Second Annual
DOE Joint Genome Institute
User Meeting

Sponsored By

U.S. Department of Energy
Office of Science

March 28–30, 2007

Marriott Hotel

Walnut Creek, California

Contents

Speaker Presentations

Abstracts alphabetical by speaker.

The JGI *Aspergillus niger* Genome Project

Scott E. Baker (scott.baker@pnl.gov)

Fungal Biotechnology Team, Chemical and Biological Process Development Group, Pacific Northwest National Laboratory, Richland, WA

Aspergillus niger is an economically important filamentous ascomycete fungus that is used in industry for its prodigious production of citric acid and a number of enzymes. The DOE Joint Genome Institute has sequenced the genome of *A. niger* ATCC 1015, a wildtype strain and the source of the first patented microbial fermentation process for citric acid production. Preliminary annotation indicates the presence of over 250 glycosyl hydrolases. These enzymes are crucial for the degradation of lignocellulosic biomass into simple sugars and other chemical building blocks. Comparison of the JGI sequence with genomic sequence from an *A. niger* strain used for protein production (CBS 513.88, DSM Heerlen, Netherlands) reveals at least two major chromosomal rearrangements that are indicative of an early speciation event and/or induced mutagenesis protocols used in industrial strain improvement programs.

A Genomic Encyclopedia for Bacteria and Archaea

Jonathan Eisen (jaeisen@ucdavis.edu)

University of California, Davis, CA

Integrated Genome-Based Studies of *Shewanella* Ecophysiology

Jim Fredrickson (Jim.Fredrickson@pnl.gov)

Pacific Northwest National Laboratory, Richland, WA

Shewanella oneidensis MR-1 is a motile, facultative γ-Proteobacterium with remarkable respiratory versatility; it can utilize a range of organic and inorganic compounds as terminal electron acceptors for anaerobic metabolism. To function and compete in environments that are subject to spatial and temporal environmental change, *Shewanella* must be able to sense and respond to such changes and therefore require relatively robust sensing and regulation systems. The overall goal of this project is to apply the tools of genomics, leveraging the availability of genome sequence for 18 additional strains of *Shewanella*, to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus. To understand these systems we propose to use genome-based approaches to investigate *Shewanella* as a *system of integrated networks*; first describing key cellular subsystems – those involved in signal transduction, regulation, and metabolism - then building towards understanding the function of whole cells and, eventually, cells within populations. As a general approach, this project employs complimentary "top-down" – bioinformatics-based genome functional predictions, high-throughput expression analyses, and functional genomics approaches to uncover key genes

as well as metabolic and regulatory networks. The "bottom-up" component employs more traditional approaches including genetics, physiology and biochemistry to test or verify predictions. This information will ultimately be linked to analyses of signal transduction and transcriptional regulatory systems and used to develop a linked model that will contribute to understanding the ecophysiology of *Shewanella* in redox stratified environments.

Spironucleus vortens and *Naegleria gruberi*, Basal Eukaryotic Genomes

Lillian Fritz-Laylin[1](fritz-laylin@berkeley.edu), W. Zac Cande,[1] and Scott Dawson[2]

[1]Department of Molecular and Cell Biology, University of California, Berkeley, CA and [2]Section of Microbiology, University of California, Davis, CA

All but a few of the microbial eukaryotic genome projects are of disease-causing organisms. Although useful for obvious medical reasons, these genome sequences do not accurately describe the general complexity and ecological significance of basal eukaryotes, due to their parasitic lifestyle. Genome sequence analysis of free-living basal eukaryotes will greatly enhance our understanding of basal eukaryotic metabolism, as well as cytoskeletal diversity and evolution, and make possible the integration of classical cytological and functional descriptions with modern genomic approaches. Through the Community Sequencing Program, we are beginning to analyze the genomes of two such protists, *Spironucleus vortens* and *Naegleria gruberi*.

Spironucleus vortens is a free living diplomonad flagellate relative of the intestinal parasite *Giardia intestinalis*. Comparative genomic analysis of the two diplomonads will allow us to identify what genes and features of cellular processes are unique to Giardia's parasitic life style and what features are shared among all diplomonads. Additionally, it will allow us to describe the metabolism of a free-living diplomonad, thought to be one of the most basal groups of eukaryotes.

Naegleria gruberi is a free-living protist that lives in soil, fresh and marine waters, as well as in some extreme environments. Its genome represents the first in its kingdom level group of eukaryotes, the heteroloboseans. This group of amoeba is thought to be vital in nutrient cycling within soil communities. Naegleria grows and divides as an amoeba, feeding on bacteria and other microbes. When stressed it becomes a flagellate, forming *de novo* a microtubule cytoskeleton including basal bodies and flagella. The genomic analysis of this interesting amoeba will facilitate insights into the evolution of amoeboid cell movement, regulation of cell differentiation, the evolution of the cytoskeleton, and mitosis and meiosis.

Sequencing on the Mycosphaerella Branch of the Fungal Evolutionary Tree

Stephen B. Goodwin[1] (sgoodwin@purdue.edu) and Gert H. J. Kema[2]

[1]USDA Agricultural Research Service, Purdue University, West Lafayette, IN and [2]Plant Research International B.V., Wageningen, The Netherlands

Mycosphaerella and its asexual derivatives comprise the largest group of plant-pathogenic fungi, both for number of species and in the diversity of hosts attacked. Many diseases

caused by species of Mycosphaerella are economically significant and require huge inputs of fungicides for effective control. These pathogens also reduce the yield and quality of biomass used for ethanol production, so can negatively affect the supplies of both food and energy. Despite these huge economic costs, species of Mycosphaerella have been ignored for genomic sequencing until recently. This changed with the completion of the genomic sequence of the Septoria tritici blotch pathogen of wheat, *M. graminicola*, and of a 7.1x draft sequence of the black Sigatoka pathogen of banana, *M. fijiensis*. Comparison of these sequences revealed a huge and unexpected difference in genome size, from about 40 Mb for *M. graminicola* to more than 74 Mb for *M. fijiensis*. This difference apparently originated in one of the ancestors to a group of Mycosphaerella species attacking banana, and is due to amplification of transposable elements rather than to genome duplications. Some of the transposons appear to be helitrons, which have not been reported previously from ascomycete fungi. Both species seem to have a slightly smaller and possibly more specialized set of cell-wall-degrading enzymes in comparison to other plant-pathogenic fungi with sequenced genomes, which may indicate a mode of pathogenesis that involves stealth rather than brute force. Unlike most other plant pathogens, species of Mycosphaerella have a long latent phase (from seven to more than 60 days) during which they appear to grow almost as endophytes, and then they switch to become pathogens. The genomic sequence will provide an unprecedented picture of the genetic changes that occur during this switch. Enzymes that are required for pathogenicity may indicate new fungicide targets and could be useful for improving ethanol production. The genomic sequences also clarified the phylogenetic relationships of this group and showed that their closest relatives are fungi in the order Pleosporales, in accordance with classical taxonomy. In addition to elucidating evolutionary biology and the molecular basis for host-pathogen interactions, the sequences are being used by a consortium of scientists worldwide to design improved methods of disease control for increased food production at reduced cost.

Genome Sequencing Reveals Ecotypes of *Rhodopseudomonas palustris*, a Phototrophic Soil Bacterium

Caroline S. Harwood (csh5@u.washington.edu)

University of Washington, Seattle, WA

Rhodopseudomonas palustris is a purple nonsulfur bacterial species of unusual versatility that has the potential to be used as a biocatalyst for hydrogen production, carbon sequestration, biomass turnover, and biopolymer synthesis. The genome of *R. palustris* strain CGA009 has been reported and consists of a 5.46 Mb chromosome with 4836 predicted protein-coding genes. Recently four additional strains of *R. palustris* have been sequenced. Each genome is approximately the same size as the strain CGA009 genome. In addition, approximately 60 to 80% (depending on the strain) of the genes from strain CGA009 were present in each strain, and these may represent the core genes of the *R. palustris* species. Conserved core metabolic characteristics include photosynthesis, nitrogen fixation and carbon dioxide fixation. Against this backdrop of conservation, whole genome comparisons among strains showed a high degree of genome rearrangement in terms of gene orders and reading directions. Furthermore, there were high numbers of genes (250 to 560 genes) that were specific to a given strain and not seen in any other strain. Based on their gene inventories, each strain is predicted to have strain-specific physiological traits. Strain CGA009 is especially well-equipped for nitrogen fixation with three nitrogenase isozymes and four sets of glutamine synthetases and strain BisB5 has expanded anaerobic aromatic degradation capabilities (e.g., phenylacetate degradation). Strain HaA2 should be well-adapted for growth in oxygen as it encodes seven different

aerobic terminal oxidases, and strain BisB18 should be able to grow well anaerobically in dark (e.g., carbon-monoxide dehydrogenase genes, three sets of pyruvate-formate lyase genes, formate-hydrogen lyase genes, and DMSO reductase genes). Finally, strain BisA53 has an expanded set of exopolysaccharide synthesis genes and readily attaches to surfaces to form biofilms. This strain also has an unusually large number (seven) of light harvesting II gene sets. Our comparative genomic analysis suggests that *R. palustris* is a dynamic species comprised of diverse ecotypes that are well adapted to specific environmental niches.

Engineering Microbes for Production of Low-Cost, Effective, Anti-Malarial Drugs

Jay D. Keasling[1,2] (JDKeasling@lbl.gov)

[1]Departments of Chemical Engineering and Bioengineering, University of California, Berkeley, CA and [2]Department of Synthetic Biology, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA

Malaria infects 300–500 million people and causes 1-2 million deaths each year, primarily children in Africa and Asia. More than half of the deaths occur among the poorest 20% of the world's population. One of the principal obstacles to addressing this global health threat is a lack of effective, affordable drugs. The chloroquine-based drugs that were used widely in the past have lost effectiveness because the *Plasmodium* parasite which causes malaria has become resistant to them. The faster-acting, more effective artemisinin-based drugs — as currently produced from plant sources — are too expensive for large-scale use in the countries where they are needed most.

We have metabolically engineered *E. coli* to produce high levels of mono-, sesqui-, and diterpenes, most notably the sesquiterpene precursor to artemisinin, amorphadiene. The result of these studies is an *E. coli* host capable of producing 1,000,000-fold higher levels of amorphadiene than the strains and expression systems that had been available previously. The engineered strain contains a heterologous mevalonate-based terpene biosynthetic pathway and an amorphadiene cyclase gene resynthesized with the *E. coli* codon usage. Recently, we cloned the final steps in the artemisinin biosynthetic pathway and engineered yeast to produce artemisinic acid at high levels. The development of this technology will eventually reduce the cost of artemisinin-based combination therapies significantly below their current price.

De novo Hybrid 454-Sanger Assembly of a Draft Genome of *Phytophthora capsici*

Stephen Kingsmore (sfk@ncgr.org)

National Center for Genome Resources, Santa Fe, NM

Phytophthora capsici is a devastating oomycete (water mold) pathogen of vegetable crops. *P. capsici* has a broad host range with recent expansion, hemibiotroph lifestyle, and is a model for *Phytophthora* genetic & functional genomic studies. *P. capsici* is an eukaryote and has a highly repetitive, 60Mb genome. With funding from DOE-CSP, USDA and NSF, we undertook *de novo* draft genome sequencing of *P. capsici,* featuring hybrid assembly of 23X 454 GS20 pyrosequencing singleton reads, 5X Sanger paired reads and

2M 454 GS20 paired reads. e will present the results of an evaluation of 454 GS20 sequencing technology, hybrid assembly using FORGE and annotation issues. In addition, we will demonstrate the use of the Alpheus™ software system for compiling a searchable SNP and indel database of *P. capsici*.

New Amplification and Cloning Tools

David Mead (dmead@lucigen.com)

Lucigen Corp, Middleton, WI

Accurate cloning and sequence assembly are hindered by several major hurdles, including: 1) restriction-site bias in BAC cloning, 2) clone bias against structure rich sequences, repeats or AT-rich DNA, 3) chimeric clones in shotgun libraries, 4) cloning of trace amounts of template, and 5) non-specific products in whole genome amplification. Our group has recently developed technologies to address each of these problems. Centromeric and other highly repetitive genomic regions are absent or vastly under-represented in typical partial-digest BAC libraries. We have developed a method to randomly shear and clone DNA fragments of >100 kb. “Random Shear” BAC libraries show uniform coverage over regions that are over- or under-represented in conventional BAC libraries. For cloning fragments of up to 30 kb, the novel “pJAZZ” linear cloning vector provides unprecedented ability to maintain regions that are unclonable in circular plasmids. Examples include large, highly AT-rich fragments of 20-30 kb and regions of di-, tri-, and tetra-nucleotide repeats. The pJAZZ linear vector has been used to sequence and assembly an AT rich genome and the results demonstrate bias free closure without the use of BAC or fosmid libraries, saving considerable time and expense. We have also developed a method of “GC Cloning” to minimize chimeric inserts in shotgun libraries. A simple tailing reaction appends a 3’-G residue to target fragments, which are ligated to a vector with a 3’-C tail. The frequency of chimeras in the resulting libraries is less than 1%, simplifying the process of sequence assembly. Another bottleneck in genomic analysis is cloning low amounts of DNA. The high efficiency of GC Cloning allows direct cloning of nanogram amounts of DNA without template amplification. For library construction from even smaller samples, e.g., isolated cells, rare microbes, or metagenomic communities, we demonstrate a method to anonymously amplify and clone picogram amounts of DNA. Inserting the amplified DNA into transcription-free pSMART vectors reduced cloning bias against toxic sequences. Finally, a new thermostable phage DNA polymerase allows isothermal whole genome amplification at elevated temperatures. Based on strand displacement from nicked DNA rather than from random primers, this amplification method eliminates the background associated with de novo synthesis from exogenous primers. These methods and vectors provide improved cloning of fragments from all sizes and from diverse sources.

Insect Symbionts: From Bacterial Genomics to Host Ecology

Nancy Moran (nmoran@email.arizona.edu)

University of Arizona, Tucson, AZ

Heritable symbionts colonizing insect hosts have arisen often, from a variety of bacterial groups. These symbioses have shaped insect evolution and affect many terrestrial ecological communities. The symbionts are usually noncultivable, but analyses of their

genomic sequences have provided insights into their roles, particularly their provisioning of nutrients to hosts. In many obligate, ancient symbioses, exemplified by that between aphids and *Buchnera aphidicola*, the symbiont genomes are extremely reduced, accelerated in rates of sequence evolution, stable in gene arrangements, biased towards adenine and thymine, and lacking in regulatory capabilities. Because these organisms are highly derived relative to free-living ancestors, more recent symbioses can provide more insight into the initial genomic changes following adoption of a symbiotic lifestyle. Many examples within Enterobacteriaceae show parallels to well-studied pathogens of mammals. For example, *Candidatus Hamiltonella defensa*, a facultative symbiont of aphids, contains Type III Secretion Systems and homologs of pathogenicity genes, but confers benefits on its hosts, including defense against parasites and resistance to heat stress. Thus, some of the same genes that permit some organisms to be pathogenic may foster beneficial symbiosis in their close relatives. Ancient nutritional symbionts of insects possess compact, highly stable genomes, but early stages of evolution of symbionts feature rearrangements, large deletions, pseudogene accumulation, phage-borne genes, and proliferation of insertion sequences. Many features of symbiont genomes are most readily explained by a high level of genetic drift and by reduced recombination within symbiont populations.

Biological Large Scale Integration

Steve Quake (quake@stanford.edu)

Stanford University, Stanford, CA

Explorations of the Human Microbiome

David Relman (relman@stanford.edu)

Stanford University, Palo Alto, CA

Complex microbial ecosystems occupy the cutaneous and mucosal surfaces of humans. Recent advances have highlighted both the tremendous diversity of these communities and their importance to host physiology, but questions remain about the ecological processes that establish and maintain the human microbiota throughout life. Furthermore, basic features of the human microbial ecosystem remain poorly described, including variability in diversity, in space and time. Host individuality imposes a strong signature on patterns of diversity. Assembly of the oral and the gut microbiota may also involve both stochastic historical events and contemporary environmental factors. Approaches that combine community ecology, molecular microbial ecology, and metagenomics may improve our understanding of health and disease within the communal human organism. By understanding the patterns of microbial genetic, genomic, and functional diversity associated with human health, we may be able to preserve and restore health more effectively. By recognizing the early signs of impending disturbance, we may be able to predict and avoid disease.

Nitrosopumilus maritimus: An Ammonia-Oxidizing Archaeon

David A. Stahl[1] (dastahl@u.washington.edu), José R. de la Torre,[1] Christopher B. Walker,[1] Martin Könnecke,[2] and Willm Martens-Habbena[1]

Department of Civil and Environmental Engineering, University of Washington, Seattle, WA and [2]Paleomicrobiology Group, University of Oldenburg, Oldenburg, Germany

Discovered over 15 years ago using cultivation-independent gene surveys, nonthermophilic members of the Crenarchaeota are now recognized as major components of microbial assemblages in diverse aquatic and terrestrial environments. Isotopic analyses of membrane lipids from marine crenarchaeotes suggested possible autotrophy. However, the absence of cultivated representatives hindered the definitive elucidation of their physiology and of their potential role in biogeochemical cycles of carbon and nitrogen. Partial genome sequence data (~1.6 Mb, ~1626 ORfs) from *Nitrosopumilus maritimus* now serves for comparison of a cultured representative to closely related environmental populations that have not been propagated in culture, a sponge symbiont (*Cenarchaeum symbiosum*) and those represented by environmental archaeal contigs (Sargasso Sea, soil). Preliminary comparative analyses have revealed significant synteny between the genomes of *N. maritimus* and *C. symbiosum*. In the partial genome we have so far encountered single copies of genes encoding a putative archaeal ammonia monooxygenase, an ammonia permease, and an rRNA operon. The autotrophic metabolism of *Nitrosopumilus* is attributable to either the 3-hydroxypropionate pathway or a reverse TCA cycle, consistent with partial pathways of CO_2 fixation reported for *Cenarchaeum symbiosum*. In addition, some capacity for utilization of fixed carbon is suggested by the presence of genes encoding putative amino acid and oligopeptide transporters. Several genes encode putative transporters for metals, including Mn, Zn and Ni.

