cellulases, suggesting that fungi play an important role in recycling carbon in the community. Three of these GHs are among the top ten most abundant fungal transcripts (out of 10,149 total transcripts) in a fungal streamer AMD biofilm. Among the most highly expressed GH families are those that hydrolyze the sugars mannose, trehalose, and galactose, which have been shown to be constituents of the extracellular matrix in the AMD biofilm. Some of the GHs have also been detected in the proteome of the AMD biofilm via tandem mass spectrometry, confirming *in situ* expression. Secreted *A. richmondensis* enzymes have adapted to the extremely acidic (pH <1), metal-rich (~200 mM Fe), and thermophilic (40-50°C) environment of acid mine drainage, suggesting that these GHs may prove useful as components of acid-stable enzyme cocktails for deconstruction of biomass feedstocks for bioenergy production.

Single-cell Genomics and Metagenomics of Microbial Dark Matter in U.S. Great Basin Hot Springs

Senthil K. Murugapiran* (senthil.murugapiran@unlv.edu),[1] **Jeremy A. Dodsworth,**[1] Paul Blainey,[2] Stephen R. Quake,[2] Susannah G. Tringe,[3] Tijana Glavina del Rio,[3] and Brian P. Hedlund[1]

[1]School of Life Sciences, University of Nevada, Las Vegas, Nevada; [2]Department of Bioengineering, Stanford University, Stanford, California; [3]DOE Joint Genome Institute, Walnut Creek, California

Coordinated metagenomic and single-cell genomics have made it possible to access the genomes and predict the metabolic capabilities of "microbial dark matter," microbes representing candidate phylum- or class-level groups of bacteria and archaea that are recalcitrant to laboratory cultivation. Previous 16S rRNA gene-based microbial censuses of sediments from two US Great Basin hot springs, Great Boiling Spring (GBS), Nevada, and Little Hot Creek 4 (LHC4), California, revealed that both had abundant populations of microbial dark matter. Cells were separated from spring sediments by Nycodenz density centrifugation, loaded into a polydimethylsiloxane (PDMS) microfluidic device, and sorted by morphology using optical tweezers. Lysis and whole genome amplification were carried out in the microfluidic device, and the resulting product was screened by PCR using primers specific for bacterial and archaeal 16S rRNA genes. Over half of the sorted cells for which PCR product was obtained had 16S rRNA genes with <85% identity to those of cultured microbes, thus likely representing novel-class or phylum-level groups. Follow-up cell sorting based on morphology allowed us to obtain multiple cells representing strain- to species-level groups in the candidate bacterial phylum OP9 (22 cells). Amplified single-cell genomes from these and other novel groups were sequenced using the 454 FLX platform with Titanium chemistry, and metagenomes from sediments and in situ cellulolytic enrichments from the host springs were sequenced using both 454 and Illumina platforms. Coordinated analysis of the resulting datasets was synergistic: metagenomic reads and contigs served as scaffolds facilitating assembly of single cell genomic data, which in turn facilitated binning and phylogenetic identification of subsets of metagenomic data. Preliminary analyses from this combined dataset gives insight into the phylogeny of these dark matter groups and the potential roles that they play in element cycling and biomass degradation in their host environments.

New Species of Heterococcus: Lifecycle, Lipid, and Genome Analysis

David R. Nelson* (nels5133@umn.edu),[1] Sinafik Mengitsu, Gail Celio, Mara Mashek,[2] Douglas Mashek,[2] and Paul A. Lefebvre[1]

[1]Department of Plant Biology, University of Minnesota, St. Paul, Minnesota; [2]Department of Food Sciences and Nutrition, University of Minnesota, St. Paul, Minnesota

A new species of *Xanthophyceae*, *Heterococcus coloradii Nelson*, was discovered among snow fields in the Rocky Mountains. Axenic cultures of *H. coloradii* were prepared, and their cellular morphology, growth, and accumulation of lipids were characterized. *H. coloradii* was found to grow at temperatures approaching

freezing and to accumulate large intracellular stores of lipids. Of particular interest was the accumulation of several long-chain polyunsaturated fatty acids known to be important for human nutrition such as eicosapentaenoic acid and palmitoleic acid. Algae that accumulate lipids in this manner have potential uses as sources of biofuels and poly-unsaturated fatty acids for human nutrition. In order to study *H. coloradii*'s repertoire of genes, genomic DNA was extracted and sequenced with the Illumina GAIIx. 72 base-pair reads were organized into a draft genome of 170 mega base-pairs with 20x coverage. Over 20,000 unique protein hits were received with a MegaBLASTp using the translated draft genome as a query. Many genes were found that are involved in lipid metabolism and cold tolerance, thus highlighting the unique biology of *H. coloradii*.

Diverse Life Styles Encoded in the Genomes of Eighteen *Dothideomycetes*

Robin A. Ohm* (raohm@lbl.gov),[1] Rosie E. Bradshaw,[2] Bradford J. Condon,[3] Nicolas Feau,[4] Bernard Henrissat,[5] Benjamin A. Horwitz,[6] Conrad L. Schoch,[7] B. Gillian Turgeon,[3] Andrea Aerts,[1] Kerrie Barry,[1] Alex Copeland,[1] Braham Dhillon,[4] Fabian Glaser,[8] Jane Grimwood,[1] Cedar Hesse,[9] Idit Kosti,[6,8] Kurt LaButti,[1] Erika Lindquist,[1] Steve Lowry,[1] Susan Lucas,[1] Robert Otillar,[1] Asaf A. Salamov,[1] Jeremy Schmutz,[1] Hui Sun,[1] Lynda Ciuffetti,[9] Richard C. Hamelin,[4] Gert Kema,[10] Christopher Lawrence,[11] Joey Spatafora,[9] Pierre J.G.M. de Wit,[10] Shaobin Zhong,[12] Stephen B. Goodwin,[13] and Igor V. Grigoriev[1]

[1]DOE Joint Genome Institute, Walnut Creek, California; [2]Institute of Molecular BioSciences, College of Sciences, Massey University, New Zealand; [3]Department of Plant Pathology & Plant-Microbe Biology, Cornell University, Ithaca, New York; [4]Faculty of Forestry Forest Sciences Centre, UBC Vancouver, BC, Canada; [5]Lab Architecture et Fonction des Macromolécules Biologiques, Aix-Marseille Université, Marseille, France ; [6]Department of Biology, Technion - IIT, Haifa, Israel; [7]NIH/NLM/NCBI, Bethesda, Maryland; [8]BKU, Technion - IIT, Haifa, Israel; [9]Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon; [10]Wageningen University and Research Centre, Wageningen, Netherlands; [11]Virginia Bioinformatics Institute & Department of Biological Sciences, Bioinformatics Facility I, Blacksburg, Virginia; [12]College of Agriculture, Food Systems, and Natural Resources, NDSU Department 2200, Fargo, North Dakota; [13]USDA–Agricultural Research Service, Purdue University, West Lafayette, Indiana

The *Dothideomycetes* class of fungi includes many pathogens that infect a broad range of plant hosts. Here, we compare genome features of 18 different members of this class, including 6 necrotrophs, 9 (hemi)biotrophs and 3 saprotrophs, and discuss genome structure, evolution, and the diverse strategies of pathogenesis. The 18 genome sequences show dramatic variation in size due to variation in transposon expansions, but less variation in core gene content. During evolution, gene order in these genomes is changed mostly within boundaries of chromosomes by a series of inversions often surrounded by simple repeats. This is in contrast to major interchromosomal rearrangements observed in other groups of genomes. Several *Dothideomycetes* contain gene-poor and TE-rich putatively dispensable chromosomes of unknown function. In the current set of organisms, biotrophs and hemibiotrophs are mostly phylogenetically separated from necrotrophs and saprobes, which is also reflected in differences between gene sets represented in each group. The 18 *Dothideomycetes* offer a rich catalogue of genes involved in cellulose degradation, proteolysis, Cys-rich small secreted proteins and secondary

metabolism, many of which are enriched in proximity of transposable elements, suggesting faster evolution because of both TE mobility and RIP effects.

Using Comparative Genomics to Examine Mechanisms of Mycoparasitism in the Genus *Elaphocordyceps*

C. Alisha Owensby* (owensbyc@science.oregonstate.edu), Kathryn Bushley, and Joseph W. Spatafora

Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon

The order Hypocreales is characterized by a dynamic evolutionary history of interkingdom host jumping, with members that parasitize animals, plants, and other fungi. The monophyly of taxa attacking members of the same kingdom has not been supported by molecular phylogenetics. For example, *Trichoderma spp.* and *Elaphocordyceps spp.* are both mycoparasitic, but are members of two different families within the Hypocreales, the Hypocreaceae and Ophiocordycipitaceae, respectively. In fact, both species are more closely related to insect pathogens, than they are to each other. Three species of *Trichoderma* have sequenced genomes, and more recently the genomes of several insect pathogens in the Hypocreales have been published (e.g. *Cordyceps militaris*, *Metarhizium anisopliae, M. acridum,*) and others are in progress (e.g., *Tolypocladium inflatum*). Comparative genomics of these taxa is providing insights into mechanisms of host specificity and evolution of secondary metabolism. The genus *Elaphocordyceps* is a group within this clade that parasitizes ectomycorrhizal truffles in the genus *Elaphomyces*, representing a yet unsampled ecology within the order. To compare the gene space of a truffle pathogen with closely related insect pathogens and more distantly related mycoparasites, we have sequenced the genome of *Elaphocordyceps ophioglossoides*. Our draft assembly of the *E. ophioglossoides* genome has a genome size of approximately 32 MB, which is similar in size to the genome of the very closely related Coleopteran pathogen *T. inflatum* (*=Elaphocordyceps subsessilis;* 30.5 MB). Here, we present an initial inventory of a group of nonribosomal peptide synthetases, known as peptaibols, found in *E. ophioglossoides*. Peptaibols, which form pores in lipid bilayers disrupting osmoregulation, are indicated as being important in mycoparasitism and have been best described from *Trichoderma spp.*. *E. ophioglossoides* has 3 putative peptaibol synthetases, the most reported from any fungus sequenced to date. While some of the adenylation domains in these genes appear to be closely related between *E. ophioglossoides* and *Trichoderma spp.*, there also appear to have been lineage specific expansions in *E. ophioglossoides*, combining to make 1 large (16 modular) peptaibol synthetase-like gene and 2 smaller (10 modular) genes. Interestingly *T. inflatum*, the insect pathogenic congener of *E. ophioglossoides*, has 3 smaller (13, 11, and 8 modular) putative peptaibol synthetases. Our work represents the first genome sequence of an *Elaphocordyceps* species infecting other fungi, and will hopefully provide insights into the mechanisms underlying mycoparasitism.

Analysis of Bacterial Community Structure in Two Different Red Sludge Treatments Using 454 Junior Pyrosequencing

Tae-Jin Park*(tjpark1@hku.hk),[1] Weijun Ding,[2] Shaoan Cheng,[2] and Frederick C. Leung[1]

[1]School of Biological Sciences, The University of Hong Kong, Hong Kong, China; [2]State Key Laboratory of Clean Energy Utilization, Department of Energy Engineering, Zhejiang University, Hangzhou, China

Currently, there is difficulty in the disposal or utilization of red sludge due to its fineness of solid particles and large quantity. In the present study, two different approaches, anaerobic-light and single-chamber air cathode microbial fuel cell were used for red sludge treatment. We investigated the bacterial community structure and abundance in aqueous solutions from the two different red sludge treatments using 454 junior pyrosequencing. Sequencing output was 12,335 reads from anaerobic-light treated red sludge (average length of 465 nt) and 4.448 reads from red sludge after the MFC treatment (average length of 441 nt). The variable region V1-V3 of the 16s rRNA genes was targeted. There was a marked high in the bacterial diversity in MFC treated red sludge sample compared to the anaerobic-light treated group. The anaerobic-light treated group was dominant in putative heavy metal-resistant and denitrifying bacteria such as *Synechococcus*, *Pseudomonas* and *Delftia*. In the MFC treated sample, a diversification and increase in abundance of putative electricity-producing bacterial communities including *Rhodopseudomonas*, *Bradyrhizobium*, *Arcobacter* and *Ruminococcus*, accounted for approximately 45 % of the total population. However, there was significant low in denitrifying γ- and β – proteobacteria. Overall, the results reflect changes in the bacterial community of red sludge in response to two different treatment environments.

A Method for Building a Flux Balance Analysis (FBA) Simulation Model Using JGI IMG/ER and BioCyc Pathologic and MetaFlux Tools

Karen Parker* (Karen.parker2000@gmail.com)**,** Jason Smith, and Nick Welschmeyer

Moss Landing Marine Labs, Moss Landing, California

The overall objective of the Flux Balance Analysis (FBA) model is to better understand the metabolic systems associated with the accumulation of triacylglycerides (TAGs) in the marine diatom *Thalassiosira pseudonana* for algal biofuel applications.

The MetaFlux FBA tool from SRI is used to simulate the steady state metabolic fluxes of metabolites under different environmental conditions. It can also be used to simulate the effect of knocking out genes. Before the FBA model can be developed, an accurate metabolic network needs to be constructed: probable enzymes need to be assigned to reactions, protein complexes need to be defined, modified proteins need to be identified, transport reactions need to be inferred and pathway holes need to be filled. JGI IMG/ER is used in conjunction with SRI BioCyc Pathologic to refine and construct the metabolic network. Published

experimental transcriptomic data (Mock 2008) and biomass data (Yu 2009) is used to refine and validate the model. This poster focuses on the methodology used to build a FBA model.

Latendresse, M., M. Krummenacker, et al. (2012). "Construction and completion of flux balance models from pathway databases." Bioinformatics.

Markowitz, V. M. (2012). "IMG: the integrated microbial genomes database and comparative analysis system." Nucleic Acids Res 40(D115-22).

Mock, T. (2008). "Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioprocesses." PNAS 105(5): 1579-1584.

Yu, E. (2009). "Triacylglycerol accumulation and profiling in the model diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* (Baccilariophyceae) during starvation." J Appl Phycol DOI 10.

Sequencing and Data Analysis of Prokaryotic 5'-Transcrept Ends

Ze Peng* (zpeng@lbl.gov)**,** Zhong Wang, Konstantinos Mavrommatis, and Feng Chen

DOE Joint Genome Institute, Walnut Creek, California

Using next generation sequencing technology has dramatically changed the way gene expression profiles are studied. However, prokaryotes' 5'-transcripted ends studies have not been performed until recently owing to the technical difficulties in enriching for mRNAs that lack poly(A) tails. We have developed a simple process of sequencing and data analysis of prokaryotic 5'-transcrept ends, including removal of rRNA by subtractive hybridization and /or exonuclease digestion; changing 5′ end of triphosphorylated RNA to a 5′ monophosphate RNA; ligation of illumina sequencing adapter P1, RT-PCR, MiSeq sequencing and data analysis using programmed Galaxy work flow. Our results presented in this study provide novel insight into the transcriptional machinery as well as better understanding of cell biology.

Discovery of Microbial Biocatalysts by Improved Enzyme Prediction

Hailan Piao*,[1] Jeff Froula,[2] Zhong Wang,[2] Hans-Peter Klenk,[3] and **Matthias Hess (**Matthias.Hess@Tricity.WSU.edu)[1,2,4]

[1]Washington State University, Systems Biology & Applied Microbial Genomics, Richland, Washington; [2]DOE Joint Genome Institute, Microbial Genomics Group, Walnut Creek, California; [3]German Collection of Microorganisms and Cell Cultures (DSMZ), Braunschweig, Germany; [4]Pacific Northwest National Laboratory, Chemical and Biological Process Development Group, Richland, Washington

DNA sequencing has been used successfully to enhance our understanding of the genetic information that is encoded in microbial genomes and to identify several microbial enzymes of industrial importance (e.g. enzymes that convert plant biomass into biofuel precursors or that render plants resistant to pathogens).