Comparative Genomics of Stalked Bacteria

Craig Stephens (cstephens@scu.edu)

Biology Department, Santa Clara University, Santa Clara, CA

Caulobacters are Gram-negative aquatic bacteria from the alpha-Proteobacteria clade, of interest to microbiologists for several reasons, including: (1) a unique dimorphic lifecycle that includes motile (flagellated) and non-motile (stalked) cell types; (2) formation of tightly adherent, stable biofilms on submerged surfaces; and (3) ubiquitous presence in nutrient-depleted (oligotrophic) habitats, suggesting possible utility for bioremediation applications in such habitats. The freshwater species *Caulobacter crescentus* is a model for examining fundamental issues of cellular development. Sequencing of the *C. crescentus* genome facilitated dramatic leaps in understanding the developmental biology of this organism, through the application of systems biology approaches. Less effort has gone into applying genomics to understanding the environmental biology of Caulobacters.

JGI recently sequenced the genomes of *Caulobacter* strain K31 (isolated from a chlorophenol-contaminated groundwater in Finland, 97% identical in 16S rRNA sequence to *C. crescentus*), and *Maricaulis maris* (isolated from Puget Sound, 91% identical in 16S rRNA sequence to *C. crescentus*), in hopes of gaining new insights in the evolution, physiology, and environmental biology of Caulobacters. The *C. crescentus* genome is just over 4 Mb in size. The completed *M. maris* genome is notably smaller (3.3 Mb), while the draft-level sequence of the K31 genome indicates a larger, 5.8 Mb chromosome. The K31 genome sequence includes two putative plasmids, approximately 0.3 and 0.2 Mb in size.

Despite both new genomes having more genes in common with *C. crescentus* than any other sequenced microbial genome, both show only patchy synteny with *C. crescentus*, suggesting extensive recombination and gene acquisition or loss. K31 in particular shows a large number of transposons and insertion sequences.

With respect to developmental biology, the newly-sequenced strains produce swarmers that appear to be structurally and functionally similar to *C. crescentus*, but the stalked cells are distinctive: *M. maris* produces straight cells with longer, thicker stalks than *C. crescentus*, while K31 produces smaller curved cells with stubby or barely visible stalks. Both strains attach to surfaces and form biofilms. Examination of putative cell cycle/developmental regulators, and known components of developmentally-regulated structures (flagellum, holdfast, pili, and cytoskeletal proteins involved in division) show generally high sequence conservation with *C. crescentus*, suggesting that processes contributing to morphological development have been evolutionarily conserved. One notable exception is that crescentin, an intermediate filament-like protein implicated in producing curvature in *C. crescentus* cells, is encoded in K31 (which also produces curved cells), but absent in *M. maris*, which has straight cells.

In terms of physiology, the K31 genome appears to encode a much larger repertoire of enzymes involved in electron-transfer reactions than either *C. crescentus* or *M. maris*, including more cytochrome c variants, respiratory nitrate reductase, and many putative dehydrogenases that are lacking in *C. crescentus*. As a groundwater resident, K31 likely faces lower dissolved oxygen levels (if not outright anoxia) on a more routine basis than the surface-dwelling *C. crescentus*, and may find more respiratory versatility advantageous. We will discuss these and other genomic insights into physiological regulation in K31 and *M. maris* in more detail.

Burkholderia

Jim Tiedje (tiedjej@msu.edu)

Michigan State University, East Lansing, MI

Brachypodium distachyon, An Emerging Model Grass

John Vogel[1] (jvogel@pw.usda.gov), Yong Q. Gu,[1] Ming-Cheng Luo,[2] Gerard R. Lazo,[1] Naxin Huo,[1] David F. Garvin,[3] Michael W. Bevan,[4] Todd C. Mockler,[5] Olin D. Anderson,[1] Daniel M. Hayden,[1] Debbie Laudencia-Chingcuanco,[1] Todd Michael,[6] Sam Hazen,[7] and Jeff Chang[5]

[1]USDA Agricultural Research Service, Western Regional Research Center, Albany, CA; [2]Department of Plant Sciences, University of California, Davis, CA; [3]USDA Agricultural Research Service, Plant Science Research Unit, St. Paul MN; [4]John Innes Centre, Norwich, UK; [5]Oregon State University, Corvallis, OR; [6]Salk Institute, La Jolla, CA; and [7]Scripps Institute, La Jolla, CA

The value of the model system approach to modern plant biology has been powerfully illustrated by the dicot species *Arabidopsis thaliana*. Unfortunately, a dicot model is not suitable for answering questions about grasses where monocot and dicot biology diverge (e.g. cell wall composition). Rice is not an ideal model either because it is a specialized semi-aquatic tropical grass whose large size, demanding growth requirements and long generation time make experiments involving large numbers of plants very expensive.

Brachypodium distachyon is a small temperate grass with all the attributes needed to be a modern model organism including simple growth requirements, fast generation time, small stature, small genome size and self-fertility. For these reasons, *Brachypodium* is rapidly emerging as a model system to study questions unique to the grasses. The emergence of *Brachypodium* as a model system is coincident with an increased need for basic research in grass biology to develop perennial grasses as a source of renewable fuel. This need has led the U.S. Department of Energy to suggest *Brachypodium* as a model system for herbaceous energy crops (Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda 2006) and has played a role in the selection of *Brachypodium* for complete genome sequencing by the DOE Joint Genome Institute in 2007. An overview of *Brachypodium*, the genomic tools developed to date (high efficiency *Agrobacterium*-mediated transformation, EST sequences, BAC end sequences, a physical map, a genetic map and insertional mutagenesis) and the status of the genome sequencing project will be presented.

Vibrio furnissii and the Search for Superior Biofuels

Lawrence P. Wackett (wacke003@umn.edu), Dave Sukovich, Janice Frias, Jennifer Seffernick, and Stephan Cameron

Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, St. Paul, MN

The need to develop renewable energy sources will require that society develop biofuel options other than ethanol. One excellent fuel option would be bio-alkanes. Alkanes comprise the major component of current petroleum-based fuels and are completely compatible with existing infrastructure. Thus, considerable interest was generated by recent reports of high-level *n*-alkane formation by the bacterium *Vibrio furnissii* M1 (1-3). In the present study, *V. furnissii* M1 was obtained, subjected to genomic analysis, and studied biochemically. The sequence of the 16S rRNA gene and repetitive (REP)-PCR showed that *V. furnissii* M1 was not identical to other *V. furnissii* strains tested, but the level of relatedness was consistent with its assignment as a *V. furnissii* strain. Pulse field gel electrophoresis showed the presence of chromosomal bands at approximately 3.2 and 1.7 Mb, similar to other *Vibrio* strains. Complete genomic DNA from *V. furnissii* M1 was sequenced with approximately 25-fold coverage. Genome annotation data will be discussed. However, in our studies, *V. furnissii* M1 did not produce appreciable levels of *n*-alkanes. *In vivo* experiments were conducted by growing *V. furnissii* M1 under different conditions, extracting with solvent, and analyzing extracts by gas chromatography-mass spectrometry. A highly sensitive assay was used for *in vitro* experiments with cell-free extracts and [^{14}C]-hexadecanol. The data are consistent with the present strain being a *V. furnissii*, similar to that described previously by Park and coworkers (1-3), but lacking the alkane-producing phenotype.

References

1. Park, M.O., M. Tanabe, K. Hirata, and K. Miyamoto. 2001. Isolation and characterization of a bacterium that produces hydrocarbons extracellularly which are equivalent to light oil. *Appl. Microbiol. Biotechnol.* **56:**448-452.
2. Park, M.O., K. Heguri, K. Hirata, and K. Miyamoto. 2005. Production of alternatives to fuel oil from organic waste by the alkane-producing bacterium, *Vibrio furnissii* M1. *J. Appl. Microbiol.* **98**:324-331.
3. Park, M.O. 2005. New pathway for long-chain *n*-alkane synthesis via 1-alcohol in *Vibrio furnissii* M1. *J. Bacteriol.* **187:**1426-1429.

The Rumen Microbiome, A View through the Fistula

Bryan A. White[1,10] (bwhite44@uiuc.edu), Jennifer M. Brulc,[1] Dionysios A. Antonopoulos,[2] Margret E. Berg,[1,10] Melissa K. Wilson,[1] Robert E. Edwards,[3, 4, 5] Edward D. Frank,[6] Joanne B. Emerson,[7] Pedro M. Coutinho,[8] Bernard Henrissat,[8] and Karen E. Nelson[7,9]

[1]Department of Animal Sciences, University of Illinois, Urbana, IL; [2]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI; [3]Fellowship for Interpretation of Genomes, Burr Ridge, IL; [4]Center for Microbial Sciences, San Diego State University, San Diego, CA; [5]The Burnham Institute, San Diego CA; [6]Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL; [7]The Institute for Genomic Research, Rockville, MD; [8]Architecture et Fonction des Macromolecules Biologiques, Universites Aix-Marseille I & II, Marseille, France; [9]Department of Biology, Howard University, Washington, DC; and [10]The Institute for Genomic Biology, University of Illinois, Urbana, IL

The complex microbiome of the rumen functions as a uniquely effective system for the conversion of plant cell wall biomass to microbial protein, short chain fatty acids, and gases including hydrogen and methane. As such, it provides a unique genetic resource for plant cell wall degrading microbial enzymes that could be used in the production of biofuels from lignocellulose. The rumen/gastrointestinal tract harbors a dense and complex microbiome, comprising an estimated 500-1000 native microbial species, of which less than 10% have been cultivated and characterized. In order to gain a greater understanding of the ecology of this lignocellulose degrading microbiome, we used comparative metagenomics to examine random sampled pyrosequence data from three fiber-associated microbiomes and one pooled sample derived from a mixture of the planktonic microbiome fraction from the same bovine rumens. We used phylotype analysis and the subsystems-based annotations available in the SEED database to gain a better understanding of the metabolic potential of these microbiomes. Even though the three animals were fed the same diet, the community structure and predicted metabolic potentials in the rumen foregut were markedly different with respect to nutrient utilization. The fiber-associated microbiome from the rumen of one of the animals showed a marked increase in the *Gammaproteobacteria*, and a shift between the distribution of *Bacteriodetes* and *Bacilli*, when compared to the other fiber-associated samples. This microbiome also contained significantly more coding sequence similarities to metabolisms that are consistent with a community that has shifted from a carbohydrate-based metabolism to a protein and amino acid based metabolism. We also compared the glycoside hydrolase and cellulosome (dockerins and cohesins) functional gene content in the different microbiomes. It appears that the rumen microbiome has evolved a huge diversity of enzymes to hydrolyze the easily available side chains of plant polysaccharides, but the hydrolysis of the more recalcitrant main chains, especially cellulose, probably resides in less abundant organisms. This has profound implications for identifying novel enzymes for the production of biofuels from lignocellulose and for comparative metagenomics.

Impact of Secondary Endosymbiosis on Eukaryotic Genome Evolution and Cell Biology: A Cryptomonad and a Chlorarachniophyte Nuclear Genome

John M. Archibald* (jmarchib@dal.ca),[1] Michael W. Gray,[1] Patrick J. Keeling,[2] Geoffery I. McFadden,[3] and Christopher E. Lane[1]

[1]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada; [2]Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada; and [3]School of Botany, University of Melbourne, Australia

The process of endosymbiosis has been a monumental force in the origin and diversification of eukaryotic life. The primary endosymbiotic origin of plastids (chloroplasts) occurred more than a billion years ago and spawned three lineages—the green algae (and their land plant relatives), red algae and glaucophytes—whose energy-generating capabilities paved the way for a transformation of the biosphere. The photosynthetic organelles of red and green algae have spread to unrelated eukaryotes by *secondary endosymbiosis*—the engulfment and retention of an algal cell inside a non-photosynthetic host. Secondary endosymbiosis has given rise to some of the most abundant and ecologically significant aquatic photosynthesizers on the planet, including the heterokonts (e.g., diatoms and giant kelp), haptophytes (e.g., *Emiliania*), and the 'red tide'-causing dinoflagellate algae, as well as a variety of eukaryotic microbes of critical importance to human health (e.g., the malaria parasite *Plasmodium*). Despite its obvious significance, very little is known about the process of secondary endosymbiosis and its impact on the molecular and cell biology of secondary plastid-containing algae. The goal of this proposal is to sequence the nuclear genomes of two microbial eukaryotes of pivotal evolutionary and cell biological significance, the cryptomonad *Guillardia theta* and the chlorarachniophyte *Bigelowiella natans*. These organisms are unique among secondary plastid-containing algae in that they still possess the nucleus (nucleomorph) and cytoplasm of their algal endosymbionts in a highly reduced and simplified form. Despite striking similarities in the size and structure of their nucleomorph genomes, the cryptomonads and chlorarachniophytes are the product of *independent* secondary endosymbiotic events involving different endosymbionts (red and green algae, respectively) and unrelated eukaryotic host cells. The limited coding capacity of cryptomonad and chlorarachniophyte nucleomorphs and plastids indicates that their nuclear genomes have been repositories for thousands of endosymbiont-derived genes throughout their evolutionary history: comparing and contrasting these sequences will provide an unprecedented window into the process of secondary endosymbiosis and the integration of their respective hosts and endosymbionts at the genetic, biochemical and cellular level. These genome sequences will also provide critical insight into the evolutionary origins of two of the six currently recognized 'supergroups' of eukaryotes, the Chromalveolates and Rhizaria, to which the cryptomonads and chlorarachniophytes belong.

Whole Genome Analysis of Functional Protein Binding Sites and DNA Methylation: Application to p53 and Low Dose Ionizing Radiation

Krassimira Botcheva, Sean McCorkle, John J. Dunn (jdunn@bnl.gov), and **Carl W. Anderson*** (cwa@bnl.gov)

Biology Department, Brookhaven National Laboratory, Upton, NY

The effects of exposure to low doses of ionizing radiation on humans results largely from changes in gene expression mediated by activation of sequence-specific DNA binding proteins (transcription factors), changes to other chromosomal proteins and epigenetic modifications to the DNA. To develop a molecular understanding of the consequences of exposures to low doses of ionizing radiation, it will be necessary to understand where radiation-activated transcription factors bind in whole genomes and how radiation induces changes in factor binding, chromosome structure and DNA methylation occur. We developed a method, Paired-End Serial Analysis of Chromatin Occupancy (PE-SACO), for genome-wide profiling of functional protein binding sites in chromatin. PE-SACO combines chromatin immunoprecipitation with paired-end Genome Signature Tags, a technique developed at BNL that associates 20/21 bp DNA sequences (tags) with chromosomal DNA loci to identify and quantify functional protein binding sites in whole genomes. For example, p53 is a transcription factor that is activated in response to DNA double-strand breaks. p53 regulates ~1500 genes, directly or indirectly, but only some of its response elements have been identified and characterized. Furthermore, p53-mediated activation or repression of transcription depends on interactions with other factors including DNA methyltransferases, histone acetyltransferases (HATs) and deacetylases (HDACs). We currently are using PE-SACO in combination with several highly sensitive methylated-CpG island recovery assays to identify occupied p53 binding sites following low dose irradiation and other genotoxic stresses and to simultaneously investigate their methylation status.

As an initial step towards characterizing the role of p53 in response to ionizing radiation, we analyzed cloned PE-SACO ditag bar-coded libraries from human IMR-90 fibroblasts exposed to zero, 0.1 Gy or 8 Gy of ionizing radiation which were then sequenced at DOE's Joint Genome Institute employing 454 sequencing technology. Bar-coding different libraries with short DNA sequences allowed multiplexing of libraries during 454 sequencing which significantly increases efficiency, decreases costs, and improves bioinformatics analyses. Since the ditags DNAs have fixed lengths, parsing of individual tag sequences is straightforward. In addition, positioning the resulting ditags on the genome with inter-tag distances that match the size distribution of the ChIP DNA fragments increases overall mapping accuracy.

Our initial 454 sequencing experiments uncovered several areas for improvements including reducing amplification bias during library construction, improving the quality of the cloning vector and generating libraries with high complexity. To address these critical parameters, we have now developed a totally in vitro method for preparing PE-SACO ditag libraries and have added additional improvements to our PE-SACO technology that should bypass several rate limiting steps in the original protocol. We are now test-sequencing these libraries with the JGI. In future studies, we will extend our chromatin characterization to include changes in DNA methylation after exposure to ionizing radiation, its potential role in mediating bystander effects, and we will show how integrating these analyses provides a systems approach to address other issues relevant to DOE's missions in energy security, carbon management and bioremediation.

Funded by the DOE's Office of Biological and Environmental Research Low Dose Research Program.

A Comparison of Phylogenetic Microarray and SSU Clone Library Data

Eoin L. Brodie* (elbrodie@lbl.gov), Todd Z. DeSantis, Gary L. Andersen, Yvette M. Piceno, Jim Bristow, Jeanine Wiener-Kronish, Susan Lynch, Judith Flanagan, and Li Weng

Lawrence Berkeley National Laboratory, Berkeley, CA

Gene Transfer Through Transformation and Conjugation in the *Roseobacter* Lineage

Ina Buchholz[1], Silke Pradella,[2] and **Irene Wagner-Döbler**[1]* (iwd@helmholtz-hzi.de)

[1]Helmholtz-Center for Infection Research, Braunschweig, Germany and [2]German Collection for Microorganisms and Cell Cultures (DSMZ), Braunschweig, Germany

The *Roseobacter* group is an abundant lineage of marine *Alphaproteobacteria* comprising up to 25 % of all bacterial cells in the sea. Based on the genome sequences of representative strains, functional genomics investigations will be carried out to study the adaptations of the organisms to their ecological niches in the ocean (http://rosy.tu-bs.de/).

Up to 7 linear plasmids have been found in phototrophic *Roseobacter* strains[1], which account for a significant fraction of their genome, but natural gene transfer has until now not been demonstrated, nor are vectors for genetic engineering available in this group. *Staleya guttiformis* and *Roseobacter litoralis* were previously shown to harbour linear plasmids of 63 kb and 120 kb length, respectively, which carry the *pufLM* genes of the photosynthesis operon[2]. To investigate natural conjugation of these plasmids we determined the antibiotic resistance profiles of the donor strains and that of closely related potential recipients on half strength marine broth. *S. guttiformis* tolerates only tetracycline, and *R. litoralis* is resistant against streptomycin, clindamycin, kanamycin, gentamycin, and tetracycline (concentrations for all antibiotics 10 µg/ml). As recipients, *Silicibacter pomeroyi*, *Oceanibulbuls indolifex*, and *Phaobacter inhibens* T5 with complementary antibiotic resistance were chosen. Because of the size of the donor plasmids, mating mixtures were incubated at least three days, resulting in possible transconjugants as judged by PCR detection of the *pufLM* genes on media selective for the recipient. – We adapted a protocol for transformation through electroporation from *Rhodobacter sphaeroides* and were able to successfully introduce the broad host range plasmid RSF1010 (size 8.9 kb) into *Oceanibulbus indolifex* and *Staleya guttiformis* LM9 at frequencies between 2.3 x 10^5 and 6.4 x 10^7 per µg of DNA. These data show that both transformation and conjugation can occur in the *Roseobacter* group and provide a starting point for the construction of vectors for genetic experiments.