Despite this success, a majority of the genes in a genome are currently annotated as genes that encode "hypothetical proteins without assigned function". To overcome this technical limitation and to identify microbial enzymes that facilitate the conversion of recalcitrant biomass into biofuel precursors, we developed an improved gene annotation algorithm. In the study presented here, we employed this gene annotation algorithm to search JGI's Integrated Microbial Genomes (IMG) database, a data repository that contains all finished microbial genomes, for genes that encode novel biomass degrading enzymes. We identified 62 genes that possess a cellulase-specific fingerprint and that have not been identified as "cellulases" by currently available annotation algorithms. A large fraction of the identified cellulase candidates has already been cloned for heterologous expression and enzymatic activity tests to verify the biomass-degrading activity of the recombinant proteins are currently underway.

Gene Ontology Terms Describe Biological Production of Methane

Endang Purwantini* (epurwant@vt.edu),[1] Trudy Torto-Alalibo,[1] Joao C. Setubal,[1,2] Brett M. Tyler,[1,3] and **Biswarup Mukhopadhyay**[1]

[1]Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Virginia; [2]Department of Biochemistry, Institute of Chemistry University of São Paulo, Brazil; [3]Center for Genome Research and Biocomputing, Oregon State University, Corvallis, Oregon

The MENGO consortium in collaboration with the community of microbiologists engaged in bioenergy research and the Gene Ontology (GO) consortium aims to develop a comprehensive set of Gene Ontology terms that will describe bioenergy related biological processes and to annotate relevant microbial genomes with appropriate GO terms. 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 bio-energy production. Such terms will aid in the comprehensive annotation of relevant genes in diverse microbial genomes. Among the 200 terms developed so far are a comprehensive set that describes processes involved in the biological production of methane/methanogenesis.

Biologically, methane is generated by methanogenic archaea from H2 + CO_2, secondary alcohol + CO_2, formate, carbon monoxide, acetate, methanol, methylamines, and methanethiols. Pathways for methanogenesis from these substrates use unusual coenzymes such as coenzyme F_{420}, methanofuran, tetrahydromethanopterin, coenzyme M, cofactor F_{430}, and coenzyme B. Methanogenesis allows efficient mineralization of biological polymers in anaerobic niches of nature and thereby plays an important role in carbon cycle. This integrated process is leveraged for the production of methane from renewable resources and for waste treatment. The MENGO team has created some terms that are useful for describing the biological processes allowing methanogenesis from carbohydrates, including the biosynthesis of relevant coenzymes. Additionally, methanogenesis related gene products of certain methanogenic archaea such as *Methanocaldococcus jannaschii*, *Methanosarcina barkeri*, *Methanosarcina thermophilla*, *Methanosaeta concilii*, *Methanopyrus kandleri*, and

Methanothermobacter marburgenensis have been manually annotated with GO terms.

Funding for the MENGO project is provided by the Department of Energy as part of the Systems Biology Knowledgebase program - grant# DE-SC000501.

Adaptive Evolution of *Saccharomyces cerevisiae* and Elucidation of Enhanced Tolerance to High Temperature and Inhibitory Compounds by Transcriptome Analysis

Xianqi Qi, Baowei Wang, Yangfeng Peng, Ran Tu, Qinhong Wang*****
(wang_qh@tib.cas.cn), and Yanhe Ma

Key Laboratory of Systems Microbial Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, China

To improve cellulosic ethanol yield and develop the cost-effective simultaneous saccharification fermentation for ethanol production, the strain should be developed with improved tolerance to high temperature and inhibitory compounds from lignocellusic biomass. Here, an industrial strain of *S. cerevisiae* Ethanol Red E491 (Ethanol Performance Group, NJ, USA) was evolved using corn cob hydrolyzate at step-increased temperature (from 37-42°C) and the resulting daughter strain, TIB-S.C Y01, produced ethanol much more rapidly than its parent in fermentations of corn cob hydrolyzate at 40°C. Adaptation improved fermentation performance of the evolved strain. Based on transcriptome analysis of parent and evolved strains via RNA sequencing, we find evidence of parallel evolution in Bsc1p, Hxk2p, Pfk27p, Mer1p, Gnd1p, and Tkl1p genes. Consistent with the complex, multigenic nature of multiple stresses, we observe adaptations in a diversity of cellular processes. Many adaptations appear to involve epistasis between different mutations, implying a rugged fitness landscape for the tolerance of high temperature and inhibitory compounds in yeast.

Comparative and Functional Genomic Analyses Reveal the Genetic Basis of a Fungal Lignocellulolytic Enzyme System Improved by Artificial Selection

Yinbo Qu* (quyinbo@sdu.edu.cn),[1,3] Guodong Liu,[1] Lei Zhang,[2] Yuqi Qin,[1,3] Gen Zou,[2] Zhonghai Li,[1] Xing Yan,[2] Xiaomin Wei,[1] Mei Chen,[1] Ling Chen,[2] Kai Zheng,[1] Jun Zhang,[2] Liang Ma,[2] Jie Li,[1] Rui Liu,[2] Hai Xu,[1] Xiaoming Bao,[1] Xu Fang,[1,3] Lushan Wang,[1] Yaohua Zhong,[1] Weifeng Liu,[1] Huajun Zheng,[4] Shengyue Wang,[4] Chengshu Wang,[2] Guo-Ping Zhao,[2,4] Tianhong Wang,[1] and Zhihua Zhou[2]

[1]State Key Laboratory of Microbial Technology, and [3]National Glycoengineering Research Center, Shandong University, Shandong University, Jinan, Shandong, China; [2]Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China; [4]Shanghai-MOST Key Laboratory of Disease and Health Genomics, Chinese National Human Genome Center at Shanghai, Shanghai, China

The whole genome of *Penicillium decumbens* was sequenced and assembled into nine scaffolds of 30.2 Mb, putatively representing eight chromosomes and one circular mitochondrial DNA. Compared with the genome of the widely used

cellulase-producing fungus *Trichoderma reesei*, the *P. decumbens* genome encodes a set of lignocellulolytic enzymes with more diverse components, particularly for cellulose binding domain-containing proteins and hemicellulases. A comparative systems biological analysis of the wild-type strain versus a cellulase hyper-producing mutant acquired through more than 30 years of mutagenesis and screening revealed the genetic basis of this improvement. Mutations in crucial transcription factors and promoter regions of related genes have resulted in a more specific and efficient enzyme system for lignocellulose degradation. The results also point out directions for further development of synergistic enzyme systems for the effective conversion of various lignocellulosic materials into fermentable sugars.

Efficient Prediction of Fungal Genes

Ian Reid* (ian_reid@gene.concordia.ca),[1] Paul M.K. Gordon,[2] Nicholas O'Toole,[1] Mike Dahdouli,[2] Mostafa Abdellateef,[2] Greg Butler,[1] Christoph W. Sensen,[2] and Adrian Tsang[1]

[1]Centre for Structural and Functional Genomics, Concordia University, Montreal, Quebec, Canada**;** [2]Visual Genomics Centre, University of Calgary, Calgary, Alberta, Canada

We are searching the genomes of thermophilic and thermotolerant fungi for useful enzymes. This effort includes sequencing the DNA of the fungi, assembling the genomes, predicting protein-coding genes, identifying, cloning and expressing interesting proteins, and characterizing the cloned proteins. Gene prediction became a bottleneck, and we developed the Snowy Owl pipeline to provide quick and accurate predictions. We tested various combinations of predictors and chose those that provided the best predictions at lowest computational cost.

Snowy Owl combines the existing *ab-initio* gene prediction programs GeneMark ES and Augustus with a novel model scoring and merging process that relies heavily on RNA-Seq transcriptome data to select the best predictions at each location in the genome. Augustus modeling is seeded with GeneMark predictions and contigs containing long ORFs assembled from RNA-Seq reads. Augustus trained on the initial models and using RNA-Seq hints generates a wide variety of gene predictions. These predictions are scored for homology to known proteins, structural soundness, and compatibility with exon locations indicated by RNA-Seq data. The set of non-overlapping models with the highest total score is selected. In addition to the high-quality "plausible" gene models, we keep track of flawed "implausible" models at locations where no better models are available, as inputs for future manual or computational improvement.

We have successfully applied the Snowy Owl pipeline to 20 ascomycete, basidiomycete, and zygomycete genomes so far; we will present results on *Aspergillus niger* and *Phanerochaete chrysosporium,* for which previous gene predictions are available on the JGI portal, and on *Thermomyces lanuginosus,* newly sequenced within our project. Comparison of the pipeline predictions to manually curated gene models (*ca*. 2000 in each genome) indicates high sensitivity and selectivity. The new transcriptome information from RNA-Seq permits refinement of earlier genome annotations.

Comparative Genome Analysis of Basidiomycete Fungi Reveals the Genetic Signatures pf Wood Degraders

Robert Riley* (rwriley@lbl.gov),[1] Asaf Salamov,[1] Emmanuelle Morin,[2] Laszlo Nagy,[3] Gerard Manning,[4] Scott Baker,[5] Daren Brown,[6] Bernard Henrissat,[7] Anthony Levasseur,[7] David Hibbett,[3] Francis Martin,[2] and Igor Grigoriev[1]

[1]DOE Joint Genome Institute, Walnut Creek, California; [2]Institut National de la Recherche Agronomique, France; [3]Department of Biology, Clark University, Worcester, Massachusetts; [4]Razavi-Newman Center for Bioinformatics, Salk Institute for Biological Sciences, San Diego, California; [5]Pacific Northwest National Laboratory, Richland, Washington; [6]USDA, ARS, MWA, NCAUR, BFPM, Peoria, Illinois; [7]AFMB UMR 6098 CNRS/UI/UII, Marseille, France;

Fungi of the phylum Basidiomycota (basidiomycetes), make up some 37% of the described fungi, and are important in forestry, agriculture, medicine, and bioenergy. This diverse phylum includes the mushrooms, wood rots, symbionts, and plant and animal pathogens. To better understand the diversity of phenotypes in basidiomycetes, we performed a comparative analysis of 35 basidiomycete fungi spanning the diversity of the phylum. Phylogenetic patterns of lignocellulose degrading genes suggest a continuum between the white rot and brown rot modes of wood decay. Patterns of protein kinases, secondary metabolic enzymes, and secreted proteins shed additional light on the broad array of phenotypes found in the basidiomycetes. We suggest that the lignocellulose gene content of an organism can be used to predict its nutritional mode, and predict *Dacryopinax sp*. as a brown rot; *Botryobasidium botryosum* and *Jaapia argillacea* as white rots.

Transcriptome Analysis of the Ectomycorrhizal Fungus *Paxillus involutus* Provide Insights into the Mechanisms of Organic Matter Degradation

Francois Rineau,[1] Doris Roth,[2] Firoz Shah,[1] Mark Smits,[3] Tomas Johansson,[1] Björn Canbäck,[1] Peter Bjarke Olsen,[4] Per Persson,[5] Morten Nedergaard Grell,[2] Erika Lindquist,[6] Igor V. Grigoriev,[6] Lene Lange,[2] and **Anders Tunlid*** (anders.tunlid@biol.lu.se)[1]

[1]Department of Biology, Microbial Ecology Group, Lund University, Lund, Sweden; [2]Department of Biotechnology and Chemistry, Aalborg University, Ballerup, Denmark; [3]Centre for Environmental Sciences, Hasselt University, Diepenbeek, Belgium; [4]Novozymes, Bagsvaerd, Denmark; [5]Department of Chemistry, Umeå University, Umeå, Sweden; [6]DOE Joint Genome Institute, Walnut Creek, California

Soils in boreal forests contain large stocks of carbon. Plants are the main source of this carbon through tissue residues and root exudates. A major part of the exudates are allocated to symbiotic ectomycorrhizal fungi. In return, the plant receives nutrients, in particular nitrogen from the mycorrhizal fungi. To capture the nitrogen, the fungi must at least partly disrupt the recalcitrant organic matter-protein complexes within which the nitrogen is embedded. This disruption process is poorly characterized. We examined how the ECM fungus *Paxillus involutus* degrade organic litter materialusing various methods including elemental analysis, FTIR, pyrolysis-GC/MS, and synchronous fluorescence. In parallel, the global pattern of gene expression was analyzed using 454 pyrosequencing and microarray analysis. The fungus partially degraded polysaccharides and modified the structure of polyphenols. The observed chemical changes were consistent with a hydroxyl

radical attack, involving Fenton chemistry similar to that of brown-rot fungi. The set of enzymes expressed by *Pa. involutus* during the degradation of the organic matter was similar to the set of enzymes involved in the oxidative degradation of wood by brown-rot fungi. However, *Pa. involutus* lacked transcripts encoding extracellular enzymes needed for metabolizing the released carbon. The saprotrophic activity has been reduced to a radical-based biodegradation system that can efficiently disrupt the organic matter-protein complexes and thereby mobilize the entrapped nutrients. We suggest that the released carbon then becomes available for further degradation and assimilation by commensal microbes, and that these activities have been lost in ectomycorrhizal fungi as an adaptation to symbiotic growth on host photosynthate.

Providing a Glimpse into the Coding Potential of Microbial Dark Matter

Christian Rinke* (crinke@lbl.gov),[1] Alex Sczyrba,[1] Natalia Ivanova,[1] Jan-Fang Cheng,[1] Janey Lee,[1] Stephanie Malfatti,[1] Ramunas Stepanauskas,[2] Jonathan A. Eisen,[1,3] Steven Hallam,[4] William P. Inskeep,[5] Brian P. Hedlund,[6] Stefan M. Sievert,[7] Wen-Tso Liu,[8] George Tsiamis,[9] Nikos Kyrpides,[1] Eddy Rubin,[1] Philip Hugenholtz,[10] and Tanja Woyke[1]

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

Genome sequencing enhanced our understanding of the metabolic capabilities and phylogenetic relations of the microbial world. The vast majority of bacterial and archaeal genomes sequenced to date are however of rather limited phylogenetic diversity as they were chosen primarily based on their physiology, reflecting the ability to be grown in culture. Since more than 99% of microorganisms are estimated to elude current culturing attempts, culture independent methods are needed to recover genomes of these largely mysterious species comprising the microbial dark matter. Applying single cell genomics we target 100 single cell representatives of uncultured Bacteria and Archaea from phylogenetic novel lineages including candidate phyla. The assembled genomes will allow us to improve our understanding of phylogenetic relations and the evolutionary diversification of microbes. Single cell genomes will also enable us to investigate the functional capabilities of these novel lineages facilitating the search for novel genes, protein families, and pathways. Furthermore we will explore the combination of single cell genomes with metagenomic data sets to screen for population heterogeneities within microbial communities, to improve phylogenetic anchoring of metagenomic data, and to attempt co-assemblies of single cells and metagenomes in order to improve single cell genome assemblies.

Construction of Illumina Sequencing Libraries for Fungal Phylogenomics and Community Metagenomics from Nanogram Amounts of Nucleic Acid

Marianela Rodriguez-Carres* (mr88@duke.edu),[1] Gregory Bonito,[1] Teresita M. Porter,[2] Andrii Gryganskyi,[1] Timothy Y. James,[3] Hui-Ling Liao,[1] and Rytas Vilgalys[1]

[1]Department of Biology, Duke University, Durham, North Carolina; [2]Department of Biology, McMaster University, Hamilton, Ontario, Canada; [3]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan

Next generation sequencing technologies are rapidly being adopted to great benefits for studies on fungal phylogenetic biology and environmental metagenomics. Most protocols for next generation sequencing technology still require microgram amounts of RNA and DNA. However, for many kinds of fungal diversity research, it is difficult to obtain sufficient amounts of high molecular weight genomic DNA and intact RNA, particularly from small amounts of tissue. In the current study we evaluated modified NuGen protocols for the generation of Illumina sequencing libraries from different types of fungal tissue of various quantities and qualities. In order to determine the minimum range of RNA needed to assess the functional diversity of the rhizosphere community of woody plants we constructed RNA libraries from root samples of *Populus* and *Pinus*. To test the effect of the amount and quality of genomic DNA in node resolution we constructed libraries from high molecular weight and also partially degraded genomic DNA from several basal fungal species (*Basidiobolus meristosporus*, *Chytriomyces hyalinus, Polychytrium aggregatum, Chytriomyces angularis, Hyaloraphidium curvatum, Monoblepharella sp.,* and *Rhizophydium sphaerotheca*). Preliminary results suggest that libraries created with as little as 1ng of total RNA and 200ng of genomic DNA generated data comparable to current next generation sequencing protocols requiring at least 100X more starting material.

Exploring the Cyanobacterial Tree Using Molecular Genomics

Kenneth Sauer* (KHSauer@lbl.gov)

Lawrence Berkeley National Laboratory, Berkeley, California; Chemistry Department, University of California, Berkeley, Berkeley, California

It is possible currently to construct a tree for cyanobacteria in 4 dimensions using molecular genomics from more than 40 taxa. Distances between pairs of taxa are measured as amino-acid mismatch numbers (N) between aligned sequences of photosynthetic reaction center proteins: 10 from Photosystem I and 20 from Photosystem II. A scheme is devised for projecting any multi-dimensional tree onto plots in 2-dimensions that retain the numerical distances between taxa. New insights include: (1) three major clades incorporate almost all sequenced cyanobacteria; (2) three thermophilic Synechococcus taxa are nearer to two Cyanothece taxa than to any other Synechococci; (3) a recent lateral-transfer event has occurred in the clade of Prochlorococci that may have resulted from a virus. Pitfalls exist in this approach involving mismatch numbers because of probable errors in the database, but the method provides a scheme for flagging those

problems. This phylogenetic tree differs significantly from published trees using Bayesian maximum-likelihood applied to sequences of a diverse ensemble of cyanobacterial proteins.

Roles of Genotype-by-Environment Interactions in Shaping the Root-associated Microbiome of *Populus*

Christopher W. Schadt* (schadtcw@ornl.gov),[1,2] Migun Shakya,[1,2] Neil Gottel,[1] Hector Castro,[1] Zamin Yang,[1] Marilyn Kerley,[1] Gregory Bonito,[3] Dale Pelletier,[1,2] Tatiana Karpinets,[1] Jesse Labbe,[1] Edward Uberbacher,[1,2] Wellington Muchero,[1] Francis Martin,[4] Gerald Tuskan,[1] Mircea Podar,[1,2] Rytas Vilgalys,[3] and Mitchel J. Doktycz[1,2]

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]University of Tennessee, Knoxville, Tennessee; [3]Duke University, Durham, North Carolina; [4]Institut National de la Recherche Agronomique, Nancy Université, France

Populus trees represents a genetically diverse, ecologically widespread riparian genus, that have potential as cellulosic feedstocks for biofuels, and contain the first tree species to have a full genome sequence. These trees are also host to a wide variety of symbiotic microbial associations within their roots and rhizosphere, thus may serve as ideal models to study the breadth and mechanisms of interactions between plants and microorganisms. However, most of our knowledge of *Populus* microbial associations to date comes from greenhouse and plantation-based trees; there have been no efforts to comprehensively describe microbial communities of mature natural populations of *Populus*. We have compared root endophyte and rhizosphere samples collected from two dozen sites within two watershed populations of *Populus deltoides* in Tennessee and North Carolina over multiple seasons. 454 pyrosequencing has been applied to survey and quantify the microbial community associated with *P. deltoides*, using primers targeting the bacterial 16S rRNA gene and the fungal 28S rRNA gene. Genetic relatedness among the *Populus* trees was evaluated using 20 SSR markers chosen for distribution across all 19 linkage groups of the *Populus* genetic map. Soil physical, chemical and nutrient status, as well as tree growth and age characteristics were also evaluated. Root endosphere and rhizosphere communities have been found to be composed of distinct assemblages of bacteria and fungi with largely non-overlapping OTU distributions. Within these distinct endophyte and rhizosphere habitats, community structure is also influenced by soil characteristics, watershed origin and plant genotype; while observed seasonal influences have been minimal. We have isolated cultures of over a thousand bacteria and fungi from these environments representing most of the dominant community members *insitu*. Many of these isolates show distinct growth-promoting phenotypes with *Populus*. These findings indicate that the characteristics of the *Populus* root/soil environment may represent a relatively strong selective force in shaping endophyte and rhizosphere microbial communities and their functions may have great importance upon the success of *Populus* sp. Forthcoming work in collaboration with JGI will explore more in depth the genetic basis of these associations within a common garden populations of *P. trichocarpa* containing over >1000 resequenced variants.

Emerging SMRT® Sequencing Technologies Facilitate Extended Readlength

Robert Sebra* (rsebra@pacificbiosciences.com), Meredith Ashby, Aruna Bhamidipati, Keith Bjornson, Colleen Cutcliffe, Ravi Dalal, John Eid, Andrei Fedorov, Jeremy Gray, Jeremiah Hanes, Pei-Lin Hsiung, David Hsu, Satwik Kamtikar, Feruz Kurbanov, John Lyle, Sarah McCalmon, Erik Miller, Emilia Mollova, Devon Murphy, Paul Peluso, Dave Rank, Kevin Travers, Sophia Wu, and Alicia Yang

Pacific Biosciences, Menlo Park, California

Pacific Biosciences® has introduced an improved sequencing chemistry that provides longer readlengths, higher throughput, and achieves consensus accuracy at lower coverage. Example data presented here shows mean readlengths up to 2400 bases using 45-minute and 3000 bases using 90-minute continuous sequencing. With faster enzyme kinetics, flexibility to formulate SMRTbell™ libraries with insert sizes up to 10,000 bases, and improved polymerase-SMRTbell complex stability, the new sequencing chemistry increases the mean readlength as compared to the current commercial kit, increasing throughput. With 5% of single enzymes yielding readlengths beyond 7500 bases, complex contiguous regions can be thoroughly assembled using less fold coverage of individual reads as long as 15,000 bases. Further, using 250 to 1000 base insert SMRTbell libraries in combination with this new sequencing chemistry, we demonstrate Circular Consensus Sequencing (CCS) within each individual molecule. As the number of passes of sequencing increases, single molecule accuracies as high as QV40 are achieved. Circular consensus sequencing affords both long readlength and higher accuracy for efficient detection of structural variants important for applications such as minor allele detection. A variety of results will be presented to demonstrate the enhanced performance of long insert and CCS sequencing technologies. While this substantial readlength continues to expand our application space, we continue to push our technology development towards higher throughput by further extending readlength while continuing to optimize our throughput and consensus accuracy. We also present results from various research endeavors focused on SMRTbell-polymerase complex purification, enhancing complex loading efficiency, and research improving signal-to-noise ratio leading to improved detection and accuracy.

MycoCosm—A Web-based Interactive Fungal Genomics Resource

Igor Shabalov* (ishabalov@lbl.gov)and Igor Grigoriev

DOE Joint Genome Institute, Walnut Creek, Califonira

MycoCosm is a web-based interactive fungal genomics resource, which was first released in March 2010, in response to an urgent call from the fungal community for integration of all fungal genomes and analytical tools in one place (Pan-fungal data resources meeting, Feb 21-22, 2010, Alexandria, VA). MycoCosm integrates genomics data and analysis tools to navigate through over 100 fungal genomes sequenced at JGI and elsewhere. This resource allows users to explore fungal genomes in the context of both genome-centric analysis and comparative genomics, and promotes user community participation in data submission,

annotation and analysis. MycoCosm has over 4500 unique visitors/month or 35000+ visitors/year as well as hundreds of registered users contributing their data and expertise to this resource. Its scalable architecture allows significant expansion of the data expected from JGI Fungal Genomics Program, its users, and integration with external resources used by fungal community.

MycoCosm is described as a featured article in the next database issue of Nuclear Acids Research (Grigoriev et al, NAR, 2011).

The Organization and Regulation of Carbon Partitioning Pathways in Diatoms

Sarah R. Smith* (sarahsmith@ucsd.edu), Raffaela M. Abbriano, and Mark Hildebrand

Scripps Institution of Oceanography, La Jolla, CA

Diatoms are one of the most ecologically successful groups of eukaryotic microalgae, and their ability to synthesize large stores of triacylglycerol (TAG) under certain conditions make them a desirable feedstock for the production of algae-based renewable biofuels. However, a major challenge in the development of diatoms and other microalgal strains for large-scale production is to optimize biomass accumulation and the production of fuel-relevant TAG. Genetic manipulation approaches are envisioned to contribute substantially towards achieving these goals, but success will require a fundamental understanding of the organization and regulation of carbon metabolic pathways in diatoms. Since diatoms arose via a secondary endosymbiosis, they have a combination of both animal and plant-like characteristics and may differ substantially from many known model organisms in the structure of their carbon partitioning pathways. The availability of diatom genomes is facilitating the characterization of these pathways both through bioinformatics-based targeting predictions and the construction of phylogenies. A comparative analysis of the genes involved in carbon partitioning metabolic pathways (glycolysis, gluconeogenesis, pyruvate metabolism) was conducted using the genomes of *Thalassiosira pseudonana*, *Phaeodactylum tricornutum*, and *Fragilariopsis cylindrus* with the goal of identifying a conserved carbon partitioning proteome. There are several unifying features of carbon metabolism in diatoms that distinguish them from other more well-characterized organisms. However, some substantial differences exist between the organization of metabolic pathways in the centric *T. pseudonana* and the pennate diatoms. Furthermore, the proportion of unique carbon partitioning genes within a given diatom genome is non-trivial (14-17%) indicating there has been modification to these core metabolic pathways as diatoms diversified. A generalized model of diatom carbon partitioning metabolism will be presented along with a summary of the differences between species. The regulatory, physiological, and evolutionary implications of the organization of diatom carbon metabolism will be addressed.

First Core Facility for Single-Cell Genomics

Ramunas Stepanauskas, Brian Thompson* (bthompson@bigelow.org), Nicole J. Poulton, Brandon K. Swan, Elizabeth Dmitrieff, Benjamin Tupper, Wendy Bellows, Erin Field, and Michael E. Sieracki

Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine

Bigelow Laboratory Single Cell Genomics Center (SCGC) is the first shared-user facility offering single cell genomic DNA recovery services to the broad scientific community (bigelow.org/scgc). The main goal of SCGC is to serve as the engine of discoveries in the areas of microbial ecology, evolution, and bioprospecting. Robust, high-throughput protocols were established for environmental sample preservation, single cell flow-cytometric cell sorting, lysis, whole genome amplification and identification, including rigorous quality control and effective data management. Since its establishment in November 2009, SCGC has contributed to cutting-edge research projects at over 30 organizations around the globe. The types of samples processed by SCGC range from marine to deep subsurface to mammalian gut content, and the type of research questions addressed range from biogeochemistry to evolution to human health. Over 300,000 individual cells have been analyzed by our high-throughput pipeline so far, providing unique access to genomic DNA from microorganisms representing over 60 phyla of bacteria, archaea and protists. Most of these microbes have never been cultivated, and single cell genomics is providing the first view of their biological features. Center's research accomplishments include discoveries of inorganic carbon fixation pathways in abundant bacterial groups in the dark ocean, detection of in situ trophic interactions of uncultured protists, identification of novel phototrophs, and others. In 2010, the center hosted its second single cell genomics workshop, which attracted over 80 leading experts in the field, postdocs, and students.

Whole Genome Analysis of Uncultured, Ammonia-oxidizing Marine Group I Archaea from the Mesopelagic Using Single-Cell Genomics

Brandon K. Swan* (bswan@bigelow.org),[1] Mark D. Chaffin,[2] Manuel Martinez-Garcia,[1] E. Dashiell P. Masland,[1] Monica Lluesma Gomez,[1] Nicole J. Poulton,[1] Michael E. Sieracki,[1] and Ramunas Stepanauskas[1]

[1]Bigelow Laboratory for Ocean Sciences, Single Cell Genomics Center, West Boothbay Harbor, Maine; [2]Colby College, Waterville, Maine

Putative ammonia-oxidizing Marine Group I (MG-I) *Archaea* are thought to play a critical role in oceanic nitrogen and carbon cycling. However, current cultured marine representatives most likely do not capture genomic variation of uncultured populations. We employed meta- and single cell genomics to investigate the genetic diversity of uncultured mesopelagic MG-I within the South Atlantic and North Pacific Gyres. PCR-screening and whole-genome sequencing of single amplified genomes (SAGs) revealed genes supporting autotrophic and heterotrophic carbon assimilation, and genes involved in the proposed ammonia oxidation pathway of *Nitrosopumilus maritimus*. Gene arrangement and content of a high-quality draft SAG was most similar to *N. maritimus* than the sponge

symbiont *Cenarchaeum symbiosum*, and all three contained a large core genome. Fragment recruitment analysis indicated this SAG was more representative of dark ocean populations than existing marine cultures. Additionally, the SAG harbored several genomic islands, potentially providing additional adaptability to dark ocean life, and may be indicative of differences between MG-I populations from the Atlantic and Pacific oceans.

Carbon Biosequestration Potential and Microbial Stimulation by Pyrolized Carbon (Biochar) in Soil

N.Taş* (ntas@lbl.gov),[1] C. Castanha,[2] K. Reichl,[3] M. Fischer,[3] E.L. Brodie,[1] M.S. Torn,[2] and J.K. Jansson[1,4,5]

[1]Ecology Department, [2]Climate and Carbon Sciences Department, Earth Sciences Division; [3]Environmental Energy Technologies; [4]Joint Genome Institute (JGI); [5]Joint Bioenergy Institute (JBEI); Lawrence Berkeley National Laboratory, Berkeley, California

Biochar (BC) is a carbon rich product that is produced by high-temperature and low-oxygen pyrolysis of biomass. BC addition to soil has been proposed as a promising method for C sequestration and enhancing soil quality. BC has been assumed to be recalcitrant in soil, but recent research shows that it is at least partly degradable by soil microbes. However, the influence of soil chemistry and environmental conditions on microbial transformation of BC is yet to be understood. This study aims to determine the microbial controls on decomposition and soil retention of BC. We performed laboratory incubation experiments to compare the potential for BC decomposition in soils from contrasting ecosystems (subtropical forest vs grassland), temperatures (ambient and elevated) and depths (surface and deep). Soil incubations with pyrolyzed ^{13}C-wood were monitored for heterotrophic respiration, activity of extracellular enzymes, and microbial community composition. ^{13}C-CO_2 measurements confirmed that BC was degraded in both depths of grassland and subtropical soils. In both grassland and subtropical soils, compared to controls, CO_2 emissions didn't increase significantly due to BC addition. Significantly more BC was mineralized in grassland soils (7.43-8.02 mg C/ gr BC) compared to subtropical soils (6.06-7.40 mg C/ gr BC). In contrast to prevailing hypotheses about temperature response, warmer temperatures did not result in more BC mineralized to CO_2 in deeper subtropical soils; warming did increase mineralization in surface soils. Additionally, in surface soils from both grassland and subtropical soils, temperature sensitivity (Q_{10}) of native organic C decomposition decreased with BC addition. Several enzymes involved in the degradation of macromolecular C in soils, including cellobiohydolases, were significantly stimulated by BC. 16S rRNA gene sequencing of microbial communities demonstrated that *Actinomycetales* were the only bacterial group enriched by BC addition in grassland soils. However, α-Proteobacteria and *Solibacteres* (Acidobacteria) were enriched in subtropical soils. The community dynamics of the active microbes in the soil, and their expression of functional genes involved in cycling of C, are currently being assessed by sequencing of total RNA. In addition, we are using the isotopic label to trace BC metabolites into microbial cells and soil organic matter fractions. With this multi-disciplinary

approach we will further investigate the effect of BC on soil microbial metabolism and CO_2 emissions.