References

1. Biebl, H. *et al. Dinoroseobacter shibae* gen. nov., sp. nov., a new aerobic phototrophic bacterium isolated from dinoflagellates. *Int. J. Syst. Evol. Microbiol.* **55**, 1089-1096 (2005).

2. Pradella, S. *et al.* Genome organization and localization of the *pufLM* genes of the photosynthesis reaction center in phylogenetically diverse marine Alphaproteobacteria. *Appl. Environ. Microbiol.* **70**, 3360-3369 (2004).

Characteristics of SHyP-CGH DNA Microarray Signatures in Known Genomes

Charles H. Cannon[1]* (chuck.cannon@ttu.edu) and Chai-Shian Kua[2]

[1]Texas Tech University, Department of Biological Sciences, Lubbock, TX and [2]Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Mengla, Yunna, China

We examine the unique DNA sequence characteristics of the genomic signatures discovered through DNA microarray-based Short Hyperdispersed Probe (SHyP) comparative genomic hybridizations (CGH) of human, mouse, rat and ramin (*Gonystylus bancanus*) genomes. These DNA sequences are characterized using three techniques: 1) converting the probe sequences into multidimensional feature vectors; 2) a 'motif' frequency and abundance analysis; and 3) mapping them onto the entire NCBI's chromosome construct database using nBLAST. The feature vector analysis revealed distinct patterns among the different groups of genomic-indicator probes, using a principal component's analysis. Each sequence feature examined, from mono- to tri-nucleotides, provided different types of information. The mouse genome encompassed the largest signature, covering the range of DNA sequence variation in a manner similar to a random set of non-hybridizing probes. The rat and human indicator probes were well-separated in the DNA vector space, no matter what type of features were used. The higher level phylogenetic clades, rodent and mammal, both contained distinctive DNA sequence characteristics, dominated by repetitive 'motifs'. Similar but less repetitive motifs formed the core of the many of the genome indicator probes. A sizable fraction also contained no recognizable motif and the maximal complimentarity among sequences is not significantly greater within rather than among groups of probes. These results indicate that the SHyP DNA microarray approach does capture unique aspects of individual genomes and the larger clades to which they belong. While these genomic signatures do contain a subset of repetitive elements, the overall complexity of the DNA sequences characteristic for any genome is not substantially lower than those presented on the entire array. Our main objective was to study genomic diversity at a higher taxonomic level than has generally been attempted. The genomic signatures captured will not only provide unique DNA sequence identifiers but will also provide information about its relationship to other species. Ultimately, the data will be placed into an open access internet resource, allowing future virtual genomic comparisons between any two organisms already present in the database. These comparisons could quickly provide hundreds of independently inherited loci to examine at subspecific and population levels among any two species. It will also provide much wider access to genome level data to biologists working in a number of different fields.

Promoting Genomics in Tropical Botanical Research

Charles H. Cannon[1] and **Chai-Shian Kua**[2]* (kuacs@yahoo.com)

[1]Texas Tech University, Department of Biological Sciences, Lubbock, TX and
[2]Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Mengla, Yunnan, China

The availability of genomic sequence data has revolutionized the study of a wide range of model organisms. With the rapid development of sequencing technology and the 'excess' capacity of some major sequencing centers, expanding the genomic approach beyond the small set of model organisms is rapidly approaching. The availability of such data, particularly for economically and ecologically important groups such as *Shorea* (Dipterocarpaceae) or the Anacardiaceae, could facilitate a wide range of activities, from the conservation of natural diversity to the development of more effective and productive plantations. While the costs and difficulties of obtaining genomic level data are going down, it is still a major undertaking and requires a community of scientists to develop and justify such plans. Recently, at the first meeting of the Asian chapter of the Association for Tropical Biology and Conservation (ATBC) in Mahabalipuram, India, we presented an initiative to tropical botanists working in the region to form a consortium to move towards the goal of obtaining genomic sequence for a few target species. This community of scientists has started discussions about the choice of species, the structure and sources of funding, the actual processing of samples, and the management and sharing of data. We are in the initial phase of developing proposals to numerous funding organisations and including opportunities like the the U.S. DOE-Joint Genome Institute's Community Sequencing Program (http://www.jgi.doe.gov/CSP/overview.html). The introduction of genomic level perspectives to biologists working and living in these tropical countries would have a tremendous capacity building effect and allow more ready exchange between scientists working on model organisms and those working with less-than-model organisms, like tropical trees. The development of genomic expertise and science among Asian tropical biologists could provide a major step forward across a wide range of applied, basic, and conservation sciences. We also combine this initiative with an ongoing DNA microarray based technique for capturing genomic signatures, potentially providing a shortcut into other organisms, which will not have their genomes sequenced any time in the near future.

Genomic Analysis of the Evolution of Antifreeze Glycoprotein Genes in Antarctic Notothenioid Fish

C.-H.C. Cheng[1]* (c-cheng@uiuc.edu), J. Nicodemus,[1] S. Silic,[1] L. Ghigliotti,[2] and E. Pisano[2]

[1]Department of Animal Biology, University of Illinois, Urbana-Champaign and
[2]Department of Biology, University of Genova, Italy

Antarctic notothenioid fishes are an adaptive radiation that dominates the fish fauna in the frigid Southern Ocean. They owe their ecological success to a key evolutionary innovation driven by the Cenozoic cooling - antifreeze glycoproteins (AFGPs) which prevent organismal freezing by binding to internalized ice crystals from ambient water and arresting their growth. AFGPs occur as a family of size isoforms consisting of tandem (ThrAlaAla-) repeats, and are encoded in a polyprotein structure by a large multigene family. We have previously deciphered the evolutionary ancestor of AFGP gene to be a

trypsinogen-like protease (TLP) gene. The repetitive AFGP coding sequence arose from iterative duplications of a 9-nt (ThrAlaAla-) coding element within the ancestral TLP gene, and the new AFGP gene was formed when most of the TLP gene was shed. We also found a chimeric TLP-AFGP gene which may be an evolutionary transitional form. We began characterizing the genomic region containing the AFGP gene family in the giant Antarctic toothfish *Dissostichus mawsoni* in hope of reconstructing the full TLP-to-AFGP transformation process, and also determining the molecular mechanism of the evolutionary expansion of the AFGP polyprotein gene family.

We constructed a BAC library for *D. mawsoni*, and FPC analysis of AFGP/TLP-positive BAC clones and *FISH* of toothfish chromosomes localized AFGP/TLP genes to one genomic region. Shotgun sequencing and sequence assembly thus far of six BAC clones of the minimal tiling path indicate the AFGP/TLP locus to exceed 0.7 Mbp, potentially larger. The locus is gene-rich, with a current total of >60 genes, encompassing five types - AFGP, TLP, chimeric AFGP/TLP, and two distinct trypsinogens. Significant segmental duplications had occurred in this locus, generating tandem arrays of these five types of genes in discernible recurring patterns. In conjunction with data search of other vertebrate genomes, an evolutionary scenario of TLP from a member of the trypsinogen family in Teleostei, and notothenioid AFGP from TLP can now be reconstructed.

The adaptive AFGP trait of notothenioid fish removes the constant physical threat of freezing death in the icy, frigid Antarctic environment, but the entire suit of biochemical and physiological processes must also be cold-adapted for proper metabolic and housekeeping functions. The toothfish BAC library and genomic analyses of its AFGP/TLP locus would provide part of the foundation for whole genome sequence acquisition for the larger goal of understanding adaptation to the extreme cold for a vertebrate ectotherm.

Genomic Sequencing of the Chestnut Blight Fungus *Cryphonectria parasitica*

Alice C. L. Churchill[2] (acc7@cornell.edu), Michael G. Milgroom[2] and **Donald L. Nuss**[1]* (nuss@umbi.umd.edu)

[1]University of Maryland Biotechnology Institute, Rockville, MD and [2]Cornell University, Ithaca, NY

Research on the chestnut blight fungus *Cryphonectria parasitica* has provided important insights into mycovirus-host interactions, virus-mediated biological control of fungal diseases (hypovirulence), fungal population genetics, mechanisms underlying fungal pathogenesis and fungal signal transduction pathways. Very recent advances with the *C. parasitica* experimental system are providing important new insights into the role of RNA silencing as an antiviral defense mechanism in fungi and the impact of viruses on fungal vegetative incompatibility systems and secondary metabolism. The availability of the *C. parasitica* genome sequence, the first for an Ascomycete tree pathogen, will greatly accelerate the efforts of an active and growing research community that address a broad range of important fundamental and applied research projects.

Genome-Wide Sequence and Functional Analysis of Early Replicating DNA in Normal Human Fibroblasts

Stephanie M. Cohen[1], Terrence S. Furey[2], **Norman A. Doggett**[3]* (Doggett@lanl.gov), and David G. Kaufman[1]

[1]Department of Pathology and Laboratory Medicine, University of North Carolina, Chapel Hill, NC; [2]Institute for Genome Sciences and Policy, Duke University, Durham, NC; and [3]Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM

The replication of mammalian genomic DNA during the S phase is a highly coordinated process that takes place in a programmed manner. Recent studies have begun to elucidate the pattern of replication timing on a genomic scale. Using a combination of experimental and computational techniques, we identified a genome-wide set of the earliest replicating sequences. This was accomplished by first creating a cosmid library containing DNA enriched in sequences that replicate early in the S phase of normal human fibroblasts. Clone ends were then sequenced and aligned to the human genome. By clustering adjacent or overlapping early replicating clones, we identified 1759 “islands” averaging 100 kb in length, allowing us to perform the most detailed analysis to date of DNA characteristics and genes contained within early replicating DNA. Islands are enriched in open chromatin, G-negative chromosomal bands, transcription related elements, and Alu repetitive elements, with an under-representation of LINE elements. In addition, we found a paucity of LTR retroposons, DNA transposon sequences, and an enrichment in all classes of tandem repeats, except for dinucleotides. An analysis of genes associated with islands revealed that nearly half of all genes in the *WNT* family, and a number of genes in the base excision repair pathway, including four of ten DNA glycosylases, were associated with island sequences. Also, we found an over-representation of members of apoptosis-associated genes in very early replicating sequences from both fibroblast and lymphoblastoid cells, suggesting that genes involved in apoptosis represent a subset of genes that have been selected to replicate very early in the S phase. All of these data suggest that the chromatin structure/gene composition of this compartment of the S phase is conserved and may be important for the maintenance of cell stability and if defective may confer genetic instability.

Using Co-Regulation to Understand Low-Dose Ionizing Radiation Responsive Genes and Pathways

Matthew A. Coleman[1]* (coleman16@llnl.gov), Feliza Bourguet[1] (bourguet1@llnl.gov), Anya Krefft,[1] Francesca Pearson,[1] and Leif E. Peterson[2]

[1]Biosciences, Lawrence Livermore National Laboratory, Livermore, CA and [2]Departments of Molecular and Human Genetics and Medicine, Baylor College of Medicine, Houston, TX

Genome-scale expression microarray data in conjunction with DNA sequence/pattern databases was used to identify and validate gene regulatory elements that may control and differentiate aspects of cellular responses in humans sensitive to ionizing radiation. Using expression data we identified over 500 radiation responsive genes (IR). We applied computational tools to identify genes that showed a pronounced pattern of expression that correlated with increasing doses of IR. This found a group of ~33 genes that were identified as potentially regulated by TP53. To further expand the TP53 regulatory network we compared sequence level elements and transcription factor binding modules

across the group of 33 genes. This identified thousands of TP53 elements across the human genome that were synergistic with other ionizing radiation responsive elements such as SP1 and CREB transcription factor binding sites. A total of three individual modules were shared between 5 of the 33 genes. These modules were used to predict other novel IR responsive genes. QPCR verified both known and predicted IR responsive genes in two human lymphoblastoid cell lines. Chromatin immuno-precipitation assays were used to validate TP53 binding for several of these proximal promoters. These novel elements and modules help define new IR responsive networks. This data also suggest that proximal promoter regulatory elements may act cooperatively with TRP53 to modulate the cells response at both high and low doses of IR. This information provides the basis for identifying susceptibility regulatory factors at the sequence level that may be involved in individual responses to IR exposures and important for understanding low dose IR effects.

This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48 with funding from the Laboratory Directed Research and Development and DOE Low Dose Radiation Research programs.

Differential Gene Expression During Thermal Stress in the Caribbean Corals *Acropora palmata* and *Montastraea faveolata*

Michael DeSalvo[1]* (mdesalvo@ucmerced.edu), Chris Voolstra[1], Shini Sunagawa[1], Collin Closek[1], Jodi Schwarz[2], Peter Brokstein[3], Mary Alice Coffroth[4], Alina Szmant[5], and Mónica Medina[1] (mmedina@ucmerced.edu)

[1]University of California, Merced; [2]Vassar University, Poughkeepsie, NY; [3]DOE Joint Genome Institute, Walnut Creek, CA; [4]State University of New York (SUNY), Buffalo; and [5]University of North Carolina, Wilmington

Thermal stress from abnormally high sea temperatures causes a syndrome in corals known as “bleaching”, in which the coral hosts lose their color due to either loss of their endosymbiotic dinoflagellates, degradation of photosynthetic pigments within the dinoflagellates, or both. Bleaching events have increased in frequency and intensity over the last decade, leading researchers to posit a connection with global warming. Understanding how corals respond to thermal stress, and their capacity for tolerance, resilience, and adaptation is crucial to realizing the fate of endangered coral reefs. Identifying causal mechanisms can strengthen conservation efforts and encourage protective legislation. Numerous molecular mechanisms underlying the bleaching stress response have been proposed, including: damage to photosystem II, disruption of the Calvin cycle, changes in protein phosphorylation patterns, tissue necrosis, apoptosis, and oxidative stress. We have undertaken a functional genomics approach to the study of coral bleaching. Generation of cDNA microarrays for *Acropora palmata* and *Montastraea faveolata* allows for rapid, high throughput assessment of differential gene expression. The *A. palmata* microarray contains 2,016 cDNAs printed in duplicate, and the *M. faveolata* microarray contains 1,536 cDNAs printed in duplicate. Bleaching experiments were conducted at the Bocas del Toro Field Station of the Smithsonian Tropical Research Institute in fall of 2006. Analyses of these samples reveal a vast number of differentially regulated genes along a time course of thermal stress, thus giving insight into the molecular processes involved in thermal stress-induced coral bleaching.

Bioaerosol Background Characterization in Public Facilities

Norman Doggett[1]* (Doggett@lanl.gov), Priya Dighe,[1] Megan Doyal,[1] T. Jason Gans,[1] Rob Leach,[1] Chris Stubben,[1] Jian Song,[1] Murray Wolinsky,[1] Alonso Castro,[2] and Kathy Creek[2]

[1]Bioscience Division and [2]Physics Division, Los Alamos National Laboratory, Los Alamos, NM

We are studying the bioaerosol background within indoor public facilities by collecting and analyzing airborne particulate and spectral sensing data. The goal of this project is to enhance and complement the development of detection and analysis protocols by providing essential information on the background microbial populations in public settings. We have conducted a detailed biological characterization of the bioaerosol samples collected during "peak" and "non-peak" spectral sensing intervals based on intrinsic fluorescence and elastic light scatter. Biological analysis involved 1) media extraction, 2) DNA quantitation, 3) 16S PCR amplification and 4) cloning and sequencing. A 24-well block protocol was developed to enable higher throughput extractions of bioaerosol filters. Extracted DNA is quantitated by quantitative real-time PCR. PCR with 16S rDNA primers provided amplified product for cloning and sequencing. 16S community sequence analysis of bioaerosol samples from public facilities within three different cities revealed different microbial background complexities between cities, however analysis of bioaerosol samples associated with "peak" and "non-peak" spectral sensing intervals within the same location did not indicate significant changes in biological complexity. These results indicate that spectral based biosensors may detect particulate noise in the environment that is not associated with alterations in biological complexity. Details of microbial community analysis will be presented.

Comparative Genomics of Chloroviruses

David D. Dunigan[1,5]* (ddunigan2@unl.edu), Garry A. Duncan[2] (gduncan@nebrwesleyan.edu), Lisa A. Fitzgerald,[3] Michael V. Graves,[4] James R. Gurnon,[1] and James L. Van Etten[1,5]

[1]Department of Plant Pathology, University of Nebraska, Lincoln, NE; [2]Biology Department, Nebraska Wesleyan University, Lincoln, NE; [3]National Institute of Standards and Technology, Gaithersburg, MD; [4]Department of Biological Sciences, University of Massachusetts, Lowell, MA; and [5]Nebraska Center for Virology, University of Nebraska, Lincoln, NE

The family Phycodnaviridae encompasses a diverse collection of large icosahedral, dsDNA viruses infecting algae with the coding capacity for hundreds of genes. These viruses have genomes ranging from 160 to 560 kb. The family consists of six genera based initially on host range and supported by sequence comparisons. While the family is monophyletic with branches for each genus, it has evolutionary roots that connect with several other families of large DNA viruses, referred to as the nucleocytoplasmic large DNA viruses (NCLDV). The genomes of members in three genera in the Phycodnaviridae have been compared (Dunigan et al. 2006, Virus Research 117, 119-132). The viruses have diverse genome structures, some with large regions of non-coding sequence and others with regions of single-stranded DNA. The genus *Chlorovirus* represents those viruses infecting freshwater green algae, including the endosymbiotic *Chlorella*-like green algae. We have sequenced and analyzed six of these viral genomes that infect three unique algal hosts, and represent a broad geographical distribution. Comparison of these viral genomes

indicate i) a conservation of size and structure, ii) a conserved motif of tRNA-encoding polycistron located near the middle of the genomes, iii) a high level of synteny within those viruses infecting the same host, but significant gene rearrangement when comparing viruses of differing hosts, iv) a large number of genes recognized to augment metabolic functions of the host cell, e. g., nucleotide metabolism, v) the genomes appear to be mosaics of genes with both prokaryotic and eukaryotic origins, vi) 60-65% of the ORFs have no homologs in the sequence databases, and vii) the G+C contents vary with the host, viii) genes of these viruses often encode the smallest proteins of their class. An example of this final point is the potassium ion channel of the virus ATCV-1, which is 82 amino acids and is fully functional when expressed in frog oocytes; this protein is just large enough to form a basic ion channel. Some phycodnaviruses are important regulators of the algal populations they infect and they are likely to be important forces in the evolution of their hosts. Not much is known about the biology of these viruses, but the consequences of these infections on the phytoplankton community can have global affects, including altered geochemical cycling and weather patterns.