Microbial ENergy processes Gene Ontology (MENGO): New Gene Ontology Terms Describing Microbial Processes Relevant for Bioenergy

Trudy Torto-Alalibo* (trudy@vbi.vt.edu),[1] Endang Purwantini,[1] Joao C. Setubal,[1,2] Brett M. Tyler,[1,3] and Biswarup Mukhopadhyay[1]

[1]Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, Virginia; [2]Department of Biochemistry, Universidade de São Paulo, Brazil; [3]Center for Genome Research and Biocomputing, Oregon State University, Corvallis, Oregon

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 GO consortium was formed in 1998 to create universal descriptors, which can be used to describe functionally similar gene products and their attributes.

MENGO (http://mengo.vbi.vt.edu), an interest group of the GO consortium seeks to expand term development for microbial processes useful for bioenergy production. Currently, there are over 200 MENGO terms added to the GO. Areas covered include carbohydrate catabolic processes, oligosaccharide binding and transport, hydrogen production and methanogenesis. Additionally, over 200 GO annotations of bioenergy relevant gene products from microbes such as *Clostridium thermocellum*, *Methanosarcina barkeri*, *Bacteroides thetaiotaomicron* and *Chlamydomonas reinhardtii* have been made. A selection of terms and annotations will be highlighted in this presentation.

The MENGO interest group will also host a workshop immediately after the DOE JGI User meeting on March 23rd at the same venue. This workshop will highlight progress made in GO term development and microbial gene annotation as well as some of the challenges encountered. Additionally, we will have an open forum to hear from participants on other bioenergy areas to be targeted for further term development and microbial genomes to be annotated.

Funding for the MENGO project is provided by the Department of Energy as part of the Systems Biology Knowledgebase program.

Evaluation of Multiplexed 16S rRNA Microbial Population Surveys Using llumina MiSeq Platform

Julien Tremblay* (jtremblay@lbl.gov), Edward S. Kirton, Kanwar Singh, Feng Chen, and Susannah G. Tringe

DOE Joint Genome Institute, Walnut Creek, California

In recent years, microbial community surveys extensively relied on 454 pyrosequencing technology (pyrotags). Recently, the Illumina sequencing platform

HiSeq2000 has largely surpassed 454 in terms of read quantity and quality with typical yields of up to 600 Gb of paired-end 150 bases reads in one 18 day run. Yet many labs still rely on pyrotags for community profiling because the HiSeq throughput exceeds their needs, the run time is long, and accumulating sufficient samples to effectively utilize a full run introduces significant delay. Illumina recently introduced the new mid-range MiSeq sequencing platform which gives an output of 1 Gb of paired-end 150 base reads in a single day run. With its moderately-high throughput and support for massive multiplexing (barcoding), this platform represents a promising alternative to 454 technology to perform 16S rRNA-based microbial population surveys.

A workflow was therefore developed to confirm that Illumina MiSeq is a suitable platform to accurately characterize microbial communities. We surveyed microbial populations coming from various environments by targeting the 16S rRNA hypervariable region V4 which generated amplicons size of about 290 bases. These amplicons were sequenced with the MiSeq platform from both 5' and 3' ends followed by *in silico* assembling using their shared overlapping part. Downstream analyses through our Itags pipeline are also described, including a novel clustering strategy generating fast and accurate distribution of bacterial operational taxonomic units (OTUs).

Our results suggest that the MiSeq sequencing platform successfully recaptures known biological results and should provide a useful tool for 16S rRNA characterization of microbial communities.

Full Length Human cDNA Sequencing on the PacBio® *RS*

Jason G. Underwood* (junderwood@pacificbiosciences.com),[1] **Lawrence Lee**,[1] Tyson A. Clark,[1] Michael Brown,[1] Dale Webster,[1] Robert Sebra,[1] Sara Olson,[2] Brenton Graveley,[2] Jonas Korlach,[1] and Kevin Travers[1]

[1]Pacific Biosciences, Menlo Park, California; [2]UConn Health Center, Genetics and Developmental Biology, Farmington, Connecticut

Transcriptome experiments to date were performed by shotgun approach whereby the full-length transcripts are assembled from small cDNA fragments derived from the original RNA. These methods can infer the overall patterns of alternative events, but cannot define which combinations of events occurred on the fully intact RNA molecules. Since eukaryotic transcripts can display rich patterns of alternative events at diverse positions along the transcript, it is likely that some aspects of co-regulation remain undetected. To address this gap, we applied the long read length capabilities of the PacBio® *RS* to sequence full-length cDNAs. We sequenced libraries generated from the breast cancer line MCF7 and human cerebellum tissue to uncover many sequences spanning the entire length of known ESTs and RefSeq entries. The long read data included direct observation of a gene fusion event known to be present in MCF7 cells and the splicing patterns of several kinases in their full-length context. We also coupled the full length cDNA approach with the Agilent SureSelect platform to detect splicing patterns of the human kinome expressed in these cell types.

To further understand how highly complex splice site choices are interconnected, we generated an amplicon library that assesses the splicing patterns of mRNAs encoding the extracellular domain of the *Drosophila* DSCAM axon guidance receptor. Sequencing of this library reveals which of the putative 19,008 extracellular domain isoforms are actually expressed and whether there are co-regulation events that were not seen in previous short-read data.

A Genome Project for *Macrocystis pyrifera* (Phaeophyta), the Largest Alga on Earth

Klaus Valentin* (Klaus.Valentin@awi.de)[1] and Mark Cock[2]

[1]Afred Wegener Institute, Bremerhaven, Germany; [2]Station Biologique Roscoff, Roscoff, France

Brown macroalgae ("kelps") dominate coastal ecosystems on rocky shores from temperate to Polar Regions. They can form huge kelp beds which provide food and shelter for a multitude of marine organisms; in that they equal terrestrial forests and are sometimes called "Marine Forests". The biology of brown algae is well studied and recently also their molecular features became a focus of marine genomics: the genome of the first brown algae, *Ectocarpus siliculosus*, has been published, large scale transcriptomic studies on *Saccharina latissima* have been completed.

Kelps are also of growing economic importance. Serving as food for humans since centuries they today provide a mulitde of substances for industry. Just recently large kelps were recognized as possible sources of biomass for biofuel production. The most promising genus is *Macrocystis*, the "giant kelp". These species can reach a size of 60 m (180 ft), well in the range of large terrestrial trees. They can grow extremely fast, up to 0.5 m a day, and they form huge standing stocks all along the North and South American West Coast.

Here we present our plan to sequence the full genome of *Macrcystis pyrifera* and to generate substantial transcriptome data. To facilitate this we have assembled an large international team consisting of phycologists , ecologists, physiologists, algal molecular biologists and bioinformaticans. Also Chilean companies interested in *Macrocystis* Biomass Production will be part of the consortium.

Polar Metagenomics and Metatranscriptomics

Klaus Valentin* (Klaus.Valentin@awi.de),[1] Christiane Uhlig,[1] Gerhard Dieckmann,[1] Stephan Frickenhaus,[1] Fabian Kilpert,[2] Andreas Krell,[1] Peter Kroth,[3] Andrew Toseland,[4] and Thomas Mock[4]

[1]Alfred Wegener Institute, Bremerhaven, Germany; [2]Hochschule, Bremerhaven, Germany;[3]University Konstanz, Konstanz, Germany; [4]University of East Anglia, Norwich, United Kingdom

Polar phytoplankton and namely Polar Sea Ice communities are especially endangered by predicted global change. At the same time they are understudied as compared to temperate and warm water phytoplankton, probably for logistic difficulties in studying and sampling. Therefore we started in 2007 with sampling of sea ice communities from Antarctic and recently also from Arctic sea ice for

genomics studies. Parallel and through JGI we finished sequencing of the first Polar eukaryote, the sea ice diatom *Fragilariopsis cylindrus*. Here we present recently finished and planned genomic and metatranscriptomic data from Polar phytoplankton communities and new developments in boinformatic tools.

The Microbiome of a Subsocial Neotropical Beetle (*Veturius* sp., Passalidae)

Gabriel Vargas-Asensio,[1] Myriam Hernandez,[2] Garret Suen,[3] Shaomei He,[4] Stephanie Malfatti,[4] Carlos Hernández,[2] Catalina Murillo,[2] Susannah Tringe,[4] Jon Clardy,[5] David Sherman,[6] **Adrián Pinto-Tomás*** (Adrian.pinto@ucr.ac.cr),[1,2] and Giselle Tamayo-Castillo[1,2]

[1] Universidad de Costa Rica, San Pedro de Montes de Oca, Costa Rico; [2]Instituto Nacional de Biodiversidad, INBio, Santo Domingo de Heredia, Costa Rica; [3]Department of Bacteriology, University of Wisconsin, Madison, Wisconsin; [4]DOE Joint Genome Institute, Walnut Creek, California; [5]Harvard Medical School, Boston, Massachusetts; [6]Life Sciences Institute, University of Michigan, Ann Arbor, Michigan

Passalid beetles (Coleoptera) feed almost exclusively on decaying wood. As opposed to the vast majority of coleopterans, these beetles exhibit a subsocial behavior, highlighted by larval parental care, with even three generations sharing the same decomposing log. Given their ecophysiology, this behavior may lead to the acquisition and sharing of microbial symbionts for efficient cellulose degradation. To test this hypothesis, we sampled 5 *Veturius* sp family groups from 5 different logs in Braulio Carrillo National Park (Costa Rica) and analyzed the microbial community associated with the gut contents and epithelium of both larvae and adults, as well as the woody gallery material in which they resided. The homogenized samples were divided into equal parts to evaluate the influence on DNA quality and observed microbial diversity of 3 different DNA extraction methods based on commercial kits and an additional sonication treatment. In total, 68 metagenomic DNA samples were sent to the Joint Genome Institute (JGI) for pyrotag sequencing. Data was analyzed using Pyrotagger and representative sequences were aligned using the galaxy/jgi platform. This alignment was employed to compare microbial communities using Fast Unifrac. Rarefaction curves were generated with Mothur. An AMOVA test was performed with calculated diversity indexes to detect significant differences amongst sample sources and DNA extraction methods. While all extraction methods generated good quality DNA, the modified MoBio method provided significantly higher diversity results, except for gallery material in which the modified MP method showed higher diversity ($p < 0.05$). Regarding microbial community structure, clustering based on Bray-Curtis similarity matrixes indicated that samples were primarily separated by source, and secondary by log of origin, and that the variance introduced by different DNA extraction methods within the same sample was smaller than the variance among different samples. In terms of composition, gallery material is dominated by Proteobacteria (52%), the adult gut by Firmicutes (57%) and the larval gut by Firmicutes (40%), Bacteroidetes (20%) and the archeal phyla Methanomicrobia (15%) and Thermoplasmata (10%). In addition, we found representation of at least other 34 bacterial phyla in these samples. Currently, we are analyzing 7 assembled and annotated metagenomes from these samples (5

larvae, pool of adult samples and pool of gallery samples), as well as the microbial community associated with two sets of Passalid eggs. Taken together, our results show that *Veturius* adults and larvae have a characteristic and significantly different gut microbial community, suggesting potential implications in the ecophysiology of these insects.

Understanding the Mechanism of Alternative Splicing Regulation by the P5SM RNA Element and Expanding Its Utility as a Conditional Splicing System

Pooja Vijayendra,[1] Geoffrey Liou,[1] Luke Latimer,[2] James Nunez,[1] and Ming C. Hammond* (mingch@berkeley.edu)[1,2]

[1]Department of Molecular & Cell Biology, University of California, Berkeley, Berkeley, California; [2]Department of Chemistry, University of California, Berkeley, Berkeley, California

A large percentage of alternatively spliced transcripts contain premature termination codons (PTCs) and are targets for nonsense-mediated decay. In diverse eukaryotic organisms, it has been shown that a regulated switch in splicing between a PTC-containing mRNA and a normally translated mRNA can control gene expression. We have recently shown that an engineered conditional splicing system can be used to regulate gene expression in plants via introduction of an alternatively spliced exon containing the *Hy*P5SM RNA element. Here we present the latest results on the mechanism of alternative splicing regulation by this RNA element and an update on current efforts to expand the utility of this conditional splicing system for applications in plant bioengineering.

Biopig: A Hadoop-based Analytic Toolkit for Large Scale Sequence Data

Kai Wang* (KaiWang@lbl.gov),[1] **Henrik Nordberg*** (hnordberg@lbl.gov),[1] Aijazuddin Syed,[1] Shane Canon,[2] and Zhong Wang[1]

[1]DOE Joint Genome Institute, Walnut Creek, California; [2]Lawrence Berkeley National Laboratory, Berkeley, California

BioPig is a sequence analysis toolkit developed at DOE Joint Genome Institute. It is built upon the Pig query language and Apache's Hadoop map-reduce framework, and therefore enables the user to develop and test parallel bioinformatics applications quickly and easily.

We introduce the BioPig framework and its design principles, and show its ability to scale to next-generation sequence data and computation. We also illustrate how BioPig scripts concisely implement common sequence analysis tasks, ranging from simple ones like k-mer statistics to more complex ones like metagenomic gene discovery. Numerical results are reported on Magellan system at NERSC and Amazon Elastic Compute Cloud. Where possible, we provide performance comparisons with alternative methods.

The DOE Systems Biology Knowledgebase: Plant Science Domain

Doreen Ware* (ware@cshl.edu),[1,2] Sergei Maslov,[4] Shinjae Yoo,[4] Dantong Yu,[4] Michael Schatz,[1] James Gurtowski,[1] Matt Titmus,[1] Jer-ming Chia,[1] Sunita Kumari,[1] Andrew Olson,[1] Shiran Pasternak,[1] Jim Thomason,[1] Ken Youens-Clark,[1] Mark Gerstein,[5] Gang Fang,[5] Darryl Reeves,[5] Pam Ronald,[6] TaeYun Oh,[6] Chris Henry,[7] Sam Seaver,[7] David Weston,[3] Priya Ranjan,[3] Musstafa Syed,[3] Miriam Land,[3] and Adam Arkin[8]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York; [2]United States Department of Agriculture (USDA)—Agriculture Research Service; [3]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [4]Brookhaven National Laboratory, Upton, New York; [5]Yale University, New Haven, Connecticut; [6]University of California, Davis, California; [7]Argonne National Laboratory, Argonne, Illinois; [8]Lawrence Berkeley National Laboratory, Berkeley, California

The Systems Biology Knowledgebase (KBase) has two central goals. The scientific goal is to produce predictive models, reference datasets, analytical tools; and to demonstrate their utility in DOE biological research related to bioenergy, carbon cycle, and the study of subsurface microbial communities. The operational goal is to create the integrated software and hardware infrastructure needed to support the creation, maintenance, and use of predictive models and methods in the study of microbes, microbial communities and plants. The plant component of the KBase will allow users to model genotype-to-phenotype relationships using metabolic and functional networks. It will also support the reconstruction of new metabolic and functional networks. To accomplish this, we will provide interactive, data-driven analysis and exploration across multiple experiments and diverse data-types. We will provide users access to comprehensive collections of 'omics datasets together with relevant analytical tools and resources.

The major goal of KBase plants is to model genotype-to-phenotype relationships through analysis and integration of genomic, transcriptomic, metabolite and phenotype measurements, and the reconstruction of metabolic and functional networks based on expression profiles, protein-DNA, and protein-protein interactions. To accomplish this goal, KBase will provide interactive, data-driven analysis, and exploration across multiple experiments and diverse data-types. Users will be provided with a platform to analyze their own experimental data, integrate publicly available data from other 'omics' platforms, and have these results incorporated into a data exploration framework.

The KBase plants effort will consist of two major components, 1) genotyping workflows and 2) data exploration and prediction tools.

Genotyping workflows: Exponential growth in digital demands has motivated extensive research into improved algorithms and parallel systems, especially for genotyping samples, monitoring expression levels, and a host of other important biological applications.

Genotyping workflows will leverage our recent development of Jnomics, as our new Hadoop-based open-source package for rapid development and deployment of cloud-scale sequence analysis tools. Jnomics provides many pre-built tools out-of-the-box that accelerate common tasks as distributed tasks spread across a cluster. New tools can be easily created using an open-source Java API, especially for large-scale genotyping and expression analysis. Because it builds on Hadoop, Jnomics tools inherit Hadoop's efficiency and scalability for very large datasets.

Furthermore, Jnomics is "file format agnostic," allowing it to seamlessly read and write most common sequence file formats making it easy for Jnomics to interface with other components.

…

Hybrid Assembly of Metagenomic and Single-Cell Genome Sequencing Data

Christopher Wei* (cjwei@ucsd.edu), Hsin-I Chiang, and Kun Zhang

Department of Bioengineering, University of California, San Diego, La Jolla, California

Metagenomics and single cell genomics are two current culture-independent methods for genomic analysis of difficult-to-culture microorganisms. Metagenomics provides sequence data of the overall microbial population, while single cell genomics potentially allows for deeper sequencing of low-density bacteria. We successfully developed a method to combine both approaches to improve assembly of single bacterial genomes. We used next generation sequencing to attain a metagenomic profile of the mouse gut microbiome. From the same microbiome we utilized fluorescence-activated cell sorting (FACS) to isolate single bacterial cells and perform polymerase cloning and sequencing on the single cell genomes. The sequence reads obtained from a single cell genome were first de novo assembled into contigs, and then these contigs were used to recruit additional reads from the metagenomic sequence data based on shared homology. Finally the contigs and recruited reads were combined and re-assembled using a modified assembling approach. We demonstrated that the re-assembled contigs result in 4-7x longer N50 length and ~3x larger total contig length than the original single cell assembly. This resulted in genome completeness as high as 81% of single cell bacterial genomes. Research into identification and sequencing of low density, unknown bacteria will greatly benefit from our combined metagenomics and single cell method by the reduced sequencing depth from a single cell genome, which would allow for higher quality genomes of unknown bacteria.

Insight from Whole Genome and Gene Specific Sequence in Classical Mutant Strains of *Neurospora crassa*

Aric E. Wiest,[1] Alex J. McCarthy,[1] Rob R. Schnittker,[1] Scott E. Baker,[2] and **Kevin McCluskey*** (mccluskeyk@umkc.edu)[1]

[1]Fungal Genetics Stock Center, University of Missouri-Kansas City, School of Biological Sciences, Kansas City, Missouri; [2]U.S. DOE Pacific Northwest National Laboratory, Richland, Washington

Building upon over 70 years of genetic and biochemical research, the plant biomass deconstructing fungus *Neurospora crassa* is providing insight into theoretical and practical issues associated with growth and manipulation of fungi for implementation of next generation biomass conversion. While classical genetics has identified over 1,000 genetic loci in *N. crassa*, the ORF associated with over 400 of these loci has not been characterized at the level of the DNA sequence. Whole genome and targeted gene sequencing is being used at the FGSC

to characterize otherwise anonymous classical mutations, providing added value to the materials in the FGSC collection. While whole genome sequencing of 19 different laboratory strains has proven to be useful for the identification of individual mutations, the insight provided by intra-species comparisons between the reference genome and newly sequenced genomes provides unique information in a variety of areas. For example, the number and characteristics of Single Nucleotide Polymorphisms, Insertions, and Deletions reveals presumably neutral variation. The ability of strains to tolerate the numerous nuclear and mitochondrial frameshift mutations caused by indels is surprising, as is the number of nonsense mutations in some strains. Additional analysis includes the distribution of a newly discovered 4-base repeat that is manifest as an insertion or deletion with regard to the reference genome. Similarly, characterization of known and genetically mapped mutations reinforces our understanding of the limits of genetic mapping. Ongoing work continues the themes of gene and mutation characterization.

McCluskey, K., Wiest, A., Martin, J., Lipzen, A., Schackwitz, W., Grigoriev, I. and S.E. Baker. 2011. Rediscovery by whole genome sequencing: classical mutations and genome polymorphisms in *Neurospora crassa*. G3: Genes|Genomes|Genetics 1:303-316 doi: 10.1534/g3.111.000307

Integrating Community Data, Activity-Specific Populations and Individual Genomes in Mapping the Biochemical Networks of Lignocellulose Degradation by Fungi and Bacteria in Forest Soils from within the Long Term Soil Productivity Experiment

Roland Wilhelm* (rwilhelm@mail.ubc.ca)**,** Kendra Mitchell, Erick Cardenas, Martin Hartmann, David Van Insberghe, Hilary Leung, Steven Hallam, and William Mohn

Department of Microbiology and Immunology, University of British Columbia, Vancouver, Canada

Forest soils are rich in high molecular weight lignocellulose compounds which provide carbon and energy for one of the world's most complex-functioning, biologically-active habitats.

The decomposition of lignocellulose, a crystalline multi-component polymer consisting of saccharide and phenolic subunits, involves a massive diversity of enzymes (over 130 families of glycosyl hydrolases alone) and a succession of microbial populations with varying metabolic capacities. With "–omic" approaches, the capacity now exists to deconstruct the complex biochemical and ecological forces driving lignocellulose decomposition. The recent discovery of novel soil bacterial taxa capable of lignin degradation (Bugg et al. 2011) demonstrates our lack of comprehensive data on the decomposer community. We have begun a multi-layered approach which integrates comprehensive pyrosequencing datasets of community genetics (16S and 18S DNA and cDNA libraries), with metagenomics, metatranscriptomics, stable isotope probing (SIP) (13C lignin, cellulose & hemicellulose) and single-cell genomics. These analyses are being performed on a large number of replicated treatments as part of a robust, long term forest regeneration experiment known as the Long Term Soil Productivity experiment (LTSP), involving hundreds of researchers and research sites across North America. Community phylogenetic libraries are used to identify

key taxa, based on the co-occurrence with taxa known to degrade lignocellulose, with specific attention to identify possible fungal and bacterial mutualism. SIP-DNA methods pinpoint novel taxa involved in the decomposition of lignocellulose components, which can then be queried against our community database to gather distribution and abundance estimates and to identify broader ecological networks. Currently, we've compiled over 130, 000 16S and 18S sequences from across six forest stands in British Columbia, with similar community data from three other regions (12 sites) being processed. We are assembling and annotating eight draft genomes from isolates corresponding to the most abundant taxa in our community library and preliminary metagenomic data from two sites will be available for presentation. Our preliminary data will be illustrative of the efficacy of our approach and the presentation of our analysis pipeline will offer insights into integrating –omic approaches in a broad environmental study. By focusing our study on LTSP sites, we will not only provide a comprehensive examination of lignocellulose-degrading soil organisms, but we have the opportunity to correlate microbial ecology to above ground productivity (soil fertility) and biodiversity. Finally, by cataloguing the wide range of naturally occurring lignocellulose-degrading enzymes in individual genomes and metagenomes, we increase the likelihood of discovering an industrially relevant enzyme capable of pre-processing woody wastes into bio-fuels or other synthetic materials.

Bugg et al. "The emerging role for bacteria in lignin degradation…" *Curr. Opin. Biotech.* **22** (2010)

Bacterial Endophytes of Conifer Buds

Emily C. Wilson* (ewilson3@ucmerced.edu**)** and A. Carolin Frank

University of California, Merced, Merced, California

Understanding the factors that contribute to successful and productive plants is key to improving renewable approaches to agriculture and forestry. Symbionts such as endophytes—bacteria and fungi inside healthy plant tissue—can play a role in plant stress protection and growth promotion and warrant further research. Our goal is to isolate conifer endophytes found across different conifer species, which may indicate a long-term mutualistic association with the host. Such isolates are more likely to be beneficial to the host, and are therefore interesting targets for downstream studies. We use layered agar culturing method with several minimal media agar types to isolate colonies from pulverized conifer tissue. Individual colonies are identified using DNA extractions, PCRs and Sanger sequencing. In addition to sequencing identification we use a laser identification method called BARDOT (bacterial rapid identification using optical scattering technology) which generates the forward light scatter pattern of each colony. These scatter patterns are unique to bacterial species down to the strain and even serovar level and the construction of a scatter pattern library will aid in rapid identification of isolates. From conifer bud tissue sampled in two locations and needle tissue in a third location we cultured several common endophytes isolates across conifer species and sampling location. We isolated *Bacillus* sp. 2_A_57_CT2 from the bud tissue of *Pinus nigra* (sampled in Merced, CA) and *Pinus ponderosa* (sampled in Yosemite Valley, CA) as well as the needle tissue of *Pinus contorta* (sampled in

Tuolumne Meadows, CA). *Micrococcus luteus* and *Bacillus firmus* were isolated from the bud tissue of *P. nigra* (Merced) and *P. ponderosa* (Yosemite Valley). Several other bacterial endophytes were isolated including *Bacillus* and *Paenibacillus* species which were not found in all conifer tissues. BARDOT scatter patterns of all isolates were collected and are being used to build a scatter pattern library for future use. Our results reveal that the same bacterial species can be isolated from conifer bud tissue sampled in different locations from different conifer species, suggesting that associations between bud endophytes and conifer hosts are conserved across different conifer species. Future work includes genome sequencing and the analysis of potential beneficial properties of each isolate.

Independent Contrast Based Phylogenetic Profiling

Dongying Wu* (DYWu@lbl.gov),[1,2] Guillaume Jospin,[2] and Jonathan A. Eisen[1,2]

[1]DOE Joint Genome Institute, Walnut Creek, California; [2]University of California, Davis, Davis, California

Non-homology functional prediction methods allow for general predictions of gene function even when functions are not known for any homologs of a gene of interest. One such non-homology method - phylogenetic profiling - involves the analysis of the joint presence or absence of two gene families in different taxa. Phylogenetic profile analysis has been used previously to infer a meaningful biological connection between multiple gene families such as presence in biological pathways and interactions between proteins in complexes. Independent contrast is a statistical method that incorporating phylogenetic information to transform correlations of biological data into values that are statistically independent and identically distributed. We report here on the use of independent contrasts for phylogenetic profiling of gene families found across the diversity of bacteria and archaea. Specifically, we have constructed 568,290 gene families for 965 complete bacteria and archaea genomes included in the IMG database. Using a phylogenetic tree of the 965 genomes, we then calculated the independent contrasts of 74,789,356 pairs of families (we focused on those that are present in at least five operational of the 965 genomes). We have grouped 13,476 gene families into 2531 clusters according to the independent contrasts between the families. Using a similar approach, we have grouped 826 Clusters of Orthologous Groups of proteins (COGs) into 226 clusters, and 186 Pfams into 75 clusters. Efforts are underway to analyze all the gene families included in the independent contrast based phylogenetic profiling clusters to shed light on the biological functions of the unknown gene families.

Metagenomic Landscape of Gut Microbial Community in the Malaria Mosquito *Anopheles gambiae*

Jiannong Xu* (jxu@nmsu.edu),[1] Phanidhar Kukutla,[1] Alexander Tchourbanov,[1] Homgmei Jiang,[2] Ying Wang,[1] Matthew Steritz,[1] Jinjin Jiang,[1] and Michael Best[1]

[1]New Mexico State University Biology Department, Las Cruces, New Mexico; [2]Northwestern University Department of Statistics, Chicago, Illinois

The mosquito gut accommodates a dynamic microbiota that is essential for various mosquito life traits. To understand the structure and functionality of the gut ecosystem, we characterized the microbiome by 16S rRNA, metagenomic and metatranscriptome sequencing. Bacterial 16S rRNA V1-3 pyroreads revealed that the gut community structures were dynamic across life stages from immature stages (larvae and pupae in aquatic habitats) to adults. Intriguingly, a blood meal drastically simplified bacterial composition in adult gut. The taxa from *Enterobacteriaceae* and *Psuedomonadaceae* were selectively enriched after a blood meal. Besides, *Elizabethkingia* sp. is a predominant inhabitant in the adult guts. The adult gut microbiome was further characterized by Illumina metagenomic and metatranscriptomic sequencing. The Illumina reads were assembled into contigs for taxonomic classification and function assignment. The community membership revealed by metagenomic assembly was consistent with that by 16S rRNA profiling. The metagenomic assembly was annotated against SEED subsystem and KEEG Orthology (KO). The link between community structure and function was investigated by community comparison between sugar fed and blood fed guts. Interestingly, despite the variability in microbial composition, a core suite of function categories were present in both communities, suggesting that the function capacity in the gut ecosystem is determined by existing genes rather than species. Such relative stability of the genetic capacity is important for the functionality of an ecosystem. Likely, a fluctuation of variation in taxonomic composition may not significantly affect functionality of the community. The functioning network in blood fed gut was assessed by transcriptome analysis. Blood meal digestion releases a substantial amount of free heme from hemoglobins. Free heme is a rich source of reactive oxygen species (ROS), which imposes oxidative stress in the gut environment. Transcriptomic analysis showed that anti-oxidant genes from both mosquito and microbes were upregulated in the blood fed gut. This indicates that symbiotic microbes contribute to the redox homeostasis in the blood fed gut, which is essential for mosquito fecundity.

Barcode Labeling of Short Reads for Detection of Large Scale Genomic Variations

Tao Zhang* (tzhang3@lbl.gov), Kevin Eng, Jeff Froula, Kanwar Singh, Matt Blow, Zhong Wang, Len Pennacchio, and Feng Chen

DOE Joint Genome Institute, Walnut Creek, California

Genomic variations, such as insertion, deletion, inversion and translocation, are common in human population and frequently found in tumor tissues. Identification of these variations is important for understanding difference between individuals and diagnosis of diseases. This goal can be achieved by mapping sequences

obtained from different individuals or tumor samples into reference genome. Although short-read sequences can be used to detect minor changes in nucleotide sequence by mapping, due to their small footprints, short reads are inadequate for identification of large genomic variations. We have developed a method for labeling of small DNA fragments derived from single large DNA molecules (up to 40kb) using nucleotide sequence-based barcodes. By barcode labeling, the linkage association was preserved, even if there was no sequence overlaps between these short reads. The linkage information can be used to extend the footprints of short-read sequences for detection of events of large genetic variations. We have tested labeling of ~1,000,000 large DNA molecules using 960 unique barcodes in micro-droplets. We found that short-read sequences containing same barcodes can be mapped into clusters in reference genome, suggesting that these short reads were derived from identical DNA templates. Mapping of these short reads into a reference genome containing simulated large scale genomic variations revealed that boundaries of these clusters were located right next to the sites of variations. Our results suggested that barcode labeling of small genomic DNA fragments can be used to increase footprints of short sequence reads for detection of large genetic variations in complex genome.