Brachypodium distachyon EST Sequencing Project

Samuel E. Fox* (foxsa@onid.orst.edu) and Todd C. Mockler

Oregon State University, Corvallis, OR

Teaching Genome Biology: a Collaborative Effort between the JGI and UC Merced

M. Pilar Francino[1]* (mpfrancino@lbl.gov), Miriam Barlow,[2] and Monica Medina[2]

[1]DOE Joint Genome Institute, Walnut Creek, CA and [2]University of California, Merced

With the aim of providing cutting-edge theoretical and experimental training in Genome Biology, UC Merced and the JGI have initiated a joint course that involves faculty from both institutions. The course started this Spring semester 2007 and is mostly attended by upper division undergraduate students participating in the UCM Biological Sciences Major. More than 20 JGI researchers from all programs and departments are participating by lecturing or organizing laboratory sessions on their topics of expertise at the UC Merced campus. The main objective of the course is to instruct students in the procedures involved in a genome project, by developing several small projects that illustrate different aspects of genome science. The course includes hands-on training in both experimental procedures and bioinformatic techniques, following the protocols in use at the JGI and including DNA preparation, library construction, sequencing, genome assembly, gene prediction and annotation. In addition, the course describes the major areas of current research in the field of Genome Biology, including Comparative Genomics, Metagenomics, Transcriptomics and Proteomics. Students participating in this course will be qualified to participate in summer internships at the JGI and will have the required skills to seek employment in genomic centers after finishing their degrees.

Phylogenomics and Protein Interactivity: Generalizing the Complexity Hypothesis

M. Pilar Francino[1]* (mpfrancino@lbl.gov), Monica Medina[2], Jenna Morgan[1] and Paramvir Dehal[1,3]

[1]Evolutionary Genomics Department, DOE Joint Genome Institute, Walnut Creek, CA; [2]School of Natural Sciences, University of California, Merced; and [3]Lawrence Berkeley National Laboratory, Berkeley, CA

In spite of the large numbers of complete genomes now available, no consensus has yet emerged regarding deep phylogenetic relationships in many branches of the tree of life. Rather, the comprehensive analysis of whole genomes indicates that several biological phenomena create discrepancies between gene and organismal history, such that the phylogeny of one gene or group of genes can't necessarily be interpreted as the phylogeny of the species that carry them. This is not surprising given that gene and species lineages are related but distinct levels of selection. Many gene lineages evolve under complicated dynamics, that entail intraspecific processes of gene duplication, loss and recombination, plus, mainly in the prokaryotic world, events of horizontal transfer between species. As a consequence, the similarity among homologous gene copies in different species depends on many factors beyond the topology of the species tree. However, even if all genes and other DNA sequences can undergo change-inducing molecular events, natural selection at the organismal level will constrain differently the possible evolutionary paths of different gene lineages. Recent analyses suggest the existence of genes that evolve in a stable and conservative manner across many organisms, avoiding gene loss, duplication, horizontal transfer and acceleration of the nucleotide substitution rate. Further, we suggest that protein interactivity is the biological glue that allows these genes to resist mutational pressure towards different kinds of change. In the context of reconstructing the deep history of life, these genes should be the most reliable phylogenetic markers. Here, we define these genes in two deep clades of the tree of life–Bacteria and Opisthokonta–and use them to resolve phylogenetic relationships within them.

Effect of Host Tree Species on Gut Microbial Community of Asian Longhorned Beetle (*Anoplophora glabripennis*)

Scott Geib[1]* (smg283@psu.edu), John Carlson,[2] Maria del Mar Jimenez-Gasco,[3] Ming Tien,[4] and Kelli Hoover[1] (kxh25@psu.edu)

[1]Department of Entomology, [2]School of Forest Resources, [3]Department of Plant Pathology, and [4]Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA

The Asian longhorned beetle is an invasive species introduced into the U.S. from China. It has a broad host range and has resulted in the loss of thousands of hardwood trees, especially maples. Like other wood-feeding insects, ALB must acquire its nutrition by degrading lignocellulose to obtain nutrients and energy. Cellulolytic enzymes in the ALB gut may originate from symbionts, the insect itself, or some combination of the two. We surveyed for bacterial and fungal community composition from the gut of mid-instar larval ALB fed in different host tree species using culture independent community analysis. PCR amplification and cloning of the total gut DNA for the 16S rDNA region for bacteria and translation elongation factor-1α region for fungi allowed us to screen for members of the microbial community that differed in insects fed in different host tree species. In larvae fed

on sugar maple or pin oak, the gut was found to harbor a rich variety of bacterial and fungal species, including several unique genera (e.g., *Cellulosimicrobium* sp. and *Fusarium* sp.) known to play a role in wood decay. Also, the complete suite of cellulolytic enzymes (endo- and exo-glucanases as well as beta-glucosidases) were detected in beetle guts using specific substrates, including an assay that provides direct evidence of lignin degradation. We plan to investigate further how the gut community adapts through the insect's life history and as a function of host tree to understand the source and evolutionary relationships of gut microbial symbionts in this important invasive species.

JGI Sequencing Projects: Statistics and Timelines

Tijana Glavina del Rio[1]* (glavinadelrio1@llnl.gov), Kerrie Barry,[1] Lynne Goodwin,[2] Miranda Harmon-Smith,[1] Susan Lucas,[1] and David Bruce[2]

[1]DOE Joint Genome Institute, Walnut Creek, CA and [2]Los Alamos National Laboratory, Los Alamos, NM

The Department of Energy's (DOE) Joint Genome Institute (JGI) is one of the major publicly funded high throughput sequencing centers. The current capacity of the Production Genomics Facility (PGF) in Walnut Creek, California is approximately three billion bases per month, generating a total of 55 million lanes this year. JGI sequencing projects are initiated through one of three peer reviewed programs: Community Sequencing Program (CSP), DOE Microbial Program and the Laboratory Science Program (LSP). This poster will present an overview of project statistics for 2006 and current projects for 2007. In 2006, the JGI processed a collection of DOE mission relevant sequencing projects ranging from prokaryotes to eukaryotes as well as several microbial communities. The poster will also describe how projects are scheduled for production sequencing and display tools used for tracking projects to their completion. Project timeline from initiation to completion will also be presented.

This work was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. DE-AC52-06NA25396.
UCRL-ABS-228496
LBNL-62524 Abs.

Bias Free Linear Vector for Cloning Recalcitrant DNA and Accelerating Sequence Finishing

Ronald Godiska,[1] Rebecaa Hochstein,[1] Sarah Vande Zande,[1] Nikolai Ravin,[2] Attila Karsi,[3] and **David A. Mead**[1]* (dmead@lucigen.com)

[1]Lucigen Corporation, Middleton, WI; [2]Centre Bioengineering, Russian Academy of Science, Moscow, Russia; and [3]Mississippi State University, Mississippi State, MS

We have developed a novel linear vector for unbiased cloning of 10-30 kb inserts in *E. coli*. This vector, termed "pJAZZ", shows unprecedented ability to maintain large inserts from very AT-rich genomes. The otherwise difficult to clone genome from *Flavobacterium columnare* (70% AT) was sequenced to seven fold coverage using the pJAZZ vector, resulting in less than 10 clone gaps left in this 3.2 Mb genome. The linear vector was able to maintain 20-30 kb fragments from *Lactobacillus helveticus* (65% AT) and 2-4 kb inserts from Piromyces species (up to 96% AT), which were unclonable in

conventional plasmids. Unlike fosmid cloning, the construction of large-insert libraries (10-20 kb) in pJAZZ is simple and robust, using standard methods of transformation and plasmid purification. We are evaluating the use of a single pJAZZ shotgun library to eliminate the need for multiple libraries, making finishing easier and more cost effective. Enhanced stability of inserts in the pJAZZ vector is attributed to both the lack of supercoiling and the lack of transcriptional interference. Torsional strain inherent to supercoiled plasmids can induce localized melting and generate secondary structures, which are substrates for deletion or rearrangement by resolvases and replication enzymes. For example, the instability of tandem repeats and palindromic sequences is presumably due to cleavage of hairpin structures or to replication slippage across the secondary structures. Most conventional plasmid vectors also induce strong transcription and translation of inserted fragments, and they allow transcription from cloned promoters to interfere with plasmid stability. As a result, certain DNA sequences are deleterious or highly unstable, leading to sequence "stacking", clone gaps, or a complete inability to construct libraries, especially from AT-rich genomes or toxic cDNAs. The transcription-free, linear pJAZZ vector also minimizes "sequence gaps" caused by secondary structures, as shown by its stable cloning of inverted repeats and di- and tri-nucleotide repeats.

In-silico Methods for Troubleshooting Genomic Shotgun Data

Eugene Goltsman* (egoltsman@lbl.gov), Vasanth Singan, Stephan Trong, Alex Copeland, and Alla Lapidus

DOE Joint Genome Institute, Walnut Creek, CA

End-sequencing of shotgun libraries of small genomic inserts is today the most popular approach to Whole Genome Sequencing (WGS). Irregularities in WGS datasets present assembly problems that are expensive and time-consuming to solve, with cloning bias, contamination and long repeats posing the biggest challenges. Shotgun assembly data exhibit well recognizable patterns that follow certain statistical models, and deviations from these models usually stem from flaws and anomalies in the input data, which in turn reflect problems in the cloning protocol, chemistries, or the DNA being sequenced. We developed several statistical and bioinformatic methods for detecting cloning bias, DNA contamination and high repeat content at early stages of the WGS project. These methods are based on analyses of i) depth of coverage distributions, ii) dynamics of iterarative assemblies and iii) GC profiles of real and simulated shotgun datasets. We identify and describe relationships between read coverage and the Poisson function and demonstrate ways to routinely identify cloning bias and contamination through these relationships. Identifying abnormal patterns in the dataset's GC profile at various levels (s.a. genome, library, plate) provided a convenient method for catching suspected contamination. Routine automated application is also discussed.

This work was performed under the auspices of the U.S. DOE of Science, Biological and Environmental Research Program, and by the University of California, LLNL under Contract No. W-7405-Eng-48, LBNL under Contract No. DE-AC02-05CH11231 and LANL under Contract No. W-7405-ENG-36.

Project Management at the Joint Genome Institute

Lynne Goodwin*, David Bruce, Kerrie Barry, Tijana Glavina del Rio, and Susannah Tringe

Joint Genome Institute, Los Alamos National Laboratory, Los Alamos, NM

As high throughput sequencing centers move from managing a small number of large projects to managing many simultaneous small projects, the ability to govern schedule, cost, quality, and project specification becomes more difficult. The JGI has implemented a formal project management system that simplifies controlling multiple small projects.

The Department of Energy Joint Genome Institute (JGI) high throughput sequencing and computational analysis group consists of teams at Oak Ridge National Laboratory, DOE Production Genomics Facility at Walnut Creek, Ca, Los Alamos National Laboratory, Lawrence Livermore National Laboratories and the Stanford University, SHGC. Historically the JGI had a small number of large projects. Today, the JGI sequencing capacity is dedicated to many small projects (< 10 MB) such as microbial, both small eubacteria and eukaryotic genomic, environmental metagenomic, and large eukaryotic projects. The JGI Project Management Office is the primary point of contact for JGI sequencing project information for both the internal and external stakeholders.

Who Forgot to Sequence *Micrococcus luteus*

Charles Greenblatt[1]* (greenbl@cc.huji.ac.il), Michael Young,[2] and Arseny Kaprelyants[3]

[1]Hebrew University, Jerusalem, Israel; [2]Institute of Biological Sciences, The University of Wales, Aberystwyth; and [3]Bakh Institute of Biochemistry, Russian Academy of Sciences, Moscow, Russia

Micrococcus luteus is as "historically" important as any bacteria. It was Fleming's pre-penicillin subject, and since lysozyme readily removes its peptidoglycan envelope, was the workhorse of protoplast studies. It has been a major focus for isoprene research – and from it isolation of important genes in the pathway have been undertaken. It is at a critical point in the evolution of the high G+C bacteria, being simpler in its genomic structure than many *Streptomyces*. Of interest to the D.O.E. is its ability to use many aromatic compounds as carbon sources and its tolerance to metals –important components of toxic wastes. *Micrococcus* spp. have been found to degrade hydrocarbons and olefinic compounds, with soil isolates of the genus capable of using biphenyl as a carbon source. An entire enzyme system is in place for phthalate degradation. *M. luteus* is known to harbor a 2.3 Kbp plasmid, pMLU1 with genes encoding enzymes for degradation of malathione and chlorpyriphos

We have found a number of *M. luteus* isolates in amber, an especially harsh and oligotrophic environment. This finding along with a number of other non-sporulating high G+C bacteria in amber focused our attention on their mechanisms of survival and dormancy. In an interesting confirmation. We became even more intrigued when we learned of resuscitation promoting factor (Rpf). Rpfs are a family of growth factors found among the high G+C bacteria which can resuscitate bacteria from a dormant state. The factor from M. luteus raises colony counts of other high G+C organisms many fold. It has important implications for detection and culturing of these organisms, most of which contain multiple gene homologues. The group includes a number of human pathogens. In *M. luteus* there is only a single rpf-like gene, which is an essential gene. The Rpf complex

is flanked by a proteins homologous to the IS1557 transposase of *Mycobacterium tuberculosis* and ArgR homologous to that from streptomycetes, corynebacteria and mycobacteria. *M. luteus* is clearly the organism of choice for further work on this novel intercellular signalling pathway.

We have joined forces with Janet Jansson's group in Sweden, which has characterized *Arthrobacter chlorophenolicus,* an organism capable of living on para-substituted phenols or phenol, as a single carbon source. *A. chlorophenolicus* has an unusual substrate specificity for para-substituted phenols (chlorophenol, bromophenol and nitrophenol), or phenol, and is able to degrade them via a novel catabolic pathway

In combination with the metal tolerance of *M. luteus* this seemed like a powerful duo for bio-remediation. The JGI has accepted the proposal to sequence both organisms, which we believe will shed light on the pathways for

i) catabolism of aromatics

ii) adaptation to high concentrations of toxic compounds and metals

iii) growth in changing temperatures

iv) evolution of these processes

The sequencing of *M. luteus* will lead to understanding metal binding (gold and strontium) and hydrocarbon degrading genes linked to methane oxidation. Insight will be gained into how bacteria shift between dormant and active states, and the adaptation of bacterial physiology to the soil environment.

Genome Sequencing and Annotation of the Acidophilic Metal-Reducing Bacterium *Acidiphilium cryptum JF-5*

Tina Gresham,[1] Mitch Day,[1] Loubin Yang,[1] Michael A. Thomas,[1] Peter P. Sheridan,[1] David Sims,[2] Paul Richardson,[2] and **Timothy S. Magnuson**[1]***** (magntimo@isu.edu)

[1]Idaho State University, Pocatello, Idaho and [2]DOE Joint Genome Institute, Walnut Creek, CA

Some metal-contaminated environments are under acid conditions, nonetheless, acidophilic iron respiring bacteria have been detected in such environments, and a potential means for microbial remediation of metals under acid conditions is feasible. The genome of the strict acidophile *Acidiphilium cryptum* JF-5 was sequenced by JGI, and a variety of interesting and significant findings were made. The main chromosome of *A. cryptum* was about 3.9 Mb in size, containing 4005 ORFs. Surprisingly, *A. cryptum* also possessed an 'extragenome' of 9 plasmids ranging in size from 1 kb to 200 kb. The genome sequence of *A. cryptum* was processed at TIGR using the Annotation Engine service. This service included automated annotation through the pipeline hosted by TIGR and the use of the Manatee manual annotation software. Manual annotation was accomplished by a group of over 30 biologists in a focused three-day period, resulting in over 3500 genes annotated. Plasmid-encoded mobility and replication related sequences were discovered, as well as sequences corresponding to mercury resistance and cation transport. On the main chromosome, genes were found encoding redox proteins thought to be important in Fe respiration. These genes included 12 genes encoding both membrane-associated and periplasmic cytochromes c. No evidence of cytochromes c containing more than 3 hemes were found. Significant differences in cytochrome c gene numbers and

predicted protein structures from *Shewanella* and *Geobacter* were noted, suggesting different mechanisms for respiratory Fe-reduction. Genes for metal transformation were discovered, including chromate reductase and arsenate reductase. Biofilm-related genes, encoding capsular exopolysaccharide synthesis and pili/flagella assembly, were present. Several genes encoding NiFe-hydrogenase were found, as well as a modified TCA cycle (glyoxylate bypass) for conservation of carbon. Carbohydrate metabolism genes were consistent with this organism's ability to grow on glucose, an unusual feature of some Fe-respiring bacteria. *A. cryptum* is thus substantially different genetically and biochemically than its neutrophilic counterparts, but is well-suited for both aerobic and anaerobic lifestyles in environments where toxic metals are prevalent.

Annotation of Eukaryotic Genomes

Igor Grigoriev* (ivgrigoriev@lbl.gov), Andrea Aerts, Alan Kuo, Robert Otillar, Asaf Salamov, Astrid Terry, and Kemin Zhou

DOE Joint Genome Institute, Walnut Creek, CA

Comparative Genomic Analysis in Aquificales

N. Hamamura[1]* (nhama@pdx.edu), E. Griffiths,[2] M. Podar,[1] K. Schweighofer,[3] R. Seshari,[4] J. Heidelberg,[5] J. Meneghin,[1] and A.-L. Reysenbach[1]

[1] Department of Biology, Portland State University, Portland, Oregon; [2]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario, Canada; [3]NASA Center for Astrobioinformatics, Moffett Field, California; [4]The J. Craig Venter Institute, Rockville, Maryland; and [5]Department of Biological Sciences, University of Southern California, Los Angeles, California

The Aquificales are a group of microaerophilic chemolithoautotrophic and heterotrophic thermophiles that are found widely distributed in both marine and terrestrial hot spring environments, where they are often the dominant visible filamentous biomass in thermal streams and play a key role in biomineralization reactions. Here we report comparative analysis of two closed genome sequences and two draft sequences of Aquificales, *Persephonella marina*, *Sulfurihydrogenibium azorense*, *S. yellowstonense* from Hydrogenothermaceae, and *Hydrogenivirga* strain 128-5-R1-1 from Aquificaceae. The closed *P. marina* genome consists of a 1,930,284 bp circular chromosome encoding for about 1987 ORFs whereas the 1,640,899 bp chromosome of *S. azorense* encodes for 1782 ORFs. A single extrachromosomal element was found in *P. marina*, which contains 53 682 bp and encodes 74 open reading frames. One notable difference amongst the genomes was the relative abundance or absence of transposases. *S. azorense* contains about 38 transposases. Likewise, 27 transposases were detected in the draft *S. yellowstonense* genome. Unlike their hyperthermophilic relative, *Aquifex aeolicus*, the closed genomes of *P. marina* and *S. azorense* revealed the presence of typical cassette of chemotaxis genes. All Aquificales fix CO_2 via the reductive TCA cycle, and are able oxidize sulfur. Although until recently all Aquifcales were thought to oxidize hydrogen with oxygen, *S. yellowstonense* is unable to use hydrogen, and this is reflected in the hydrogenase repertoires of the different genomes. Many features of the genomes confirm the reported metabolic properties and provide clues to potential novel metal reducing pathways.

Unprecedented Bacterial Diversity in a Hypersaline Microbial Mat

J. Kirk Harris[1]* (jjharris@colorado.edu), Jeffrey Walker,[1] Julita Madejska,[2] Falk Warnecke,[2] Victor Kunin,[2] Karla Ikeda,[2] Kerrie Barry,[2] John Spear,[3] Phil Hugenholtz,[2] and Norman Pace[1]

[1]University of Colorado, Boulder, CO; [2]DOE Joint Genome Institute, Walnut Creek, CA; and [3]Colorado School of Mines, Golden, CO

Microbial mats have been considered simple model systems for studying microbial ecology and biogeochemistry. Some of the most extensive mat communities known inhabit ~200 square kilometers of evaporative ponds in North America's largest salt works, the Exportadora de Sal in Guerrero Negro, Baja California Sur, Mexico. These mats have been studied extensively for decades, and were considered simple microbial ecosystems based on microscopic observations and cultivation studies. However, recent cultivation independent, DNA sequence-based analyses showed that the Guerrero Negro mats are among the most diverse microbial ecosystems so far studied. These results, based on a survey of ~1600 ribosomal RNA gene sequences, suggested this environment harbored extensive additional diversity. Thus, we proposed a large-scale ribosomal RNA analysis combined with a metagenomic analysis to further characterize this diversity of this unique environment.

Approximately 120,000 additional ribosomal RNA sequences and ~100 Mbp of metagenomic sequences were determined as part of this CSP project. Resulting ribosomal RNA sequences are representative of ~75 bacterial divisions, up from 45 divisions observed in the original study. This includes ~30 new bacterial divisions (15 in original study). Comparison of the ribosomal RNA and metagenomic sequence data shows a similar phylogenetic distribution. This study provides new perspective on bacterial phylogenetic diversity in general.

Insights into Stress Ecology and Evolution of Microbial Communities from Uranium-Contaminated Groundwater Revealed by Metagenomics Analyses

Christopher L. Hemme[1]* (hemmecl@ou.edu), Ye Deng,[1] Terry Gentry,[2] Liyou Wu,[1] Matthew W. Fields,[3] Chris Detter,[4] Kerrie Barry,[4] Miriam Land,[5] Nikos Krypides,[4] David Watson,[5] Paul Richardson,[3] James Bristow,[3] Terry Hazen,[6] James Tiedje,[7] Eddy Rubin,[3] and Jizhong Zhou[1] (jzhou@ou.edu)

[1]Institute for Environmental Genomics, Department of Botany and Microbiology, University of Oklahoma, Norman, OK; [2]Department of Soil Sciences, Texas A&M University, College Station, TX; [3]Department of Microbiology, Montana State University, Bozeman, MT; [4]DOE Joint Genome Institute, Walnut Creek, CA; [5]Biological Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [6]Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA; and [7]Center for Microbial Ecology, Michigan State University, East Lansing, MI

Due to the uncultivated status of the majority of microorganisms in nature, little is known about their genetic properties, biochemical functions, and metabolic characteristics. Although sequence determination of the microbial community 'genome' is now possible with high throughput sequencing technology, the complexity and magnitude of most microbial communities make meaningful data acquisition and interpretation difficult.

Therefore, we are sequencing groundwater microbial communities with manageable diversity and complexity (~10-400 phylotypes) at the U.S. Department of Energy's Natural and Accelerated Bioremediation Research (NABIR)-Field Research Center (FRC), Oak Ridge, TN. The microbial community has been sequenced from a groundwater sample (FW106) contaminated with very high levels of nitrate, uranium and other heavy metals and pH ~3.7. Consistent with trends expected in stressed ecosystems, the metagenome reveals a community of low species and strain diversity dominated by a single *Frateuria*-like γ-proteobacteria with other γ- and β-proteobacteria present at low proportions. Metabolic reconstruction reveals specific adaptations to the geochemical conditions of FW106 including genes encoding metal resistance (*czcABC*, *czcD*, *cadA*, *merA*, *arsB*), denitrification, and solvent resistance (1,2-dichloroethene, acetone, butanol). In addition to the presence of these specific genes, certain resistance genes also appear to be overrepresented in the metagenome including genes from nitrate/nitrite transport (*narK*) and metal translocation (*czcABC*, *czcD*, *cadA*), likely due to a combination of gene duplication and lateral gene transfer. A screen for positive selection shows most of these genes to be under strong negative selection, suggesting that in the short term at least, the overabundance of these transporters provide a positive fitness benefit to the cell by increasing the rate of ion transport. SNP analysis revealed a low level of polymorphism with the overwhelming majority of SNP representing unique changes within the assembled reads, suggesting that the strains in the sample are largely clonal. A model is presented for the evolution of microbial communities under high-stress conditions.

Genomics of Cellulosic Ethanol-Producing Bacteria

Christopher L. Hemme[1]* (hemmecl@ou.edu), Matthew W. Fields,[2] Qiang He,[3] Zhiguo Fang,[1] J. C. Detter,[4] Kerrie Barry,[4] Alla Lapidus,[4] Cliff S. Han,[5] Paul Richardson,[4] Terry Hazen,[7] Eddy Rubin,[5] and Jizhong Zhou[1] (jzhou@ou.edu)

[1]Institute for Environmental Genomics, Department of Botany and Microbiology, University of Oklahoma, Norman, OK; [2]Department of Microbiology, Montana State University, Bozeman, MT; [3]Department of Civil and Environmental Engineering, Temple University, Philadelphia, PA; [4]DOE Joint Genome Institute, Walnut Creek, CA; [5]Los Alamos National Laboratory, Los Alamos, NM; and [6]Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA

Recent global fluctuations in the supply and demand of petroleum have prompted a call for increased research into biologically-derived fuels (biofuels). Of particular interest are those processes involving production of biofuels from cellulosic biomass. To this end, the genomes of three strains of ethanol-producing bacteria (*Thermoanaerobacter ethanolicus* 39E, *T. ethanolicus* X514 and *Clostridium cellulolyticum*) have been sequenced. Strain 39E was isolated from a Yellowstone hot spring and is relatively well-characterized. Strain X514 is a metal-reducing bacteria isolated from the deep subsurface and is predicted to have been geographically isolated from 39E for ~250 MY. Metabolic reconstruction reveals insights into the carbon metabolism and niche adaptation of the two strains. Both strains are capable of metabolizing glucose and xylan to ethanol with a novel bifunctional secondary alcohol dehydrogenase serving as the terminal enzyme in the pathway. Slight differences are noted in the carbon metabolism of the two strains, including a complete KDPG metabolism pathway in 39E and the lack of a complete methylglyoxal shunt in X514. A survey of unique genes between the strains reveals lineage-specific gene expansions in the two strains including individual unique sugar transporter profiles and an increased number of P-type metal translocating ATPase genes in X514. In contrast to *Thermoanaerobacter*, *C. cellulolyticum* is capable of degrading a variety of cellulosic

materials including cellulose, xylan, pectin, mannan and chitin. *C. cellulolyticum* employs a large extracellular cellulosome complex to degrade these materials and comparisons with a previously sequenced genome of *C. thermocellum* suggests a significant diversity in cellulosome composition. To complement this research, a request for sequencing the genomes of an additional 20 ethanol-producing Clostridia strains has been approved by JGI. Strains were chosen from among the genera *Clostridium*, *Thermoanaerobacter*, *Thermoanaerobacterium* and *Acetivibrio* based on prior knowledge, phylogeny, unique physiology and industrial applications. The expansion of the genomic database of industrially-important Clostridia is expected to provide substantial benefits in the understanding of this class of organisms.

Diversity and Environmental Distribution of the Termite Group I Phylum, and Isolation of *Elusimicrobium minutum* gen. nov. sp. nov.

Daniel Herlemann* (herlemann@mpi-marburg.mpg.de), Oliver Geissinger, and Andreas Brune (brune@mpi-marburg.mpg.de)

Department of Biogeochemistry, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany

The bacterial candidate phylum Termite Group I (TG-1) presently consists mostly of "Endomicrobia", which have been characterized as endosymbionts of flagellate protists occurring exclusively in the hindgut of termites and wood-feeding cockroaches. Although the presence of TG-1 bacteria outside these insect guts is largely undocumented, public databases contain a growing number of distantly related 16S rRNA gene sequences obtained from other habitats. Using specifically designed primers based on the existing dataset, we explored the diversity and environmental distribution of the TG-1 phylum. Phylogenetic analysis revealed that the sequences obtained in this and previous studies form several deep-branching lineages. The "Endomicrobia" remain a separate lineage distinct from another, probably non-endosymbiotic, lineage that contains clones from the intestinal tracts of insects and the cow rumen. This lineage also includes *Elusimicrobium minutum* gen. nov. sp. nov., the first isolate from the TG-1 phylum, an anaerobic bacterium with a fermentative metabolism, colonizing the gut of scarab beetle larvae. Other deep-branching lineages consist of clones from various soils, sediments, and polluted aquifers. Although the abundance of TG-1 bacteria in most habitats other than termite and wood-feeding cockroach guts seems to be relatively low, the primer sets designed in this study will be useful to detect representatives of this phylum also in other environments. The genomes of *Elusimicrobium minutum* strain Pei191 and "*Candidatus* Endomicrobium trichonymphae", a representative of the endosymbiotic lineage (see poster by W. Ikeda-Ohtsubo and A. Brune) are presently being sequenced (CSP_777133 by A. Brune, J. Leadbetter, M. Friedrich, and P. Hugenholtz).

Evaluation of the Illumina Genome Analysis System

David W. Hillman*, Mary Ann Pedraza, Jean Zhiying Zhao, Sirisha Sunkara, Robert Egan, Feng Chen, and Paul Richardson

DOE Joint Genome Institute, Walnut Creek, CA

The Illumina 1G Genome Analysis System permits simultaneous sequencing of millions of DNA fragments using a reversible 4-color terminator based sequencing chemistry. The JGI Technology Development Group has been validating the process, equipment, and data output using a variety of microorganisms as test cases. Currently the system produces 2-3 million reads for 8 samples per run with readlengths from 26-36 bases

The process begins with gDNA fragmentation, ligation to system specific dsDNA adapters, size selection of the library, and an 18 cycle PCR amplification. Next libraries are hybridized to complementary primers on a specialized glass slide, the "Flowcell", using the fluidics of the "Cluster Station". Using isothermal bridge amplification, the Cluster Station amplifies isolated molecules on the flowcell to produce each "cluster" of about 1000 identical molecules. After conversion into ssDNA, sequencing primer is hybridized to the cluster. The flowcell is then transferred to the "Analyzer" which is a highly automated fluidic, microscopy, and image capture system. During each synthesis cycle 4 fluorescent nucleotides flow into the flowcell lanes and one fluorescent nucleotide is incorporated into each cluster. After washing and imaging, the newly incorporated fluorescent group is cleaved from each nucleotide and the 3' terminus of the molecule is unblocked to allow the next cycle of synthesis.

Illumina's software consists of 3 modules. The image analysis program called Firecrest is used to extract data from the photographic images. The basecalling module, Bustard, determines the base at each cycle. The 3rd module, Gerald, aligns the sequences to a reference sequence and measures the error rate. Gerald also produces a number of statistical parameters that are used in troubleshooting.

The preliminary results of our validation show that the process requires a minimum of effort for sample preparation and running the instrument. Throughput results have been variable but data quality appears very good and the system is robust. Hardware and software improvements continue and we have focused on protocol development and documentation. We have stressed the efforts of changing run parameters, library preps, and cluster densities in relation to overall throughput and sequence accuracy.

A number of challenges encountered during the study are discussed. The results of our validation show promise in a variety of applications and the system generates large amounts of data with a minimum of labor and expense.

Linking Undergraduate Research to the Discovery of Novel Diazotrophs

A. M. Hirsch[1,2]* (ahirsch@ucla.edu), S. Lee,[3] P. L. DeHoff,[2] N. Milani,[1] J. Alvarado,[4] J. Tzeng,[4] M. Valdés,[5] R. Simons,[2,3] and E. Sanders-Lorenz[3] (erinsl@microbio.ucla.edu)

[1]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles; [2]Molecular Biology Institute, University of California, Los Angeles; [3]Department of Microbiology, Immunology and Molecular Genetics, University of California, Los

Angeles; [4]Department of Ecology and Evolutionary Biology, University of California, Los Angeles; and [5]Departmento de Microbiología, Instituto Politécico Nacional, Mexico City

Students working together in four-person teams in an MIMG laboratory course entitled 'Undergraduate Research in Microbial Molecular Biology, Ecology, Diversity, and Evolution' looked for diazotrophs in different soil types in the Mildred E. Mathias Botanical Garden at UCLA. They collected soil around the roots of several plants including the legume *Gleditsia* and the gymnosperm *Podocarpus*. Returning to the lab, they processed the soil samples for analysis by plating the soil suspensions on various nitrogen-free media. DNA, extracted both directly from soil as well as from bacteria growing in nitrogen-free media, was subjected to PCR with 16S rDNA primers, and sequenced. 16S rDNA analysis showed that two strains (N27 and N13), which grow under nitrogen-free conditions, are *Streptomyces*, and are in the process of being analyzed for the presence of the *nifH* gene.

Undergraduate students have also analyzed novel diazotrophs originally isolated in Mexico and shown to have *nifH* (Valdés et al., 2005). Based on 16S rDNA analysis, these actinomycetes (strains L5 and 7702) were classed as members of *Micromonospora*, a genus hitherto not known to have nitrogen-fixing bacteria. Western blot analysis, with antibodies specific to the nitrogenase complex (NifHDK), have shown that these strains, as well as N27 and N13, express proteins that cross-react with this antibody, and thus further support the hypothesis that these bacteria fix nitrogen utilizing the nitrogenase complex. Undergraduate researchers are currently examining strains L5 and 7702 for the presence of the *nifD* and *nifK* genes.

Through the Teather and Wood (1982) plate assay, we have also determined that strain L5 secretes cellulase, an important clue as to its potential method of entry into plant roots. We are currently developing quantitative assays to measure the amount of cellulase activity exhibited by each of the diazotrophic actinomycetes.

In addition to yielding possible insights into mechanisms of invasion of plants by potentially symbiotic bacteria, that these bacteria have both nitrogen-fixing and cellulase enzyme activity makes them fruitful organisms for future studies into sustainable energy production. This group of microbes may also have use in the transformation of agricultural byproducts, which now are either burned or wasted, instead of being utilized for improving soil fertility or animal feed.

References

1. Teather, R.M. and Wood, P.J. 1982. Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Appl. Environ. Microbiol.* 43:777-780.
2. Valdés. M., Pérez, N.-O., Estrada de los Santos, P., Caballero-Mellado, J., Normand, P. and Hirsch, A.M. Non-*Frankia* actinomycetes isolated from surface-sterilized roots of *Casuarina equisetifolia*. *Appl. Environ. Microbiol*. 71:460-466.

Diversity of Arsenate-Respiring Bacteria Isolated from Extreme Environments: Reconciling Observations of their Biochemistry with Partially Annotated (or Un-annotated) Genomes

S. E. Hoeft,[1] J. Switzer Blum,[1] R. Tabita,[2] B. Witte,[2] J. F. Stolz,[3] and **R. S. Oremland**[1]* (roremlan@usgs.gov)

[1]U.S. Geological Survey, Menlo Park, CA; [2]The Ohio State University, Columbus, OH; and [3]Duquesne University, Pittsburgh, PA

Arsenate-respiring bacteria have been implicated in the mobilization of arsenic in groundwater by virtue of carrying out the dissimilatory reduction of the less mobile arsenate oxyanion to the more mobile and toxic arsenite oxyanion. Despite the relevance this process has with regard to global water quality and its affect on human health, there is a dearth of information with regard to the genomes of these microorganisms. We have been engaged for a number of years in studying the microbial arsenic cycle in two alkaline/hypersaline, arsenic-rich soda lakes in California, Mono Lake (pH = 9.8; salinity = 90 g/L; As = 0.2 mM) and Searles Lake (pH = 9.8; salinity = 340 g/L; As = 4 mM). Three arsenate-respirers have been isolated from Mono Lake:

Bacillus arseniciselenantis (strain E1H), *B. selenitireducens* (strain MLS10) both low G+C Gram positives, and strain MLMS-1 of the δ-Proteobacteria (Switzer Blum et al., 1998; Hoeft et al., 2004). The two bacilli are obligately anaerobic heterotrophs that respire a number of other electron acceptors in addition to arsenate (e.g., nitrate, selenate/selenite, tellurite, thiosulfate, sulfur). The arsenate reductase of strain MLS10 has been characterized (Afkar et al., 2003). A full genome sequencing of MLS10 was undertaken by JGI, but it has not been completed. Strain MLMS-1 (*Dearsenobacterium thioautotrophicum* canditus) is of interest because its phenotype is obligately-chemo-autotrophic (sulfide or sulfur as its electron donor) and obligately linked to arsenate as its electron acceptor. Preliminary genome analysis by JGI has identified homologs of genes for nitrogen fixation and components of methane mono-oxygenase, although it is doubtful that the latter is expressed in such a capacity as the organism is a strict anaerobe. Missing from the genome is a RubisCo enzyme for HCO_3^{2-} fixation and it is possible that a reductive TCA cycle is operative for bringing inorganic C into cellular material. Unfortunately, the full genome for MLMS-1 has not been posted owing to difficulties in assembling its components into one coherent contig.

Finally, strain SLAS-1, an extremely halophilic and alkaliphilic member of the *Haloanaerobiales* was isolated from Searles Lake (Oremland et al., 2005) and molecular techniques have also detected its *in situ* presence within the sediments (Kulp et al., 2006). This microbe is of particular interest because it is capable of both heterotrophic (lactate) and chemo-autotrophic (sulfide) growth using arsenate as its electron acceptor. Radiolabeling experiments with ^{14}C-bicarbonate have demonstrated its capacity for inorganic carbon fixation, but the means by which it achieves this (e.g. RubisCo) have not as yet been elucidated. Strain SLAS-1 would be an excellent choice for a future full genome annotation.