Identification of Ionic Liquids-Tolerant Cellulases for Biomass Processing

Tao Zhang,[1] Supratim Datta,[2] Jerry Eichler,[3] Natalia Ivanova,[1] Seth D. Axen,[1] Cheryl A. Kerfeld,[1,4] Feng Chen,[1] Nikos Kyrpides,[1] Philip Hugenholtz,[1,2,5] Jan-Fang Cheng,[1] Kenneth L. Sale,[2] Blake Simmons* (BASimmons@lbl.gov),[2] and Eddy Rubin[1]

[1]DOE Joint Genome Institute, Walnut Creek, California; [2]The Joint BioEnergy Institute, Emeryville, California; [3]Department of Life Sciences, Ben Gurion University of the Negev, Beersheva, Israel; [4]Department of Plant and Microbial Biology, University of California, Berkeley, California; [5]Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences & Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia

Ionic liquids (ILs) are effective solvents for biomass pretreatment. It is known that ILs can have inhibitory effect on fungal cellulases, making the digestion of cellulose inefficient in the presence of ILs. The identification of IL-tolerant enzymes that could be produced as a cellulase cocktail would reduce the costs and water use requirements of the IL pretreatment process. Due to their adaptation to high salinity environments, halophilic enzymes are hypothesized to be good candidates for screening and identifying IL-resistant cellulases. Using a genome-based approach, we have identified and characterized a halophilic cellulase (*Hu*-CBH1) from the halophilic archaeon, *Halorhabdus utahensis*. *Hu*-CBH1 is present in a gene cluster containing multiple putative cellulolytic enzymes. Sequence and theoretical structure analysis indicate that *Hu*-CBH1 is highly enriched with negatively charged acidic amino acids on the surface, which may form a solvation shell that may stabilize the enzyme, through interaction with salt ions and/or water molecules. *Hu*-CBH1 is a heat tolerant haloalkaliphilic cellulase and is active in salt concentrations up to 5 M NaCl. In high salt buffer, *Hu*-CBH1 can tolerate alkali (pH 11.5) conditions and, more importantly, is tolerant to high levels (20% w/w) of ILs, including 1-allyl-3-methylimidazolium chloride ([Amim]Cl). Interestingly, the tolerances to heat, alkali and ILs are found to be salt-dependent,

suggesting that the enzyme is stabilized by the presence of salt. Our results indicate that halophilic enzymes are good candidates for the screening of IL-tolerant cellulolytic enzymes.

From Community Structure to Functions: Metagenomics-Enabled Predictive Understanding of Temperature Sensitivity of Soil Carbon Decomposition to Climate Warming

Jizhong Zhou*(jzhou@ou.edu),[1] Liyou Wu,[1] Kai Xue,[1] Lei Cheng,[1] Mengting Yuan,[1] Jin Zhang,[1] Ye Deng,[1] Joy D. Van Nostrand,[1] Zhili He,[1] Ryan Penton,[2] Jim Cole,[2] James M. Tiedje,[2] Rosvel Bracho-Garrillo,[3] Edward A.G. Schuur,[3] Chengwei Luo,[4] Konstantinos Konstantinidis,[4] Xia Xu,[1] Dejun Li,[1] and Yiqi Luo[1]

[1]Institute for Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma; [2]Center for Microbial Ecology, Michigan State University, East Lansing, Michigan; [3]Department of Biology, University of Florida, Gainesville, Florida; [4]Center for Bioinformatics and Computational Genomics and School of Biology, Georgia Institute of Technology, Atlanta, Georgia

Determining the response, adaptation and feedback mechanism of biological communities to climate change is critical to project future states of the earth and climate systems, but poorly understood in microbial communities. To provide system-level, predictive mechanistic understanding of the temperature sensitivity of soil carbon decomposition to climate warming by using cutting-edge integrated metagenomic technologies, we have carried out our studies at two contrasting long-term experimental facilities, the tundra ecosystems in Alaska and the temperate grassland ecosystems in Oklahoma.

Approximately 1670 Pg (billion tons) of soil carbon are stored in the northern circumpolar permafrost zone. Permafrost thaw, and the microbial decomposition of previously frozen organic carbon, is considered one of the most likely positive feedbacks from terrestrial ecosystems to the atmosphere in a warmer world. Our experimental results showed that the areas that thawed over the past 15 years had 75% more annual losses of old C compared to minimally thawed areas, but had overall net ecosystem C uptake as increased plant growth offset these losses. In contrast, sites that thawed decades earlier lost an additional 25% more old C annually, which contributed to overall net ecosystem C release despite increased plant growth. These findings were mirrored by the warming experiment where increased plant uptake appears to compensate for microbial release of carbon, at least in the three years of warming that we have observed. Metagenomic sequencing and GeoChip analyse showed that the microbial community compositions and structure were dramatically altered by natural thawing or experimental warming. Together, these data document significant losses of soil C with permafrost thaw that, over decadal time scales, overwhelms increased plant C uptake at rates that could make permafrost a large biospheric C source in a warmer world, similar in magnitude to current C fluxes from land use change.

We have also used integrated metagenomic technologies to analyze the responses of microbial communities in a long-term (10 years) experimental warming grassland ecosystem in Oklahoma. Our results showed that microorganisms play crucial roles in regulating soil carbon (C) dynamics through three primary feedback

mechanisms: (i) shifting microbial community composition, which most likely led to the reduced temperature sensitivity of heterotrophic soil respiration, (ii) differentially stimulating genes for degrading labile but not recalcitrant C so as to maintain long term soil C stability and storage, and (iii) enhancing nutrient cycling processes to promote plant nutrient use efficiency and hence plant growth.

…

Poster Presentations

Attendees

Current as of February 27, 2012

Eric Ackerman
Sandia National Laboratory
eackerm@sandia.gov

Ed Allen
DOE Joint Genome Institute
eaallen@lbl.gov

Iain Anderson
DOE Joint Genome Institute
IJAnderson@lbl.gov

Adam Arkin
Lawrence Berkeley National Lab
aparkin@lbl.gov

Scott Baker
Pacific Northwest National Laboratory
scott.baker@pnl.gov

Petr Baldrian
Institute of Microbiology ASCR
baldrian@biomed.cas.cz

Susan Baldwin
Univeristy of British Columbia
sbaldwin@mail.ubc.ca

Massie Ballon
DOE Joint Genome Institute
mlballon@lbl.gov

Nicholas Ballor
Joint BioEnergy Institute
nrballor@lbl.gov

Jody Banks
Purdue University
banksj@purdue.edu

Richard Baran
Lawrence Berkeley National Lab
RBaran@lbl.gov

Kerrie Barry
DOE Joint Genome Institute
kwbarry@lbl.gov

Ellen Beauchamp
Williams College
emb3@williams.edu

Steven Benner
Foundation for Applied Molecular Evolution
sbenner@ffame.org

Randy Berka
Novozymes, Inc.
ramb@novozymes.com

Avijit Biswas
University of Southern Mississippi
Biswas.Avijit@usm.edu

Jaime Blair
Franklin & Marshall College
jaime.blair@fandm.edu

Matthew Blow
DOE Joint Genome Institute
mjblow@lbl.gov

Harvey Bolton
Pacific Northwest National Laboratory
harvey.bolton@pnnl.gov

Gregory Bonito
Duke University
gmb2@duke.edu

Magnolia Bostick
Clontech Laboratories, Inc.
magnolia_bostick@clontech.com

Siobhan Brady
University of California, Davis
sbrady@ucdavis.edu

Jennifer Bragg
US Department of Agriculture
jennifer.bragg@ars.usda.gov

William Brazelton
East Carolina University
wbrazelton@gmail.com

Natalie Breakfield
Univ. of North Carolina, Chapel Hill
nbreakfield@gmail.com

Thomas Brettin
Oak Ridge National Laboratory
brettints@ornl.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
dcbruce@lbl.gov

Gregory Butler
Concordia University
gregb@cs.concordia.ca

Heike Bücking
South Dakota State University
heike.bucking@sdstate.edu

Craig Cary
University of Waikato
caryc@waikato.ac.nz

Leong Keat Chan
DOE Joint Genome Institute
leongchan@lbl.gov

Patricia Chan
University of California, Santa Cruz
pchan@soe.ucsc.edu

Jay Chen
Oak Ridge National Laboratory
chenj@ornl.gov

Amy Chen
DOE Joint Genome Institute, LBNL
IMAChen@lbl.gov

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

Dylan Chivian
Lawrence Berkeley National Lab
DCChivian@lbl.gov

Mansi Chovatia
DOE Joint Genome Institute
mrchovatia@lbl.gov

Virginia Chow
University of Florida
vmchow@ufl.edu

Julianna Chow
DOE Joint Genome Institute
jchow@lbl.gov

Kris Christen
Oak Ridge National Laboratory
christenks@ornl.gov

Melissa Christopherson
University of Wisconsin, Madison
melissachristopherson@gmail.com

George Chuck
Plant Gene Expression Center, UCB
georgechuck@berkeley.edu

Scott Clingenpeel
DOE Joint Genome Institute
srclingenpeel@lbl.gov

Devin Coleman-Derr
University of California, Berkeley
colemanderr@berkeley.edu

Robert Cottingham
Oak Ridge National Laboratory
cottinghamrw@ornl.gov

David Cowley
Pacific Northwest National Laboratory
david.cowley@pnnl.gov

Attendees

Benjamin Crary
University of Wisconsin
benjamin.crary@gmail.com

Pedro Crous
CBS Fungal Biodiversity
Centrep.crous@cbs.knaw.nl

Aaron Darling
Genome Center, UC-Davis
aarondarling@ucdavis.edu

Chris Daum
DOE Joint Genome Institute
daum1@llnl.gov

Brian Davison
Oak Ridge National Laboratory
davisonbh@ornl.gov

Paramvir Dehal
Lawrence Berkeley National Lab
psdehal@lbl.gov

Ed DeLong
MIT
delong@mit.edu

Shweta Deshpande
DOE Joint Genome Institute
SDeshpande@lbl.gov

Stephen DiFazio
West Virginia University
spdifazio@mail.wvu.edu

Daniel Drell
US Department of Energy
Daniel.drell@science.doe.gov

Irina Druzhinina
Vienna University of Technology
druzhini@mail.zserv.tuwien.ac.at

Rob Egan
DOE Joint Genome Institute
rsegan@lbl.gov

Nadine El Said
American University in Cairo
nadine86@aucegypt.edu

Sawsan ELgogary
American University in Cairo
s.elgogary@aucegypt.edu

Bryan Ervin
California State University, Chico
bervin@mail.csuchico.edu

Luke Evans
Northern Arizona University
luke.evans@nau.edu

Lasse Feldhahn
Helmholtz-Center for Env. Research
lasse.feldhahn@ufz.de

Marsha Fenner
DOE Joint Genome Institute
MWFenner@lbl.gov

Klaus Fiebig
Ontario Genomics Institute
klausfiebig@gmail.com

Erin Field
Bigelow Laboratory for Ocean Sciences
efield@bigelow.org

Cheryl Foust
Oak Ridge National Laboratory
foustcb@ornl.gov

Jed Fuhrman
University of Southern California
fuhrman@usc.edu

Melany Funes
Williams College
msf2@williams.edu

Sean Gallaher
University of California, Los Alamos
gallaher@chem.ucla.edu

Justin Gallivan
Emory University
justin.gallivan@emory.edu

Jinnie Garrett
Hamilton College
jgarrett@hamilton.edu

George Garrity
Michigan State University
garrity@msu.edu

Gaurav Ghag
University of Southern Mississippi
gaurav.ghag@eagles.usm.edu

Mohamed Ghazy
American University in Cairo
mghazy@aucegypt.edu

Maria Ghirardi
NREL
maria.ghirardi@nrel.gov

Trevor Ghylin
University of Wisconsin
Madisontrevor.ghylin@gmail.com

David Gilbert
DOE Joint Genome Institute
degilbert@lbl.gov

Carol Giometti
Argonne National Laboratory
csgiometti@anl.gov

Stephen Giovannoni
Oregon State University
steve.giovannoni@oregonstate.edu

N. Louise Glass
University of California, Berkeley
lglass@berkeley.edu

Fred Gmitter
University of Florida
Citrus Research and Education Center
fgg@crec.ifas.ufl.edu

Bob Goldberg
University of California, Los Angeles
bobg@ucla.edu

Stephen Goodwin
US Dept. of Agriculture, Purdue Univ.
sgoodwin@purdue.edu

Lynne Goodwin
Los Alamos National Laboratory
lynneg@lanl.gov

Sean Gordon
US Department of Agriculture
seangordon07@gmail.com

Igor Grigoriev
DOE Joint Genome Institute
ivgrigoriev@lbl.gov

Stephen Gross
DOE Joint Genome Institute
smgross@lbl.gov

Masood Hadi
Sandia National Laboratory, JBEI
mzhadi@sandia.gov

Ming Hammond
Lawrence Berkeley National Lab
mingch@berkeley.edu

James Han
DOE Joint Genome Institute
jkhan@lbl.gov

Miranda Harmon-Smith
DOE Joint Genome Institute
MLHarmon-Smith@lbl.gov

Erik Hawley
Washington State University Tri-Cities
ehawley.amge@gmail.com

Shaomei He
DOE Joint Genome Institute
she@lbl.gov

Sylvie Herrmann
Helmholtz-Center for Env. Research
sylvie.herrmann@ufz.de

Russell Herwig
University of Washington
herwig@uw.edu

Jaqueline Hess
Harvard University
jhess@fas.harvard.edu

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

David Hibbett
Clark University
dhibbett@clarku.edu

Nan Ho
Las Positas College
nho@laspositascollege.edu

Cindi Hoover
DOE Joint Genome Institute
cahoover@lbl.gov

Mandy Hsia
US Department of Agriculture
mandy.hsia@ars.usda.gov

Susan Hua
DOE Joint Genome Institute
shua@lbl.gov

Miriam Hutchinson
University of New Mexico
yesterdaymail@gmail.com

William Inskeep
Montana State University
binskeep@montana.edu

Erin Jarvis
UC Berkeley
erinjarvis@berkeley.edu

Thomas Jeffries
University of Wisconsin Madison
twjeffri@wisc.edu

Jerry Jenkins
DOE Joint Genome Inst, HudsonAlpha
jjenkins@hudsonalpha.org

Charles Jensen
Biomed Central
charles.jensen@biomedcentral.com

Richard Jorgensen
Langebio
rajorgensen@langebio.cinvestav.mx

Shawn Kaeppler
University of Wisconsin
smkaeppl@wisc.edu

Don Kang
DOE Joint Genome Institute
ddkang@lbl.gov

Ulas Karaoz
Lawrence Berkeley National Lab
UKaraoz@lbl.gov

Matthew Kaser
Bell & Associates
mkaser@bell-iplaw.com

Lisa Kegg
DOE Joint Genome Institute
lrkegg@lbl.gov

Gert Kema
Plant Research International B.V.
gert.kema@wur.nl

Richard Kerrigan
Sylvan Research
rwk@sylvaninc.com

Cris Kinross
Epicentre, an Illumina company
cristine.kinross@epicentre.com

Valentin Klaus
AWI
valentin@awi.de

Annegret Kohler
INRA
annegret.kohler@nancy.inra.fr

Tom Kristensen
University of Oslo
tom.kristensen@imbv.uio.no

Sunita Kumari
Cold Spring Harbor Labs
kumari@cshl.edu

Victor Kunin
Taxon Biosciences
kunivi@gmail.com

Dwight Kuo
Lawrence Berkeley National Lab
dkuo@lbl.gov

Cheryl Kuske
Los Alamos National Laboratory
kuske@lanl.gov

Nikos Kyrpides
DOE Joint Genome Institute
NCKyrpides@lbl.gov

David Lai
DOE Joint Genome Institute
dlai@lbl.gov

Robert Landick
Univ. of Wisconsin, Madison, GLBRC
landick@bact.wisc.edu

Christina Lanzatella-Craig
US Department of Agriculture
christina.craig@ars.usda.gov

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

Sarah Lebeis
University of North Carolina
lebeis@live.unc.edu

Edwin Lee
Genencor
edwin.lee@danisco.com

Hana Lee
University of California, Berkeley
hanalee@berkeley.edu

Dacia Leon
EBI/UC Berkeley
dleon@lbl.gov

Daniel Linder
US Forest Service
dlindner@wisc.edu

Tomas Linder
Swedish Univ. of Agricultural Sciences
tomas.linder@slu.se

Erika Lindquist
DOE Joint Genome Institute
ealindquist@lbl.gov

Anna Lipzen
DOE Joint Genome Institute
alipzen@lbl.gov

Siqing Liu
US Department of Agriculture
siqing.liu@ars.usda.gov

Connie Lovejoy
IBIS / Laval University
Connie.lovejoy@bio.ulaval.ca

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

Susan Madrid
Dupont Industrial Biosciences
susan.madrid@danisco.com

Stephanie Malfatti
DOE Joint Genome Institute
SAMalfatti@lbl.gov

Francis Martin
INRA
fmartin@nancy.inra.fr

Jeffrey Martin
DOE Joint Genome Institute
jamartin@lbl.gov

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

Rebekah Mathews
DOE Joint Genome Institute
rpmathews@lbl.gov

Konstantinos Mavrommatis
DOE Joint Genome Institute
KMavrommatis@lbl.gov

Kevin McCluskey
Fungal Genetics Stock Center
mccluskeyk@umkc.edu

Katherine McMahon
University of Wisconsin
tmcmahon@engr.wisc.edu

Xiandong Meng
DOE Joint Genome Institute
xiandongmeng@lbl.gov

Sabeeha Merchant
University of California, Los Angeles
merchant@chem.ucla.edu

Folker Meyer
Argonne National Laboratory
folker@anl.gov

Jenifer Milam
University of Southern Mississippi
jenifer.milam@eagles.usm.edu

Attendees

Florence Mingardon
JBEI/TOTAL
fmingardon@lbl.gov

Beth Mole
Univerisity of California, Santa Cruz
bethmariem@gmail.com

Mary Ann Moran
University of Georgia
mmoran@uga.edu

Emmanuelle Morin
INRA
emmanuelle.morin@nancy.inra.fr

Paul Morris
Bowling Green State University
pmorris@bgsu.edu

Annika Mosier
UC Berkeley
annika.mosier@gmail.com

Wellington Muchero
Oak Ridge National Laboratory
mucherow@ornl.gov

Biswarup Mukhopadhyay
Virginia Tech
biswarup@vt.edu

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

Donald Natvig
University of New Mexico
dnatvig@gmail.com

David Nelson
University of Minnesota
nels5133@umn.edu

Jessica Nguyen
Eureka Genomics
jessica@eurekagenomics.com

Angela Norbeck
Pacific Northwest National Lab / EMSL
angela.norbeck@pnnl.gov

Trent Northen
Lawrence Berkeley National Lab
trnorthen@lbl.gov

Robin Ohm
DOE Joint Genome Institute
raohm@lbl.gov

Millie Olsen
Montana State University
millie.thornton@msu.montana.edu

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

Michiel Op De Beeck
Hasselt University
michiel.opdebeeck@uhasselt.be

Aidee Orozco
Brown Forman Corporation
aideeorozco@b-f.com

Catherine Owensby
Oregon State University
owensbyc@science.oregonstate.edu

Krishnaveni Palaniappan
DOE Joint Genome Institute, LBNL
kpalaniappan@lbl.gov

Tae-Jin Park
University of Hong Kong
tjpark1@hku.hk

Karen Parker
Moss Landing Marine Labs
karen.parker2000@gmail.com

Weronika Patena
Carnegie Institution for Science
patena@gmail.com

Yuri Pena
Brown-Forman
yuripena@b-f.com

Ze Peng
DOE Joint Genome Institute
zpeng@lbl.gov

Lin Peters
DOE Joint Genome Institute
lgpeters@lbl.gov

Hailan Piao
Washington State University Tri-Cities
hailan.piao@tricity.wsu.edu

Samuel Pitluck
DOE Joint Genome Institute
s_pitluck@lbl.gov

Darren Platt
Amyris Biotechnologies
platt@amyris.com

Ram Prasad Pokhrel
Global Environment Research Pvt. Ltd.
ger.nepal@gmail.com

Amy Powell
Sandia National Laboratories
ajpowel@sandia.gov

Abhishek Pratap
DOE Joint Genome Institute
apratap@lbl.gov

James Preston
University of Florida
jpreston@ufl.edu

Endang Purwantini
Virginia Tech
epurwant@vt.edu

Yinbo Qu
Shandong University
quyinbo@sdu.edu.cn

Pablo Rabinowicz
US Department of Energy
pablo.rabinowicz@science.doe.gov

Preethi Ramaiya
Novozymes, Inc.
pira@novozymes.com

Ian Reid
Concordia University
ian_reid@gene.concordia.ca

Aaron Richardson
Codexis
aaron.richardson@codexis.com

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

Loren Rieseberg
University of British Columbia
lriesebe@mail.ubc.ca

Monica Riley
Marine Biological Lab
mriley301@comcast.net

Christian Rinke
DOE Joint Genome Institute
crinke@lbl.gov

Barbara Robbertse
NCBI
robberts@ncbi.nlm.nih.gov

Evan Roberts
University of Southern Mississippi
evanw.roberts@gmail.com

Marianela Rodriguez-Carres
Duke University
mr88@duke.edu

Dan Rokhsar
DOE Joint Genome Institute
DSRokhsar@gmail.com

Margie Romine
Pacific Northwest National Laboratory
margie.romine@pnl.gov

Francisco Ruiz-Dueñas
CIB, CSIC
fjruiz@cib.csic.es

Douglas Rusch
Center for Genomics & Bioinformatics
doug.rusch@gmail.com

Tamer Said
American University in Cairo
tsaid@aucegypt.edu

Jennifer Salazar
Argonne National Laboratory
jsalazar@anl.gov

Kenneth Sauer
Lawrence Berkeley National Lab
KHSauer@lbl.gov

Richard Sayre
New Mexico Consortium, LANL
rsayre@newmexicoconsortium.org

Christopher Schadt
Oak Ridge National Laboratory
schadtcw@ornl.gov

Michael Schatz
Cold Spring Harbor Laboratory
mschatz@cshl.edu

Gary Schroth
Illumina, Inc.
gschroth@illumina.com

Rekha Seshadri
SGI
rseshadri@syntheticgenomics.com

Joao Setubal
University of São Paulo
setubal@vbi.vt.edu

Henry Shaw
Lawrence Livermore National Lab
shaw4@llnl.gov

Lan-Xin Shi
University of California, Davis
shi@ucdavis.edu

Maria Shin
Eureka Genomics
maria@eurekagenomics.com

Jan Fredrik Simons
Ion Torrent
jan.simons@lifetech.com

Steven Singer
Lawrence Berkeley National Lab
SWSinger@lbl.gov

Yaron Sitrit
Ben-Gurion University
sitrit@bgu.ac.il

Jeffrey Skerker
University of California, Berkeley
skerker1@gmail.com

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

Douglas Smith
Synthetic Genomics
dgsmith@syntheticgenomics.com

Christina Smolke
Stanford University
csmolke@stanford.edu

Joseph Spatafora
Oregon State University
spatafoj@science.oregonstate.edu

Lisa Stein
University of Alberta
lisa.stein@ualberta.ca

Ramunas Stepanauskas
Bigelow Laboratory for Ocean Sciences
rstepanauskas@bigelow.org

Garret Suen
University of Wisconsin-Madison
gsuen@wisc.edu

Matthew Sullivan
University of Arizona
mbsulli@email.arizona.edu

Hui Sun
DOE Joint Genome Institute
hsun@lbl.gov

Sirisha Sunkara
DOE Joint Genome Institute
ssunkara@lbl.gov

Brandon Swan
Bigelow Laboratory for Ocean Sciences
bkswanny@gmail.com

Jennifer Talbot
University of Minnesota
jmtalbot@umn.edu

Mika Tarkka
Helmholtz-Center for Env. Research
mika.tarkka@ufz.de

Angela Tarver
DOE Joint Genome Institute
amtarver@lbl.gov

Neslihan Tas
Lawrence Berkeley National Lab
ntas@lbl.gov

Tootie Tatum
DOE Joint Genome Institute
oltatum@lbl.gov

Michael Thelen
JBEI
mthelen@llnl.gov

John Thomas
CodeSlurry
meetings@codeslurry.com

Brian Thompson
Bigelow Laboratory for Ocean Sciences
bthompson@bigelow.org

Claire Ting
Williams College
cting@williams.edu

Julien Tremblay
DOE Joint Genome Institute
jtremblay@lbl.gov

Heather Trumbower
Navigenics
heather.trumbower@gmail.com

Adrian Tsang
Concordia University
tsang@gene.concordia.ca

Anders Tunlid
Lund University
anders.tunlid@mbioekol.lu.se

Gillian Turgeon
Cornell University
bgt1@cornell.edu

Brett Tyler
Oregon State University
brett.tyler@oregonstate.edu

Daniel Van der Lelie
RTI International
vdlelied@rti.org

Tracy Vence
Genome Technology
tvence@genomeweb.com

Rytas Vilgalys
Duke University
fungi@duke.edu

Juan VillaRomero
University of California, Berkeley
villaromero@berkeley.edu

Axel Visel
DOE Joint Genome Institute
avisel@lbl.gov

John Vogel
US Department of Agriculture Center
john.vogel@ars.usda.gov

Christian Voolstra
King Abdullah Univ. of Science & Tech.
christian.voolstra@kaust.edu.sa

Qinhong Wang
Tianjin Inst of Industrial Biotechnology
Chinese Academy of Sciences
wang_qh@tib.cas.cn

Zhong Wang
DOE Joint Genome Institute
zhongwang@lbl.gov

David Ward
Montana State University
umbdw@montana.edu

David Weston
Oak Ridge National Laboratory
dweston.tn@gmail.com

Helen Whitaker
BioMed Central
helen.whitaker@biomedcentral.com

H. Wiley
Pacific Northwest National
LaboratorySteven.Wiley@pnnl.gov

Eske Willerslev
University of Copenhagen
ewillerslev@snm.ku.dk

Gordon Wolfe
California State Univ. Chico
gwolfe2@csuchico.edu

Attendees

Jennifer Wortman
Broad Institute
jwortman@broadinstitute.org

Tanja Woyke
DOE Joint Genome Institute
twoyke@lbl.gov

Stan Wullschleger
Oak Ridge National Laboratory
wullschlegsd@ornl.gov

Fangfang Xia
Argonne National Laboratory
fangfang@mcs.anl.gov

John Xu
New Mexico State University
jxu@nmsu.edu

Hugh Young
US Department of Agriculture
hugh.young@ars.usda.gov

Xiao-Ping Zhang
University of California, Davis
xpzhang@ucdavis.edu

Yubo Zhang
DOE Joint Genome Institute
yubozhang@lbl.gov

Tao Zhang
DOE Joint Genome Institute
tzhang3@lbl.gov

Zhiying Zhao
DOE Joint Genome Institute
zyzhao@lbl.gov

Carl Zimmer
Independent Journalist
carl@carlzimmer.com

Gerald Zon
geraldzon@sbcglobal.net

Author Index

Abbriano, Raffaela M. 65
Abdellateef, Mostafa 59
Ackerman, Eric 44
Adams, Paul 40
Aerts, Andrea 53
Amasino, Richard 37
Anderson, Olin 16
Anderton, Amy 22
Aranda, Manuel 13
Arif, Chatchanit 13
Arkin, Adam 1, 13, 16, 18, 25, 50, 73
Ashby, Meredith 64
Axen, Seth D. 79
Aylward, Frank O. 21
Baker, Scott 42, 60, 74
Baldrian, Petr 14
Ballor, Nicholas R. 15
Banfield, Jill 51
Banks, Jody 1
Bao, Xiaoming 58
Baran, Richard 16
Barry, Kerrie 37, 53
Beam, J. 47
Beaulaurier, John 42
Bellows, Wendy 66
Benner, Steven A. 1
Berka, Randy M. 44
Bernstein, H. 47
Best, Michael 78
Bhamidipati, Aruna 64
Bjornson, Keith 64
Blainey, Paul 52
Blow, Matt 78
Bonito, Gregory 62, 63
Bowen, Benjamin 16, 25
Bracho-Garrillo, Rosvel 80
Bradshaw, Rosie E. 53
Brady, Siobhan 2
Bragg, Jennifer 16, 22
Bramhacharya, Shanti 21
Brandt, Craig C. 36
Brazelton, William J. 17
Brenner, Steven 25
Brettin, Tom 18
Brodie, E.L. 67
Brooks, Scott C. 36
Brown, Daren 60
Brown, Michael 69
Brown, R. 47
Brumm, Phillip J. 21
Bruno, Kenneth S. 42
Bryant, Doug 37
Budak, Hikmet 37
Bullerjahn, George 29
Bulsara, Nadeem 25
Buscot, Francois 31
Bushley, Kathryn 54
Butler, Greg 59
Caicedo, Ana 37
Campbell, Dustin J. 46
Campbell, James H. 36
Canbäck, Björn 60
Canon, Shane 18, 72
Cardenas, Erick 75
Carlson, R. 47
Carroll, Susan 36
Cary, S.C. 19
Castanha, C. 67
Castro, Hector 63
Catalan, Pilar 37
Celio, Gail 52
Chaffin, Mark D. 66
Chandonia, John-Marc 25
Chen, Feng 48, 56, 68, 78, 79
Chen, Ling 58
Chen, Mei 58
Cheng, Jan-Fang 61, 79
Cheng, Lei 80
Cheng, Shaoan 55
Cheng, Xiaoliang 40
Chia, Jer-ming 73
Chiang, Hsin-I 74
Chin, Jason 42
Chivian, Dylan 13, 25, 50
Choi, Cindy 48
Chow, Virginia 20
Christopherson, Melissa 21
Chuck, George 22
Ciuffetti, Lynda 53
Clardy, Jon 71
Clark, Tyson A. 69
Clingenpeel, Scott 23, 38
Cock, Mark 70
Cole, Jim 80
Collett, James 42
Condon, Bradford J. 53
Copeland, Alex 53
Coradetti, Sam 3
Cottingham, Robert 13, 18, 25, 50
Cowan, D.A. 19
Craig, James 3
Crary, Benjamin 24
Criddle, Craig S. 36
Culley, David E. 42
Currie, Cameron R. 21
Curry, John D. 25
Cutcliffe, Colleen 64
Dahdouli, Mike 59
Dalal, Ravi 64
Dangl, Jeff 23
Datta, Supratim 79
Davison, Brian 13
de Wit, Pierre J.G.M. 53
DeAngelis, Kristen 38
Dehal, Paramvir 13, 25, 50
del Rio, Tijana Glavina 39, 52
Deng, Kai 40
Deng, Ye 80
Desai, Narayan 18, 50