References

1. Switzer Blum et al. (1998) Arch. Microbiol. 171: 19 – 30.
2. Hoeft et al. (2004) Appl. Environ. Microbiol. 70: 2741 – 2747.
3. Afkar et al. (2003) FEMS Microb. Lett. 226: 107 – 112.
4. Oremland et al. (2005) Science 308: 1305 – 1308.
5. Kulp et al. (2006) Appl. Environ. Microbiol. 72: 6514 – 6526.

Comparison of Protocols for Isolating Large Insert Clone DNA that is Suitable for High Throughput Library Construction

Karla Ikeda* (kmikeda@lbl.gov), Eileen Dalin (E_Dalin@lbl.gov), and Jan-Fang Cheng

U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA

JGI has been generating sequences from numerous large insert size clones each year. These targeted sequences allow collaborators to study regions of interest without having to sequence the entire genomes. As part of the assembly QC process, JGI would also sequence a set of fosmids or BACs randomly selected from all eukaryotic genomes to aid in the assembly. Last year JGI has isolated DNA from about 1,200 large insert clones, constructed libraries from each clone, and sequenced them.

We have been using Qiagen® Plasmid Maxiprep protocol to isolate DNA from BAC/fosmid clones, which is laborious and time consuming. In the attempt to find a more efficient way of isolating DNA from these clones, we have compared several protocols including the GenElute™ HP Plasmid Maxiprep from Sigma, the BACMAX™ DNA Purification Kit from Epicentre, and the Edge BioSystems FosPrep™ 96 Fosmid Prep. An overview of each protocol will be presented. We will also provide the comparison of costs, amount of time, DNA yield, quality, and the suitability for library construction from using these DNA isolation protocols.

This work was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. DE-AC52-06NA25396.
UCRL-ABS-228583
LBNL-62528 Abs.

A Symbiont within a Symbiont: “Endomicrobia”, A Class of Intracellular Bacteria in the TG-1 Phylum, Exclusively Colonize Termite Gut Flagellates

Wakako Ikeda-Ohtsubo* (ikeda@mpi-marburg.mpg.de) and Andreas Brune (brune@mpi-marburg.mpg.de)

Department of Biogeochemistry, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany

“Endomicrobia” are a separate class of uncultivated bacteria in the candidate phylum Termite Group I (TG-1). They occur exclusively in the cytoplasm of anaerobic flagellates in the guts of termites and wood-feeding cockroaches, often at high density (>1000 cells per flagellate), and ultrastructural evidence suggests they are obligate endosymbionts. A full-cycle-rRNA approach revealed that the high phylogenetic diversity of “Endomicrobia” in termite guts reflects the generic affiliation of the host flagellates. Parallel phylogenetic analyses of *Trichonympha* flagellates and their symbionts “Endomicrobia” showed congruence of the corresponding SSU rRNA trees, indicating cospeciation of this symbiotic pair. “*Candidatus* Endomicrobium trichonymphae” (CET), a representative of “Endomicrobia” occurring in the flagellate *Trichonympha* of the termite *Zootermopsis nevadensis*, has been selected as a model organism for further investigations. Since obligate symbionts are generally recalcitrant to cultivation, we established a procedure to prepare the genomic DNA of CET from enriched symbiont cells by physical separation,

which removed most of the contaminating bacteria originating from the termite gut content. The genomes of CET and *Elusimicrobium minutum* strain Pei191, the first isolate from the TG-1 phylum (see poster by D. Herlemann, O. Geissinger, and A. Brune) are presently being sequenced (CSP_777133 by A. Brune, J. Leadbetter, M. Friedrich, and P. Hugenholtz), which will provide further insights into the nature and evolutionary status of this intriguing symbiosis.

Why is *Arthrobacter chlorophenolicus* an Interesting Sequencing Target?

Janet K. Jansson* (janet.jansson@mikrob.slu.se)

Department of Microbiology, Swedish University of Agricultural Sciences, Genetics Center, Uppsala, Sweden.

Arthrobacter chlorophenolicus is an actinobacterium that was isolated from soil enriched with increasing concentrations of 4-chlorophenol (4-CP). When isolated it was demonstrated to be able to grow on and degrade unusually high concentrations of 4-CP (up to 350 ppm) and related p-substituted phenolic compounds and phenol.
A. chlorophenolicus can degrade mixtures of phenolic compounds and it does so sequentially with the preference: 4-nitrophenol, 4-chlorophenol/4-bromophenol and phenol, although this is inversely related to their toxicity. Kinetic measurements have shown that this preference is not due to diauxie, but to the rate of uptake of the compounds across the cell membrane. *A. chlorophenolicus* utilizes a novel route for degradation of 4-chlorophenol via hydroxyquinol as an intermediate. A 4-chlorophenol-degradation operon has been cloned and sequenced. Many of the genes in the operon are duplicated and analysis of %GC content indicate that the duplicated genes arose via lateral gene transfer. When one of the key enzymes in the degradation pathway (hydroxyquinol dioxygenase) is disrupted, the cells are no longer able to grow on 4-CP or 4-NP, but exhibit better growth on phenol than the wild type strain. This indicates that this organism uses different pathways for degradation of phenol compared to substituted phenols and suggests that they are regulated differently. *A. chlorophenolicus* can degrade 4-CP at 5°C and during temperature switches between 5°C and 28°C. The cells were tagged with the *gfp* gene, encoding green fluorescent protein, enabling them to be specifically monitored in soil incubated at different temperatures. The physiological status of the cells was determined using viability stains to stain live and dead cell fractions and the stained fractions were enumerated in soil by flow cytometry. Recently it was found that the cells survive better in soil in a presumably dormant state at 5°C compared to 28°C. The *A. chlorophenolicus* genome sequence data will be used to enable us to better understand cold soil survival by actinobacter and how this property can be used to enhance bioremediation rates in temperate climates. In addition, the *A. chlorophenolicus* genome will be compared to that of a related actinobacterium, *Micrococcus luteus*, that also has unusual survival capabilities (*see separate poster presented by Charles Greenblatt*). We are particularly interested in elucidation of the physiological status of the non-spore-forming dormant state in actinobacteria.

Comparative Genomics of Two Biosynthetically Rich Marine Actinomycetes

Paul R. Jensen* (pjensen@ucsd.edu), Kevin Penn, Dan Udwary, and Bradley S. Moore

Scripps Institution of Oceanography, Center for Marine Biotechnology and Biomedicine, University of California, San Diego

Actinomycetes are frequently cultured from marine sediments yet little is known about their ecological roles in the marine environment or how they differ from non-marine strains. Comparative genomics of the recently described marine actinomycetes *Salinispora tropica* and *S. arenicola* reveal circular chromosomes of 5.1 and 5.6 mb, respectively. The approximately 1000 unique genes possessed by each of these species provide insight into potential ecological differences, while the more than 1500 common genes that distinguish them from non-marine actinomycetes hold clues to the genetic basis of marine adaptation in Gram-positive bacteria. Remarkably, approximately 0.5 mb of each genome is devoted to secondary metabolite production, which includes the most diverse array of biosynthetic mechanisms observed to date in bacteria. These genes are assembled into large biosynthetic clusters that are frequently associated with mobile genetic elements and can be used to predict the type of molecule produced and, in some cases, its exact chemical structure. Bioinformatic analysis of a mixed polyketide synthase/non-ribosomal peptide synthetase cluster in the *S. tropica* genome was used to solve the structure of one compound (salinilactam) and resolve ambiguities in contig assembly that hindered genome closure. Dramatic differences in the secondary metabolites produced by these two closely related and co-occurring species imply horizontal gene transfer as the mechanism of gene acquisition and support the hypothesis that these compounds have important ecological functions that may include opportunities for niche differentiation.

Microbial Finishing Informatics System

Pat Kale* (kale1@llnl.gov) and Stephan Trong

DOE Joint Genome Institute, Walnut Creek, CA

Mycosphaerella graminicola Sequencing Heads Towards the First Finished Genome of a Filamentous Plant Pathogenic Fungus

Gerrit H. J. Kema[1]* (gert.kema@wur.nl), Larry D. Dunkle,[2] Alice C. L. Churchill,[3] Jean Carlier,[4] Andy James,[5] Manoel T. Sousa, Jr.,[6] Pedro Crous,[7] Nicolas Roux,[8] Theo A. J. van der Lee,[1] Alexander Wittenberg,[1] Erika Lindquist,[9] Igor Grigoriev,[9] Jim Bristow,[9] Jane Grimwood,[9] and Stephen B. Goodwin[2] (sgoodwin@purdue.edu)

[1]Plant Research International B.V., Wageningen, The Netherlands; [2]USDA Agricultural Research Service, Purdue University, West Lafayette, IN; [3]Department of Plant Pathology, Cornell University, Ithaca, NY; [4]UMR BGPI, CIRAD, Montpellier, France; [5]CICY, Merida, Mexico; [6]EMBRAPA Genetic Resources and Biotechnology, Brasília, Brazil; [7]Fungal Biodiversity Centre, Utrecht, The Netherlands; [8]Bioversity, Montpellier, France; and [9]DOE Joint Genome Institute, Walnut Creek, CA

Mycosphaerella is one of the largest genera of plant pathogenic fungi with more than 1,000 named species, many of which are important pathogens causing leaf spotting diseases in a

wide variety of crops including cereals, citrus, banana, eucalypts, soft fruits, and horticultural crops. A few species of Mycosphaerella cause disease in humans and other vertebrates. An international project was initiated to sequence the genomes of *M. graminicola* and *M. fijiensis*, two of the most economically important pathogens of wheat and banana, respectively, along with 40,000 ESTs from *M. fijiensis* and the related maize pathogen *Cercospora zeae-maydis*, through the Community Sequencing Program sponsored by the U.S. DOE-Joint Genome Institute. The 9x *M. graminicola* sequencing is complete and was made public November 1, 2006 following automated and manual annotation. Due to the very good assembly statistics as well as a >2000-marker DArT linkage map that was aligned to the genome, JGI decided to finish the *M. graminicola* genome at the Stanford Human Genome Center. The majority of chromosomes have been sequenced completely including both telomeres. These data indicate that *M. graminicola* has both the largest chromosome number and the smallest chromosome sizes recorded among filamentous ascomycetes. Detailed analyses of progeny isolates showed the plasticity of the *M. graminicola* genome, as a significant number of chromosomes can be missed and potentially contain redundant information for pathogenicity and mating. The repetitive content of the individual chromosomes is significantly but disproportionally larger on the <2Mb chromosomes indicating that at least some may be considered supernumerary. Major factors for pathogenicity, however, are located on the larger chromosomes and mostly towards the telomeres. More than 1200 genes have been annotated manually and indicate that this pathogen contains smaller gene families compared to other well-studied filamentous plant pathogenic fungi, which may reflect its incognito life style. The *M. fijiensis* EST sequencing has resulted in more than 30,000 ESTs and the genome sequencing resulted in an approximately 70Mb genome size that contains a substantial amount of repetitions. Interestingly, the genome sequence showed a double CG-peak that is absent in the *M. graminicola* genome and appears to be specific for Mycosphaerella pathogens of banana. The current status of both sequencing projects will be discussed.

High Quality Microbial Finishing at JGI

Alla Lapidus[1]* (ALapidus@lbl.gov), Eugene Goltsman,[1] Steve Lowry,[1] Hui Sun,[1] Alicia Clum,[1] Brian Foster, Stephan Trong,[1] Pat Kale,[1] Alex Copeland,[1] Patrick Chain,[2] Cliff Han,[3] Tom Brettin,[3] Jeremy Schmutz,[4] and Paul Richardson[1]

[1]DOE Joint Genome Institute, Walnut Creek, CA; [2]JGI-Lawrence Livermore National Laboratory, Livermore, CA; [3]JGI-Los Alamos National Laboratory, Los Alamos, NM; and [4]JGI-Stanford, Stanford, CA

The value of complete microbial genome sequence is established and appreciated by scientific community. A finished genome represents the genome assembly of high accuracy and quality (with no gaps), verified and confirmed through a number of computer and lab experiments. Several yeas ago JGI has established a set of high standards for the final microbial assembly and has been strictly following them thereafter.

More than 100 microbial projects have been completed since that time within the framework of the JGI's portfolio (DOE GTL program, DOE Microbial program and the Community Sequencing Program). Progress in DNA sequencing technology, design of new vectors for library construction, improvements in finishing strategies and tools, as well as the availability of a number of assemblers and advanced methods for OFR finding and genome annotation have significantly reduced the time required for genome closure. Despite this fact, complexity and speed of genome closure depends on the quality of DNA

received, the whole genome shotgun libraries produced from this DNA, GC content of the genome, the size and frequency of identical or nearly identical repetitive structures, and the amount of regions that can not be cloned or had to clone in *E. coli*. The whole genome finishing/assembly improvement pipeline will be presented showing the lab approaches and computational finishing techniques developed and implemented at JGI for finishing the large number of microbial projects in the queue. We also will present our progress in completing metagenomic projects. A number of projects for which the combination of different sequencing technologies (Sanger and 454) and finishing strategies were used will also be presented.

Survey of the *Brachypodium distachyon* Genome: A Potential Model Species for the Grasses

Gerard R. Lazo[1]* (lazo@pw.usda.gov), Yong Q. Gu,[1] Naxin Huo,[1] John P. Vogel,[1] Ming-Cheng Luo,[2] and Olin D. Anderson[1]

[1]USDA Agricultural Research Service, Western Regional Research Center, Albany, CA and [2]Department of Plant Sciences, University of California, Davis, CA

The plant *Brachypodium distachyon* has demonstrated potential to serve the grass species as a model organism, having many desirable physical, genetic, and genomic attributes. In many ways, the attributes which made *Arabidopsis thaliana* a good model system for dicotyledonous plants, also holds true for *B. distachyon*. Several grass species are currently under consideration to be developed as a major source for bioenergy in the United States, and it is possible that *B. distachyon* can be used as a tool to develop an understanding of the biology for grasses, and to design strategies for crop improvement. Sequencing of the entire *B. distachyon* genome is underway by the U.S. Department of Energy – Joint Genome Institute (DOE-JGI) Community Sequencing Program. In preparation for this genome information, a preview of the *B. distachyon* genome was sampled through sequencing.

A preliminary screening of sequences from 20,440 expressed sequence tags (ESTs) of five cDNA libraries derived from developing seed head, stems plus sheath, root, leaf, and callus assembled into a collection of 2872 different gene candidates. A survey of over 60,000 BAC-end sequences (BES) from three deep-coverage BAC libraries helped to reveal the general composition, structure, and organization of the *B. distachyon* genome. This essentially represented about ten percent of the estimated 355 Mb genome. About forty percent of the BES had good matches to NCBI dbEST expressed sequences suggesting a high density of transcriptional regions; less common in complex grass species such as *Triticum aestivum* (bread wheat). In comparison with other related grass species, *B. distachyon* appeared closely related to wheat and barley, and distantly related to rice, corn, sorghum, and sugarcane, respectively.

Using repetitive element databases, including plant family Poaceae references, only ten percent of the BES were masked or annotated as matching repetitive sequences. A significant portion of the matches were to retrotransposons of the Copia and Gypsy class of retroelements. There were other simple repeat and satellite DNA sequences, and some that may be characterized as new elements. The *B. distachyon* genome will provide easier access to genes in more complex grass genomes.

The BAC libraries are currently being used to construct physical maps for the *B. distachyon* genome. Identified gene-related BES compared *in silico* to the sequenced rice

genome, have identified conserved and divergent regions based on gene colinearity between *Brachypodium* and rice. Some BES also have significant matches to mapped wheat ESTs to be used in marker development. The prospects are promising for the use of *B. distachyon* as a model species for the grasses.

Phylogenetic Diversity of a Soil Microbial Community

To Hang Shela Lee,[1] Karen Flummerfelt,[1] Ann Hirsch,[2] Robert Simons,[1] and **Erin Sanders-Lorenz**[1]* (erinsl@microbio.ucla.edu)

[1]Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, CA and [2]Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, CA

Changes in biotechnology are driving a nationwide research initiative to transform undergraduate science education, in which the goal is to devise teaching methods that generate excitement about the discovery process central to scientific research. To incorporate this initiative into its curriculum, UCLA's Department of Microbiology, Immunology, and Molecular Genetics launched a project-based laboratory course entitled 'Undergraduate Research in Microbial Molecular Biology, Ecology, Diversity, and Evolution' in which students explore microbial diversity within environmental samples. The project presented on this poster represents the work that a single student participating in the course accomplished in a single 10-week quarter. Here, the goal was to reconstruct the phylogeny of a bacterial community, cultivated from soil collected from the base of a cabbage tree (*Cussonia natalensis*) in the UCLA Botanical Garden, using small subunit ribosomal RNA (SSU rRNA) genes. Genomic DNA was harvested from 24 microbial isolates, then subjected to PCR and sequencing of SSU rRNA genes. DNA sequences were compared to those of classified organisms in both the Genbank database using BLAST and the RDP (http://rdp.cme.msu.edu/). A more detailed phylogenetic picture of the soil community was obtained by building an evolutionary distance-based tree. The analysis revealed that many of the isolates belong to well-characterized taxonomic groups, including the class *Gammaproteobacteria* and the phyla *Firmicutes* and *Actinobacteria*. Other isolates belong to classes having few cultivated representatives such as the *Flavobacteria* and *Sphingobacteria*. Because soil microorganisms are prolific manufacturers of antibiotic substances, isolates were screened for antibiotic production. Two produced antibiotics inhibiting growth of an indicator bacterial strain. Isolates were also surveyed for resistance to antibiotics. Seven displayed resistance to a single drug, while one showed multiple-resistance. These phenotypic observations demonstrate a range of ecological strategies employed by microbes in this habitat. However, the cultivation-dependent analysis of the soil community clearly did not reflect the phylogenetic diversity expected from a survey of a terrestrial environment. Instead, the tree biased those organisms that were best fit for growth under the environmental conditions selected for the study. Thus, in current renditions of the course, students purify total DNA representing the entire metagenome directly from the same soil sample concomitantly subjected to cultivation-dependent analysis. Following PCR amplification and sequencing of SSU rRNA genes in environmental clones, students generate a phylogenetic tree representing the 'uncultivatable' bacterial population, albeit the small number of clones studied is not sufficient to accurately reflect the full diversity of bacteria present within the soil environment. Moreover, taxonomic classification of the environmental sequences proves to be challenging due to the limited diversity of well-characterized strains available for comparison in Genbank and RDP. Nevertheless, students are still able to ascertain how well the isolates represent the diversity of microorganisms portrayed in the environmental

trees, finding taxa rarely overlap between the two trees. Thus, the two single-gene methodologies used to explore the composition of microbial communities complement each other nicely in an instructional laboratory setting, being limited only by the number of DNA sequences available to students when conducting their analyses.

Sequencing Metabolically Diverse *Pyrobaculum* and Related Hyperthermophilic Archaeal Species

Todd M. Lowe[1]* (lowe@soe.ucsc.edu), Jennifer F. Biddle,[2] Aaron Cozen,[1] Zhidan Zhang,[2] Sorel T. Fitz-Gibbon,[3] Christopher H. House,[2] and Chad Saltikov[4]

[1]Department of Biomolecular Engineering, University of California, Santa Cruz, CA; [2]Department of Geosciences, Pennsylvania State University, University Park, PA; [3]IGPP Center for Astrobiology, University of California, Los Angeles, CA; [4]Department of Environmental Toxicology, University of California, Santa Cruz, CA

Members of the *Pyrobaculum* genus are widespread and abundant in geothermal environments, and represent a unique clade among the Archaea because its members respire toxic metals such as arsenic and selenium, sulfur compounds, iron, nitrate, and oxygen. In contrast to the crenarchaeal model *Sulfolobus*, which is an acidophilic obligate aerobe, *Pyrobaculum* provides an alternative crenarchaeal system to investigate biological processes in non-acidic, aerobic/anaerobic, high-temperature environments. The only member of this genus sequenced prior to 2006, *Pyrobaculum aerophilum*, is not easily cultured, which has slowed research in this compelling genus. By COG analysis, 39% of *P. aerophilum* genes have no orthologs and/or no assigned function, representing a potentially rich source of new extremophile biology. Because other *Pyrobaculum* species may provide a better experimental model, and comparative genomics has been shown to expedite model organism research, the *Pyrobaculum* Consortium (Fitz-Gibbon, House, Lowe, Saltikov) proposed to sequence four species closely-related to *P. aerophilum* (*Pyrobaculum calidifontis* [*Pca*]*, Pyrobaculum islandicum* [*Pis*], *Pyrobaculum arsenaticum* [*Par*], *Thermoproteus neutrophilus*), plus one slightly more distant member within the same family (*Caldivirga maquilingensis,* [*Cma*]).

Genomic DNA was prepared by members of the *Pyrobaculum* Consortium to be sequenced by the Joint Genome Institute as part of the 2006 Community Sequencing Program. Currently, two genomes are complete (*Pca, Pis*) and two are in quality draft stage (*Par, Cma*). All genomes are 1.8-2.1 million base pairs in size, although G/C content is surprisingly diverse for members of the same family, ranging from 57.1% (*Pca*) to 49.6% (*Pis*) to 43.1% (*Cma*). In addition to protein gene prediction and annotation provided by Oak Ridge National Labs, our lab has created five-way full-genome alignments within the Archaeal Genome Browser (http://archaea.ucsc.edu/), improved non-coding RNA gene annotation, and used a variety of comparative methods to observe gene gain/loss patterns to study genome evolution and operon conservation with respect to species' physiology.

One intriguing result is the large number of genome insertions, approximately 150-200 per genome. A majority of insertion regions include apparent multi-gene operons that often have no detectable sequence similarity to proteins from related species. *Pyrobaculum* operons also appear to be much shorter on average relative to all other archaea and bacteria. These observations suggest a high natural rate of horizontal transfer of multi-gene DNA segments, potentially from a large reservoir of unstudied hyperthermophile crenarchaeal viruses (no plasmids have been identified from any of these species). The

high rate of new operon acquisition may help explain why *Pyrobaculum* species can employ variable combinations of five different forms of respiration (most living things respire on only one oxidant).

Comparative analysis of transfer RNA genes also reveals the highest rate of new intron insertion for *any* species known to date. The origin and mechanism of these insertions is unknown, but may be related to the unusual genome dynamics observed for protein coding genes. These new genomes have already opened new lines of investigation in *Pyrobaculum* research, and provide an unprecedented degree of comparative power in continued analyses of gene content, operon structure, regulatory sequences, and whole-genome evolution of hyperthermophilic archaea.

Microbial Genomics, Comparative and Functional Analyses at JGI-LLNL

Stephanie Malfatti* (malfatti3@llnl.gov), Lisa Vergez, Maria Shin, Mari Christensen, Jeff Elliott, Dorothy Lang, and Patrick Chain

Lawrence Livermore National Laboratory, Joint Genome Institute, Livermore, CA

Since the initiation of the Department of Energy's Joint Genome Institute, and as part of the DOE's Microbial Genome Program, the LLNL microbial genomics group has, as part of the JGI, been involved in various aspects of delivering finished genomes and performing detailed analyses, including comparative genomics, for publication purposes. Though we have only published 18 of our ~50 finished bacterial genomes, we are currently in the final phases of analysis for an additional 10 and are in various stages of annotation and comparative analyses for another 16 microbial genomes. Here, we outline our data management and finishing processes, and also present a few of our more recently completed microbial genomes, including *Sinorhizobium medicae* WSM419, *Pseudomonas putida* F1 and *Psychrobacter* sp. PRwf-1.

This work was performed under the auspices of the U.S. DOE by the University of California, LLNL under Contract No. W-7405-Eng-48.

Effective Training on JGI's Integrated Microbial Genomes System

Mary E. Mangan,[1] Jennifer Williams,[1] Scott M. Lathe[1] (slathe@openhelix.com), Kenneth J. Klima,[2] George Taniwaki,[2] and **Warren C. Lathe III***[1] (wclathe@openhelix.com)

[1]OpenHelix, LLC. Bellevue, Washington and [2]Hebert Research. Seattle, Washington

Concurrent with the accumulation of sequence data since the beginning of the Human and other genome projects, there has been an exponential rise in the number of databases and analysis resources. One example of this rise is the annual database issue of Nucleic Acid Research (NAR). In 1996 it reported just over 50 resources [1]. The latest issue [2] included 968 resources, 110 more than the previous year [3]. Yet, these online resources are only being used to a small portion of their potential benefit due to lack of training. The need for training has become more acute as data accumulate and new resources continue to grow in novelty, speed and complexity [4]. OpenHelix tested which of several methods would be effective in training researchers on the use of these genome resources and found that online pre-recorded tutorials are highly effective and as effective as live lecture-type training.

Using this research as a springboard, OpenHelix developed and now provides online training materials on a large number of resources, including the Integrated Microbial Genomes system. The IMG training materials are funded by JGI and free of charge to IMG users. They include an online pre-recorded tutorial, and PowerPoint slides with a full, suggested script, handouts and exercises for download. Researchers can use these materials to train themselves or as aids in teaching a course or seminar.

References

1. (1996) Entire Issue. Nucleic Acids Research (24) 1-252. .
2. Galperin MY (2007) The Molecular Biology Database Collection: 2007 update. Nucleic Acids Research (34): D3-D4.
3. Galperin MY (2006) The Molecular Biology Database Collection: 2006 update. Nucleic Acids Research (33): D3-D5.
4. Collins FS, Green ED, Guttmacher AE, Guyer MS (2003) A vision for the future of genomics research. Nature Genetics 422: 835-847.

Chimera Free Cloning of Single DNA Inserts Using “GC Cloning”

David Mead* (dmead@lucigen.com), Rebecaa Hochstein, Keynttisha Jefferson, Ronald Godiska, and Spencer Hermanson

Lucigen Corporation, Middleton, WI

The efficiency of shotgun DNA sequencing depends to a great extent on the quality of the random libraries used. It is important to minimize chimeric inserts to facilitate accurate sequence assembly. Single-insert clones are usually ensured by ligating asymmetric linkers, typically with a BstXI recognition site, to the insert DNA. Subsequently, the excess linkers must be completely removed from the insert DNA before ligation to a vector containing complementary BstXI ends. We have developed a novel “linker free” cloning strategy to eliminate the need for linker addition and removal in constructing high-quality shotgun libraries. It is based on a new GC cloning technology and a unique DNA end blocking chemistry developed at Lucigen. The pSMARTGC vector contains a single 3’-C overhang, which is compatible with the single 3’-G overhang added to blunt ended DNA using PyroPhage DNA polymerase or many other non-proofreading polymerases. The unique combination of a C tailed vector and G tailed insert blocks the ligation of multiple fragments. This protocol is robust and showed five to ten-fold higher yields of clones compared to previous protocols, and is significantly faster than TA cloning. The level of chimerism is ~ 1% in the library, and the background of clones without an insert was <1%. Another important benefit is the ability to construct complex libraries using 10-100 ng insert without compromising the sequence coverage of level of empty background. The procedure is very rapid, as libraries were completely processed in a day. High copy, low copy, single copy and linear vector versions of the GC cloning vectors with chimera free capabilities have been constructed.

Genomic Resources for Two Reef-Building Caribbean Corals, *Acropora palmata* and *Montastraea faveolata*

Mónica Medina[1]* (mmedina@ucmerced.edu), Jodi Schwarz,[2] Peter Brokstein,[3] Shini Sunagawa,[1] Michael DeSalvo[1] (mdesalvo@ucmerced.edu), Chris Voolstra,[1] Collin Closek,[1] Mary Alice Coffroth,[4] and Alina Szmant[5]

[1]University of California, Merced; [2]Vassar University, Poughkeepsie, NY; [3]DOE Joint Genome Institute, Walnut Creek, CA; [4]State University of New York (SUNY), Buffalo; and [5]University of North Carolina, Wilmington

Coral reefs are among the most diverse and important ecosystems on Earth. Around 500 million people depend on coral reefs for food, coastal protection and other resources; however coral reefs are threatened by the declining health of the marine environment. Coral bleaching and disease are causing the world's reefs to suffer drastic declines in coral cover. At the heart of the reef ecosystem is a symbiosis between coral hosts and endosymbiotic dinoflagellates (*Symbiodinium* spp.). The importance of this symbiosis to coral physiology and ecosystem-wide nutrient cycling is well documented, but little is known about the cellular and molecular mechanisms by which the partners establish and regulate the symbiosis. We have developed cDNA libraries for different life stages of the Caribbean corals *Montastraea faveolata* and *Acropora palmata*, and these libraries are being used to print gene expression microarrays. cDNA libraries and microarrays are also being developed for two *Symbiodinium* strains that infect *M. faveolata* and *A. palmata*. By experimentally infecting aposymbiotic coral larvae with dinoflagellates, we will measure gene expression changes that are correlated with the onset and maintenance of symbiosis. These observations will provide new knowledge and avenues of research to help protect these endangered organisms. We are also conducting experiments to address: 1) the molecular basis of coral bleaching; 2) the mechanisms of calcification in corals; 3) the molecular basis of coral diseases; 4) speciation genes in the *M. annularis* species complex; and 5) speciation genes between *A. palmata* and *A. cervicornis*.

Genes Involved in Pyrene Degradation are Clustered in PAH-Degrading Mycobacteria

Charles Miller[1,2]* (cdmiller@biology.usu.edu)**,** Yanna Liang,[2] Zhang Chun,[1] Ronald Sims,[2] and Anne Anderson[1,2]

[1]Department of Biology and [2]Department of Biological and Irrigation Engineering, Utah State University, Logan, UT

Three polycyclic aromatic hydrocarbon (PAH)-mineralizing mycobacterium isolates, JLS, KMS and MCS, were isolated from a wood preservative-contaminated land-treatment unit where active PAH remediation was occurring. The genomic sequencing of these mycobacteria has recently been completed by the Joint Genomes Institute. These isolates have large genomes (~ 6 Mb), are GC rich (~68%) and have extensive numbers (66-80) of dioxygenase genes likely involved in aromatic ring-hydroxylation and opening, processes essential for degradation. These isolates differ in plasmid content with KMS having a linear and circular plasmid, MCS having a linear plasmid and JLS having no plasmids. 2D-gel electrophoresis shows a proteome for strain KMS of about 450 peptides in the PI 4-7 range with 15-18 proteins that are upregulated and 5-7 newly produced in the presence of pyrene. The enzymes involved in degradation of pyrene appear similar to those in mycobacterium isolates studied by other groups. Genes corresponding to the pyrene-

induced proteins are clustered in both the genome and plasmids for isolate KMS. Furthermore, large regions (up to 16 kb) containing these genes are duplicated in the genome and plasmids.

Sequence Analysis of Megaplasmids from the Dioxin Mineralizing Bacterium *Sphingomonas wittichii* RW1

Todd R. Miller[1]* (trmiller@jhsph.edu), Steven L. Salzberg,[2] and Rolf U. Halden[1]

[1]Center for Water and Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD and [2]University of Maryland Center for Bioinformatics and Computational Biology, College Park, MD

Bacteria phylogenetically related to *Sphingomonas* species are frequently found associated with the degradation of toxic aromatic hydrocarbons in the environment. The bacterium, *Sphingomonas wittichii* RW1 is a model organism for aromatic hydrocarbon degradation as it mineralizes dibenzo-*p*-dioxin, dibenzofuran and co-oxidizes several of their polychlorinated and hydroxylated congeners. The genome of *S. wittichii* RW1 was recently sequenced by the Joint Genome Institute in an attempt to better understand the genetic basis for aromatic hydrocarbon degradation in this organism. During assembly of the genome, two circular megaplasmids were discovered. Here we present an analysis of the genetic organization and physiological functions encoded by the two megaplasmids. The larger plasmid, pSW1 (315 kb), codes for putative proteins involved in carbohydrate and fatty acid metabolism, chemotaxis, conjugal transfer, and camphor resistance. Plasmid pSW2 (225 kb) codes for all proteins previously shown to be required for dioxin degradation: ferredoxin (Fdx1), the dioxin dioxygenase subunits (DxnA1/A2) encoded on a 24 kb region, hydrolase (DxnB), putative receptor protein (DxnC), ferredoxin (Fdx3) and proteins of the 4-hydroxysalicylate/hydroxyquinol degradative pathway (DxnD, DxnF, DxnE, DxnGH, and DxnI). In addition, pSW2 encodes for several mono- and dioxygenases of unknown function, and proteins involved in catechol degradation, as well as elements of a type IV secretion system. Both plasmids contain multiple insertions sequences, transposases and extant phage-related genes.

Characterization of the Drosophila Gut Microbiome

Jenna Morgan* (JLMorgan@lbl.gov), Holly Huse, Deborah Kimbrell, and Jonathan Eisen

Lawrence Berkeley National Laboatory, Berkeley, CA

Use of Near-Neighbor PCR to Close Scaffold Gaps in Microbial Genomes

A. Christine Munk* (cmunk@lanl.gov), Yan Xu, Avinash Kewalramani, Riley Arnaudville, Roxanne Tapia, Thomas S. Brettin, and Cliff S. Han

Joint Genome Institute, Los Alamos National Laboratory, Los Alamos, New Mexico

One of the challenges of finishing microbial genomes is large numbers of uncloned regions (scaffold gaps). Gaps with no clone links require many expensive pcr reactions to connect scaffolds. Ordering and orienting scaffolds with no clone links is difficult unless a closely-related genome is available to identify possible links. This abstract describes near-neighbor PCR (nnPCR), a method of closing these 'scaffold' gaps with a minimum of pcr reactions using closely related finished genomes.

In the manual process of near-neighbor PCR, scaffold ends are identified visually using the Consed program's Assembly view and map of scaffolds. Blast is performed using as query a fasta file containing alternatively 1) the sequence of all unordered contigs >2kb and >10 reads, or 2) the terminal 10kb of each scaffold end. A closely-related-genome downloaded from Genbank is used as subject. The blast output is parsed and a tab-delimited file is produced which can be opened as a spreadsheet. Links between contigs are identified and pcr primers are chosen and paired according to the observed links. PCR is performed and products are end-sequenced and assembled to close scaffold gaps. If PCR products are >1500 bp, they are shattered, subcloned and subclones are end-sequenced and assembled to make a consensus sequence to close scaffold gaps.

In one microbial finishing project with 14 uncaptured gaps considered, 9 out of 14 pcr reactions were successful, and 9 gaps were closed. In another project with 23 uncaptured gaps, 8 out of 19 pcr reactions were successful and 5 gaps were closed. This can be compared to combinatorial PCR, which would require 378 and 703 PCR reactions respectively.

Software to automate and improve this procedure has been developed and is currently being tested. This nnPCR software uses the Consed autoreport function to generate a file which contains a map of contigs in scaffolds, and creates a fasta file for the blast query with sequence from the contigs at scaffold ends. nnPCR uses megablast to search all finished genomes currently in Genbank for hits within the same organism within a 20 kb range. The software chooses primers and pairs them, and submits a work order directly to the lab.

For the project with 14 uncaptured gaps considered, 12 out of 14 pcr reactions chosen by the nnPCR software were successful. For the project with 23 uncaptured gaps, 3 of 6 pcr reactions were successful.

Several microbial genomes have been significantly improved using near-neighbor PCR. This method has significantly reduced the number of PCR reactions that would be required if combinatorial PCR were necessary.

Possible improvements to the software are being tested and results will be included in the poster.

Standard cDNA Library Construction at DOE Joint Genome Institute

Dean Ng* (dng@lbl.gov), Mei Wang, Edward Kirton, Doug Smith, Gladys McCay, Ali Qureshi, Erika Lindquist (EALindquist@lbl.gov), and Paul Richardson

DOE Joint Genome Institute, Walnut Creek CA

cDNA libraries are generated by the JGI to provide expressed sequence tag (EST) sequences for both Eukaryotic whole genome shotgun projects and as EST only projects. The primary goal is to generate high quality cDNA libraries with maximum diversity for gene discovery and to assist with genome annotation, where EST sequences are used for gene model validation and to train gene calling algorithms. In order to maximize gene discovery, acquisition of multiple RNA samples from different tissues, developmental stages or growth conditions may be required.

The standard cDNA library construction protocol does not involve an amplification step, utilizes a size selection step, and produces clones residing in a Gateway vector for ease of subsequent transfer to alternative vectors. The clones themselves are a valuable resource and may be further used by researchers as a gene collection.

In order for the JGI to generate cDNA libraries, collaborators must provide ~500ug of high quality Total RNA along with quantitative and qualitative measurements. Occasionally samples have been found to contain components which erroneously inflate the quantitative measurements. Since high quality RNA is critical to make diverse cDNA libraries, accurate RNA measurements are required. Lithium chloride precipitation has improved many of these samples without significant loss of RNA and therefore is used as part of the standard RNA preparatory procedure.

In order to eliminate abundant ribosomal RNA species and enrich for mRNA, JGI uses poly-adenylated RNA (polyA RNA) purification involving a magnetic bead-oligo dT based process to select polyA RNA. The recovered polyA RNA is used as the template for first strand cDNA synthesis using reverse transcriptase and an oligo dT-NotI primer. Second strand synthesis is performed with *E. coli* DNA ligase, polymerase I, and RNaseH followed by end repair using T4 DNA polymerase. A SalI adaptor is ligated to the cDNA, which is subsequently digested with NotI followed by size selection using gel electrophoresis. Two size ranges of cDNA are cut out of the gel, low with a range of ~600-2kb and high ranging up from ~2kb. Size selected inserts are directionally ligated into the SalI and NotI digested vector pCMVsport6. The ligation is transformed into T1 phage resistant cells and library complexity (total number of clones) is determined.