Deutsch, Samuel 40
Deutschbauer, Adam M. 16
Deviod, Scott 18
Dhillon, Braham 53
Dibble, Dean 22
Dieckmann, Gerhard 70
Dier, Paul 25
DiFazio Stephen P. 26
Ding, Weijun 55
Disz, Terry 18
Dmitrieff, Elizabeth 66
Dodsworth, Jeremy A. 52
Doktycz, Mitchel J. 63
Doonan, John 37
Druzhinina, Irina S. 27
Dubrovsky, Joseph G. 49
Edgar, Robert C. 28
Edgar, Robyn 29
Egan, Rob 29
Eichler, Jerry 79
Eid, John 64
Eisen, Jonathan A. 61, 77
El Dorry, Hamza 34
Elgogary, Sawsan 30
El-Said, Nadine 34
Emerson, David 32
Eng, Kevin 78
Fang, Gang 73
Fang, Xu 58
Feau, Nicolas 53
Fedorov, Andrei 64
Feldhahn, Lasse 31
Fernández-Fueyo, Elena 31
Ferreira, Ari 30, 34
Field, Erin 32, 66
Fischer, M. 67
Fofanov, Viacheslav Y. 25, 28
Frank, A. Carolin 76
Frickenhaus, Stephan 70
Froula, Jeff 56, 78
Gallivan, Justin P. 2
Galloway, Michael 18
Garrett, Jinnie M. 33
Garvin, David 37
Gerstein, Mark 73
Ghazy, Mohamed 34
Ghylin, Trevor 35
Gihring, Thomas M. 36
Giovannoni, Stephen J. 3
Glaser, Fabian 53
Glass, Elizabeth M. 13
Glass, N. Louise 3
Gmitter Jr., Fred G. 4
Goldberg, Robert 5
Goodwin, Lynne A. 21
Goodwin, Stephen B. 53
Gordon, Paul M.K. 59
Gordon, Sean 16, 37
Gottel, Neil 63
Graveley, Brenton 69
Gray, Jeremy 64
Green, Stefan J. 36
Grell, Morten Nedergaard 60
Griesbach, Meghan 33
Grigoriev, Igor 31, 38, 42, 44, 48, 51, 53, 60, 64
Grimwood, Jane 53
Gross, Stephen 38, 48
Grosse, Ivo 31
Gryganskyi, Andrii 62
Gryndler, Milan 14
Gu, Yong 16
Gunter, Lee E. 26
Gurtowski, James 73
Hake, Sarah 22
Hallam, Steven 61, 75
Hamelin, Richard C. 53
Hammel, Kenneth E. 31
Hammond, Ming C. 72
Hanes, Jeremiah 64
Hartmann, Martin 75
Hasterok, Robert 37
Hawley, Erik 39
Hazen, Samuel 37
Hazen, Terry 15
He, Shaomei 39, 71
He, Zhili 80
Hedlund, Brian P. 52, 61
Heins, Richard 40
Hellsten, Uffe 26
Henrissat, Bernard 42, 53, 60
Henry, Chris 13, 25, 73
Hernández, Carlos 71
Hernandez, Myriam 71
Herrmann, Sylvie 31
Hess, Jaqueline 41
Hess, Matthias 39, 42, 56
Hesse, Cedar 53
Hettich, Robert 51
Hibbett, David 5, 60
Hildebrand, Mark 65
Hon, Lawrence 42
Hoover, Cindi A. 43
Horwitz, Benjamin A. 53
Hršelová, Hana 14
Hsiung, Pei-Lin 64
Hsu, David 64
Hugenholtz, Philip 61, 79
Hutchinson, Miriam I. 44
Inskeep, William P. 47, 61
Ivanova, Natalia 61, 79
Jakubauskaite, Agne 33
James, Timothy Y. 62
Jansson, Janet 15, 67
Jardine, Phil 36
Jay, Z. 47
Jennings, R. 47
Jeunger, Thomas 37
Jewell, Kelsea A. 21
Jiang, Homgmei 78
Jiang, Jinjin 78
Johansson, Tomas 60
Jorgensen, Richard 45
Jospin, Guillaume 77
Kadry, Maha 30
Kaeppler, Shawn 6
Kamtikar, Satwik 64
Kangasniemi, Ariel 46
Karp, Peter D. 45
Karpinets, Tatiana 63

Keegan, Kevin 50
Keller, Keith 25, 50
Kema, Gert 53
Kerfeld, Cheryl A. 79
Kerley, Marilyn 63
Kilpert, Fabian 70
Kinross, Cris 43
Kirton, Edward S. 68
Kits, K. Dimitri 46
Klammer, Aaron 42
Klenk, Hans-Peter 56
Kohler, Annegret 31
Konstantinidis, Konstantinos 80
Korlach, Jonas 42, 69
Koshinsky, Heather 25, 28
Kosti, Idit 53
Kostka, Joel E. 36
Kozik, Alexander 49
Kozubal, M. 47
Krell, Andreas 70
Kreuzer, H. 47
Kroth, Peter 70
Kukutla, Phanidhar 78
Kumari, Sunita 73
Kuo, Alan 48
Kurbanov, Feruz 64
Kurth, Florence 31
Kyrpides, Nikos 61, 79
Labbe, Jesse 63
LaButti, Kurt 53
Land, Miriam 18, 73
Lange, Lene 60
Latendresse, Mario 45
Latimer, Luke 72
Lawrence, Christopher 53
Lazo, Gerard 16
Lee, Lawrence 42, 69
Lee, C.K. 19
Lee, Janey 61
Lefebvre, Paul A. 52
Leung, Frederick C. 55
Leung, Hilary 75
Levasseur, Anthony 60
Li, Chenlin 22
Li, Dejun 80
Li, Jie 58
Li, Zhonghai 58
Liao, Hui-Ling 62
Lindholm-Perry, Amanda ... 25
Lindquist, Erika 48, 53, 60
Liou, Geoffrey 72
Lipton, M. 47
Lipzen, Anna 37
Liu, Guodong 58
Liu, JingTao 25
Liu, Rui 58
Liu, Weifeng 58
Liu, Wen-Tso 61
Liyou, Wu 80
Lluesma Gomez, Monica 66
Lopez-Valle, Mayra 49
Lowe, Kenneth 36
Lowry, Steve 53
Lucas, Susan 53
Lundberg, Derek 23
Luo, Chengwei 80
Luo, Yiqi 80
Luong, Khai 42
Lyle, John 64
Ma, Liang 58
Ma, Yanhe 58
Mackie, Rod I. 42
Magnuson, Jon K. 42
Malfatti, Stephanie 39, 61, 71
Manning, Gerard 60
Manzaneda, Antonio 37
Marei, Narguess 30
Martin, Francis 31, 60, 63
Martin, Jeffrey 38, 48
Martin, Joel 26, 37
Martínez, Angel T. 31
Martínez, María Jesús 31
Martinez-Garcia, Manuel 66
Mashek, Douglas 52
Mashek, Mara 52
Masland, E. Dashiell P. 66
Maslov, Sergei 13, 18, 25, 50, 73
Matvienko, Marta 49
Mavrommatis, Konstantinos 56
McCalmon, Sarah 64
McCarthy, Alex J. 74
McClellan, David A. 32
McCluskey, Kevin 74
McCormick, Michael L. 33
McDonald, I.R. 19
McKay, Robert Michael 29
McMahon, Katherine 35
McMahon, Katherine D. 24
Mead, David 21
Mehlhorn, Tonia L. 36
Meilan, Rick 22
Mengitsu, Sinafik 52
Messing, Joachim 37
Meyer, Folker 13, 18, 50
Miller, Christopher 51
Miller, Erik 64
Mitchell, Kendra 75
Mitchell-Olds, Thomas 37
Mock, Thomas 70
Mockler, Todd C. 37
Moeller, Joseph A. 21
Mohn, William 75
Mollova, Emilia 64
Moran, J. 47
Morgan, Hebatulla 30
Morin, Emmanuelle 60
Morris, Paul 29
Mosier, Annika 51
Moulton, Steve 18
Muchero, Wellington 26, 63
Mukhopadhyay, Biswarup 57, 68
Munk, A. Christine 21
Mur, Luis 37
Murillo, Catalina 71

Murphy, Devon 64
Murugapiran, Senthil K. 52
Nagy, Laszlo..... 60
Napsucialy-Mendivil, Selene..... 49
Natvig, Donald O. 44
Nelson, David R. 52
Nieu, Rita 22
Nong, Guang 20
Nordberg, Henrik 72
North, Gretchen..... 38
Northen, Trent 16, 40
Novichkov, Pavel 18, 25
Nunez, James..... 72
Oh, TaeYun 73
Ohm, Robin 51, 53
Olsen, Peter Bjarke..... 60
Olson, Andrew 73
Olson, Bob..... 18
Olson, Dan..... 18
Olson, Sara 69
Orozco-Hernandez, Aidee ... 38
Osman, Ihab 30
Otillar, Robert..... 53
O'Toole, Nicholas..... 59
Overbeek, Ross..... 18, 25
Overholt, Will..... 36
Owensby, C. Alisha..... 54
Pan, Chongle 51
Parchert, Kylea J. 44
Parello, Bruce..... 18
Park, Tae-Jin..... 55
Parker, Karen..... 55
Partida-Martinez, Laila..... 38
Pasternak, Shiran..... 18, 73
Pelletier, Dale 63
Peluso, Paul 64
Peña-Ramirez, Yuri 38
Pendery, Elizabeth..... 33
Peng, Yangfeng 58
Peng, Ze..... 56
Pennacchio, Christa P..... 26
Pennacchio, Len 26, 37, 78
Penton, Ryan 80
Persson, Per.....60
Phuntumart, Vipa 29
Piao, Hailan.....56
Pinto-Tomás, Adrián.....71
Podar, Mircea.....63
Porter, Teresita M. 62
Poulton, Nicole J.....66, 66
Powell, Amy J.....44
Preston, James F.....20
Price, Morgan N.....16
Priest, Henry 37
Pringle, Anne 41
Purwantini, Endang.....57, 68
Qi, Xianqi.....58
Qin, Yuqi 58
Qu, Yinbo.....58
Quake, Stephen R.....52
Quest, Daniel.....18
Ranjan, Priya.....26, 73
Rank, Dave.....64
Recht, Sabine 31
Redfern, Joanna L. 44
Reeves, Darryl.....73
Reichl, K. 67
Reid, Ian.....59
Rice, John D.....20
Rieseberg, Loren 6
Riley, Robert.....60
Rineau, Francois 60
Rinke, Christian 61
Roder, Cornelia 13
Rodgers-Melnick, Eli 26
Rodriguez-Carres, Marianela.....62
Rokhsar, Dan.....26, 37
Romine, M. 47
Ronald, Pam.....73
Roth, Doris.....60
Rubin, Eddy 40, 61, 79
Rubinelli, Peter 22
Ruegg, Thomas 15
Ruiz-Dueñas, Francisco J.....31
Rusch, D.....47
Said, Tamer.....34
Salamov, Asaf..... 53, 60
Salazar, Jennifer..... 13
Sale, Ken..... 40, 79
Sauer, Kenneth..... 62
Sayed, Ahmed..... 34
Sayre, Richard 7
Schackwitz, Wendy 26, 37
Schadt, Christopher W... 36, 63
Schatz, Michael 7, 73
Schmutz, Jeremy..... 53
Schnable, James..... 48
Schnittker, Rob R. 74
Schoch, Conrad L. 53
Schrenk, Matthew O. 17
Schuur, Edward A.G. 80
Schwartz, Christopher..... 37
Sczyrba, Alex..... 61
Seaver, Sam 73
Sebra, Robert 64, 69
Sensen, Christoph W. 59
Seriachik, Andrew 33
Setubal, Joao C. 57, 68
Shabalov, Igor..... 64
Shah, Firoz..... 60
Shakya, Migun..... 63
Sherman, David 71
Shin, Maria 25
Shishkova, Svetlana..... 49
Siam, Rania..... 34
Sieracki, Michael E..... 66, 66
Sievert, Stefan M. 61
Simmons, Blake..... 15, 22, 40, 44, 79
Simpson, June..... 38
Singer, Steven..... 15, 51
Singh, Anup..... 40
Singh, Kanwar 48, 68, 78
Singh, Seema 22
Skrede, Inger..... 41
Slavov, Gancho T. 26
Smith, Jason..... 55
Smith, Sarah R. 65
Smits, Mark 60
Smolke, Christina 8

Spatafora, Joseph W. 53, 54
St. John, Franz 20
Stein, Lisa Y. 46
Stepanauskas, Ramunas .. 9, 32, 61, 66, 66
Steritz, Matthew..... 78
Stevens, Rick 13, 18, 25, 50
Suen, Garret..... 21, 71
Sullivan, Matthew B. 9
Sun, Hui..... 53
Sun, Jianping 3
Sun, Lan..... 22
Swan, Brandon K..... 66, 66
Syed, Aijazuddin 72
Syed, Musstafa 73
Tamayo-Castillo, Giselle 71
Tarkka, Mika T. 31
Taş, N. 67
Tchourbanov, Alexander 78
Thallman, R. Mark..... 25
Thomas, Brian 51
Thomason, Jim 73
Thompson, Brian 66
Thrash, J. Cameron..... 3
Tiao, G. 19
Tiedje, James M..... 80
Titmus, Matt 73
Tobias, Christian..... 22
Torn, M.S..... 67
Torto-Alalibo, Trudy 57, 68
Toseland, Andrew..... 70
Tran-Gyamfi, Mary 40
Travers, Kevin 64, 69
Travers, Michael..... 45
Tremblay, Julien 68
Trimble, William 50
Tringe, Susannah 23, 38, 39, 42, 47, 52, 68, 71
Tsang, Adrian 59
Tsiamis, George..... 61
Tu, Ran 58
Tuna, Metin 37
Tunlid, Anders..... 60
Tupper, Benjamin..... 66
Turgeon, B. Gillian..... 53
Tuskan, Gerald 26, 63
Tyler, Brett M.....57, 68
Tyler, Ludmila..... 37
Uberbacher, Edward..... 63
Ugartechea-Chirino, Yamel. 49
Uhlig, Christiane 70
Underwood, Jason G. 69
Urbanová, Michaela 14
Valant, Valerie 33
Valentin, Klaus.....70, 70
Van Insberghe, David..... 75
Van Nostrand, Joy D. 80
Vargas-Asensio, Gabriel 71
Vijayendra, Pooja..... 72
Vilgalys, Rytas 62, 63
Visel, Axel..... 38
Vogel, John 16, 22, 37
Voolstra, Christian R. 13
Waldrop, Mark 39
Wang, Baowei 58
Wang, Chengshu 58
Wang, Kai 72
Wang, Lushan 58
Wang, Mei 48
Wang, Qinhong 58
Wang, Shengyue..... 58
Wang, Tianhong..... 58
Wang, Wenqin..... 37
Wang, Ying 78
Wang, Zhong.....29, 38, 48, 56, 56, 72, 78
Ware, Doreen 13, 73
Watson, David B. 36
Webster, Dale..... 69
Wei, Chia-Lin..... 48
Wei, Christopher 74
Wei, Xiaomin 58
Welschmeyer, Nick 55
Weston, David.....13, 73
Wiest, Aric E. 74
Wilhelm, Roland 75
Wilke, Andreas..... 50
Wilkening, Jared 18, 50
Willerslev, Eske 10
Wilson, Emily C.....76
Windham-Myers, Lisamarie.....39
Wolfe, Benjamin 41
Woo, Hannah 15
Woyke, Tanja.....23, 38, 61
Wu, Dongying.....77
Wu, Jiajie 16
Wu, Sophia.....64
Wu, Wei-Min 36
Wubet, Tesfaye 31
Wullschleger, Stan D. 10
Xu, Hai.....58
Xu, Jiannong 78
Xu, Xia.....80
Xue, Kai 80
Yan, Xing.....58
Yang, Alicia 64
Yang, Zamin 36, 63
Yehia, Ayman 34
Yilmaz, Suzan.....40
Yiong, Yi 3
Yoo, Shinjae.....73
Youens-Clark, Ken 73
Yu, Dantong.....18, 73
Yuan, Mengting 80
Zhang, Gengxin 36
Zhang, Jin.....80
Zhang, Jun.....58
Zhang, Kun 74
Zhang, Lei.....58
Zhang, Tao 78, 79
Zhao, Guo-Ping.....58
Zheng, Huajun 58
Zheng, Kai 58
Zhong, Shaobin.....53
Zhong, Yaohua.....58
Zhou, Jizhong.....80
Zhou, Zhihua.....58
Zimmer, Carl.....11
Znameroski, Elizabeth 3
Zou, Gen 58

Notes