Library quality is first assessed by randomly selecting a subset of clones and PCR amplifying the cDNA inserts with vector primers (PCR-QC). The number of clones with inserts is determined and the approximate size range of inserts is estimated. Provided the cDNA library passes the PCR-QC, a random set of clones is selected for sequencing. Clones are picked, inoculated into 384 well plates, and grown overnight. Each clone is amplified using rolling circle amplification and the 5’ and 3’ end of each insert is sequenced using vector specific primers and Big Dye chemistry. Library quality is reevaluated to assess insert rate, unusable sequence (eg. rRNA, mitochondrial, noncellular) and library diversity. These data are used to guide which and to what depth cDNA libraries are sequenced.

This work was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No.

DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. DE-AC52-06NA25396. UCRL-ABS-228566

Identification of Quantitative Trait Loci Affecting Virulence in the Basidiomycete *Heterobasidion annosum* s. l.

Åke Olson[1]* (ake.olson@mykopat.slu.se), Mårten Lind,[1] Kerstin Dalman,[1] Bo Karlsson,[2] and Jan Stenlid[1] (jan.stenlid@mykopat.slu.se)

[1]Department of Forest Mycology and Pathology, Swedish University of Agricultural Sciences, Uppsala, Sweden and [2]Skogfors, Ekebo, Svalöv, Sweden

Quantitative trait loci (QTL) for virulence on one-year-old Pinus sylvestris and two-year-old Picea abies seedlings were identified and positioned on a genetic linkage map of *Heterobasidion annosum sensu lato* (s.l.), a devastating root rot pathogen on conifers. The virulence among 102 progeny isolates was analysed using two measurements: lesion length around and fungal growth in sapwood from a cambial infection site.In general we found negative virulence effects of hybridiation although this was contradicted on winter hardened spruce seedlings. On *P. abies*, both measurements identified several partially overlapping QTLs on linkage group (LG) 15 of significant logarithm of odds (LOD) values ranging from 2.77 to 3.85. On *P. sylvestris*, the lesion length measurement also identified a QTL (LOD 3.09) on LG 15. Moreover, QTLs on two separate smaller LGs, with peak LOD values of 3.24 and 4.58 were identified for fungal sapwood growth and lesion lengths respectively. The QTL probably represent loci important for specific as well as general aspects of virulence on *P. sylvestris* and *P. abies*. This is the first report of virulence located in a genetic linkage map of the *Heterobasidion spp.* genome. The mapping of virulence QTLs forms the basis for positional cloning and identification of the corresponding virulence genes, which may give answers to the so far largely unknown factors involved in the infection process of *H. annosum s.l.*

The Genome of *Ostreococcus lucimarinus*

B. Palenik* (bpalenik@ucsd.edu),[1] J. Grimwood,[2] P. Rouze,[3] H. Moreau,[4] I. Grigoriev,[5] and The Ostreococcus Genome Consortium

[1]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA; [2]Stanford/JGI, Palo Alto, CA; [3]Laboratoire Associé de l'INRA (France), Ghent University, Ghent, Belgium; [4]Observatoire Océanologique, Laboratoire Arago, Unité Mixte de Recherche 7628, Centre National de la Recherche Scientifique-Université Pierre et Marie Curie-Paris 6, BP 44, 66651 Banyuls sur mer cedex, France; and [5]U.S. DOE Joint Genome Institute, Walnut Creek, CA

Picophytoplankton with cell diameters less than 2 microns play a significant role in biogeochemical processes, primary productivity, and food webs. Within this size class, the smallest known eukaryotes, at around 1 micron, are *Ostreococcus tauri* and related species. The genome of *Ostreococcus lucimarinus* has recently been completed and compared to that of *O. tauri*. This comparison reveals surprising differences in gene conservation and synteny across orthologous chromosomes in the two species. In addition, the genome of O. lucimarinus provides insights into the metal metabolism of these organisms. In particular, *O. lucimarinus* is predicted to have a large number of selenoproteins. Selenoenzymes are more catalytically active than similar enzymes lacking selenium and thus the cell may require less of that protein. This genome was sequenced by

the DOE Joint Genome Institute (JGI) and analyzed by an international consortium as listed in the poster.

Automating the JGI's Community Sequencing Program

Jayant Patil[1]* (jmpatil@lbl.gov), Joni Fazo[2] (jbfazo@lbl.gov), Art Kobayashi,[2] Annette Greiner,[1] Arkady Voloshin,[1] Pat Kale,[2] Marsha Fenner,[1] David Bruce,[3] Jan-Fang Cheng,[1] and Jim Bristow[1]

[1]DOE Joint Genome Institute, Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [2]DOE Joint Genome Institute, BACE Division, Lawrence Livermore National Laboratory, Livermore, CA; and [3]DOE Joint Genome Institute, Los Alamos National Laboratory, Los Alamos, NM

The Community Sequencing Program (CSP) provides the scientific community at large with access to high-throughput sequencing at the Department of Energy's Joint Genome Institute (JGI) for projects of relevance to DOE missions. Sequencing projects are chosen based on scientific merit--judged through independent peer review--and relevance to issues in global carbon cycling, alternative energy production, and bioremediation.

Each year the JGI issues a call for proposals to participate in the program. Established in 2004 as a manual process, we have since moved to an automated system that manages the proposal submission, review and notification process. Scientists from around the world submit proposals through our website that are later reviewed on-line by our technical and peer review committees. The website is supported by the development team and managed with the on-line tool RequestTracker.

The automated CSP application is a J2EE Internet database application based on a three-tier architecture:

1. The *Presentation Laye*r consists of web user-interface components built in Java servlets, Java Server Faces (JSF) and Java Server Pages (JSP).
2. The *Business Layer* consists of JSF beans and Java command classes.
3. The *Data Access Layer* is an object to relational mapping between data access components developed with the open-source technology Hibernate, and an Oracle database.

The process for managing CSP has been refined each year since the program inception in 2004. The team responsible for developing the CSP application and defining the program requirements has also evolved. We adopted an iterative software development process in Fall of 2006 to adeptly and effectively manage CSP application development and the program as a whole.

Links:
http://www.jgi.doe.gov/CSP/
http://jgi.doe.gov:8080/csp/

Fosmid Ditags as a New Technology Developed at JGI

Ze Peng* (ZPeng@lbl.gov), Ilya Malinov, Doug Smith, Feng Chen, Paul Richardson, Len A. Pennacchio, and Jan-Fang Cheng

U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA

Paired end reads from large insert DNA libraries are essential for detecting chromosome rearrangements as well as connecting sequence scaffolds of draft genomes. However, fosmid and BAC end sequencing remains challenging as well as expensive. Ditag sequencing of fosmid ends represents a cost effective way to generate paired end sequences from large genomic fragments. We present results from several ditag libraries from human, fungi, and bacteria, which were sequenced using 454 technology. Several software tools were developed to analyze the resulted ditag sequences. These tools have been used to (1) create suffix arrays of the reference genomes; (2) filter, trim, and prepare the paired 18mer ditag sequences for analysis; (3) search for 18mer strings for matches; and (4) score the chromosome locations of ditag pairs.

For testing the accuracy and sensitivity of detecting chromosome rearrangements, we have generated 235,394 unique ditag pairs from a breast cancer genome. These fosmid sequence tags represent about 3.1-fold clone coverage of the genome. We have identified 59 rearrangements including 13 translocations, 23 deletions, and 23 inversions. Of those, 14 have been previously detected by an independent approach of using the BAC end sequence profiling data. We are in the process of generating ditags from a green microalga (Micromonas pusilla NOUM17), a deuteromycete fungus (Trichoderma virens Gv29-8), a poplar rust (Melampsora larici-populina 98AG31), and 3 prokaryotes. So far we have found that most ditag pairs could be localized to the draft genomes with the predicted distances and some ditags helped connecting sequence scaffolds to improve the continuity of the assemblies.

The ditag technology in conjunction with the 454 sequencing provides a high throughput approach to assist shotgun sequence assemblies and characterize cancer genomes.

This work was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.

LBNL-62522 Abs.

Microarrays + NanoSIMS: Linking Microbial Identity and Function with 'NanoSIP'

Jennifer Pett-Ridge[1]* (pettridge2@llnl.gov), Eoin Brodie,[2] Peter Weber,[1] Paul Hoeprich,[1] Philip Banda,[1] Ian Hutcheon,[1] and Gary Andersen[2]

[1]Lawrence Livermore National Laboratory, Livermore, CA and [2]Lawrence Berkeley National Laboratory, Berkeley, CA

In order to predict how microbes may react under given environmental conditions, or be engineered to perform useful functions, it is essential to understand the relationships between their molecular and metabolic profiles. Indeed, our need to understand both the identity and functional capacity of microorganisms is increasing as researchers seek to: a) understand spatial and metabolic relationships within complex microbial communities, b)

exploit microbial traits for bioengineered fuel cells and cellulose conversion to biofuels, and c) utilize microbes to remediate contaminated sites.

We are addressing these goals by developing a new methodology, "NanoSIP", combining the power of re-designed oligonucleotide microarrays with nano-scale secondary ion mass spectrometry (NanoSIMS) analyses in order to link the identity of microbes to their functional roles. Building upon the established concept of stable isotope probing (SIP), we are isotopically labeling microbial nucleic acids by growing organisms on ^{13}C enriched substrates. When hybridized to a high density oligonucleotide microarray we can use the high spatial resolution and high sensitivity of the NanoSIMS to detect isotopic enrichment in ribosomal RNA fragments identified through fluorescent hybridization to a newly engineered oligonucleotide microarray. This approach will allow us to directly link microbial identity and function.

The NanoSIMS is an imaging secondary ion mass spectrometer with the unprecedented combination of high spatial resolution, high sensitivity and high mass specificity. It has 50 nm lateral resolution and is capable of detecting 1 of every 200 carbon atoms in a sample while excluding isobaric interferences. We have previously used the NanoSIMS to document isotopic and elemental variations in tiny bioparticles such as *Bacilllus* spores, bacterial cells and lipid bilayers. Since the spot or feature size on a microarray is typically microns in diameter, and can contain millions of copies of an oligonucleotide probe, the NanoSIMS has the detection capability to resolve array spots labeled with ^{12}C rRNA from those labeled with ^{13}C rRNA.

We are currently in the 'proof-of-concept' phase of method development and are testing the technique using pure cultures of ^{13}C-labelled microbes. Using environmental isolates from soil, we cultured 2 strains each of fungi, gm (+) bacteria, gm (-) bacteria and actinomycetes with ^{13}C-glucose. Cultures were repeatedly subsampled during exponential phase growth in order to generate a set of samples with a range of isotopic enrichments. We have extracted DNA from these isolates, sequenced the 16S/ITS region and generated 25-mer oligonucleotide probes for each organism. This probe set has been printed onto high density oligonucleotide microarrays using the NimbelGen synthesizer in the LLNL-Livermore Microarray Center (LMAC). The arrays we are using are newly engineered to have a more conductive surface and higher reproducibility relative to traditional glass/silane microarrays. These advances allow us to successfully analyze microarray slides with a nano-secondary ion mass spectrometer (NanoSIMS), generating isotopic and elemental abundance images of the array surface, and indicating which organisms utilized the isotopically labeled substrate. We intend to apply the method to complex microbial communities found in biofilms and soils in the near future.

Genetic Questions that We Expect to Answer by Sequencing the Genome of the White-Rot Basidiomycete *Pleurotus ostreatus*

Antonio G. Pisabarro* (gpisabarro@unavarra.es) and Lucía Ramírez

Public University of Navarre, Pamplona, Spain

The sequencing of the *P. ostreatus* genome been scheduled for 2007 by the JGI. *P. ostreatus* has been studied as edible fungus, as model for lignin degradation and as source of poly-aromatic hydrocarbon degrading enzymes. However, the genetic questions that can be addressed using this organism as model are far beyond that. This organism is dikaryotic in all its life cycle except at the basidia. This allows separating the two nuclei present in a

mature sexually competent individual to study both parental haplotypes. Electrophoretic analyses of *P. ostreatus* have revealed prominent length polymorphisms in homologous chromosomes. The sequence of the two sets of homologous chromosomes will make available the two haplotypes of a mature organism for the first time. This will permit to determine the minimal synteny between homologous chromosomes and to study the meiotic behaviour of structurally polymorphic chromosomes. In *P. ostreatus* various quantitative loci (QTLs) have been identified whose genetic architecture is poorly known. The sequence of the two haplotypes will permit to study this architecture and to evaluate the contribution of other genomic regions interacting with it. Finally, the study of the contributions of the two haplotypes to subcellular structures (nucleus, mitochondria) will be possible, and this will open the door to deeper analyses of the intercommunication between the two nuclei (genomes) present in a sexually competent organism.

NCBI – The Final Destination For Sequenced Data

Samuel Pitluck* (s_pitluck@lbl.gov)

DOE Joint Genome Institute, Walnut Creek, CA

Almost all of the data sequenced at the JGI will eventually be deposited at the National Center for Biotechnology Information (NCBI). We will discuss the various types of data that are deposited at NCBI, and what types of information collaborators should provide to make this process go smoothly. We will describe our interaction with NCBI at various times during the sequencing process. As an example, we will describe the contacts with NCBI from start to finish for a microbial submission.

Jazz – The JGI Genome Assembler

Nik Putnam, Jarrod Chapman, Dan Rokhsar, **Isaac Ho*** (iyho@lbl.gov), Serge Dusheyko, Craig Furman, Sara Ting, Hank Tu, and Michael Zhang

DOE Joint Genome Institute, Walnut Creek, CA

The Jazz whole-genome shotgun assembler was used to assemble several high-profile genomes, including *Fugu rubripes*, *Ciona intestinalis*, and *Xenopus tropicalis*. It is the in-house assembler of choice for medium-to-large genomes (> 10Mb). It uses an overlap-layout-consensus paradigm based on the Celera assembler. Unique aspects of the assembler include: early use of “rectangles”, a custom graphical viewer, a spring energy maximization model for scaffold layout, and several engineering features including a stage-based modular design that allows for easy restarting within any given stage. Recent efforts have begun to overhaul the data structures of Jazz to support novel algorithms and visualization techniques.

The Evolution of Symbiosis in the Phylum *Bacteroidetes* –Whole Genome Analysis of the Symbionts '*Candidatus Amoebophilus asiaticus*' and '*Candidatus Cardinium hertigii*'

Stephan Schmitz-Esser[1]* (schmitz-esser@microbial-ecology.net), Martha S. Hunter,[2] and Matthias Horn[1]

[1]Department of Microbial Ecology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria and [2]Department of Entomology, University of Arizona, Tucson, AZ

Symbiosis is a tremendously creative force in evolution. Prokaryotes in particular are critical partners with eukaryotes: humans rely on symbiotic bacteria in their gut and on their skin; plants depend on symbiotic bacteria and fungi, and many if not most arthropods carry intracellular bacterial symbionts, just to name a few examples. The processes and interactions underlying and shaping symbiotic associations are, however, still poorly known.

The highly diverse bacterial phylum *Bacteroidetes* contains only two major endosymbiont lineages described to date, one of them being represented by the sister taxa '*Candidatus* Amoebophilus asiaticus', a symbiont of free-living amoebae, and '*Candidatus* Cardinium hertigii', a symbiont of arthropods. Both symbionts are obligately intracellular and cannot be cultivated in cell-free media. In this project we will analyze representative strains of these two elusive and largely uncharacterized symbionts. While many intracellular symbiont genomes have been sequenced in the *Proteobacteria*, the genomes of '*Cand.* Amoebophilus asiaticus' and '*Cand.* Cardinium hertigii' will be among the first symbiont genomes in the *Bacteroidetes.* Comparative genome sequence analysis will yield the first comprehensive insights into their biology. We will be able to determine molecular interaction mechanisms between the symbionts and their respective host cells, and to compare these to mechanisms found in other intracellular bacteria.

Comparative genome analysis of '*Cand.* Amoebophilus asiaticus' and the chlamydia-related amoeba symbiont *Protochlamydia amoebophila* will help to elucidate how largely unrelated bacteria belonging to different phyla have adapted to intracellular life in the same eukaryotic host (*Acanthamoeba*). These analyses will greatly benefit from the anticipated availability of the *Acanthamoeba castellanii* genome, which is currently being sequenced at the NHGRI.

'*Cand.* Cardinium hertigii' manipulates host reproduction in three different ways, two of which, cytoplasmic incompatibility, and parthenogenesis-induction, had been exclusively attributed to *Wolbachia pipientis*, a symbiont in the *Alphaproteobacteria*. Comparative genomics of '*Cand.* Cardinium hertigii' with *Wolbachia* will shed new light on the genes necessary for reproductive host manipulation.

From a phylogenetic perspective, the availability of the '*Cand.* Amoebophilus asiaticus' and the '*Cand.* Cardinium hertigii' genome sequences will allow us to investigate the evolution of the intracellular life style within the *Bacteroidetes* and the adaptation to different eukaryotic hosts. We will also be able to analyze the process of genome reduction in this phylum – a key process during evolution of intracellular bacteria, currently almost exclusively studied using *Proteobacteria* as examples.

Genome Improvement and Finishing of Eukaryotic Genomes

J. Schmutz* (jeremy@shgc.stanford.edu), J. Grimwood, JGI-SHGC Group Members, and R. M. Myers

Joint Genome Institute, Stanford Human Genome Center, Stanford University School of Medicine, Palo Alto, CA

The Stanford Human Genome Center (SHGC) began a collaboration with the Joint Genome Institute (JGI), concentrating on finishing the DOE portion of the human genome in 1999. Since the completion of the sequencing of the human genome, the JGI has rapidly expanded its scientific goals in several DOE mission-relevant areas. At the JGI-SHGC, we have kept pace with this rapid expansion of projects and currently play key project roles in the JGI organization. Our greatest contribution to JGI science has been assessing, assembling, improving, and finishing eukaryotic whole genome shotgun (WGS) projects for which the large-scale shotgun sequence is generated at the Production Genomic Facility (JGI-PGF). We follow this by combining the draft WGS with genomic resources generated at JGI-SHGC (including BAC end sequences, genetic maps and full-length cDNA sequences) to produce an improved genomic sequence. For eukaryotic genomes important to the DOE mission, we then add further information from directed experiments to produce reference genomic sequences that are publicly available for any scientific researcher. To date we have produced improved or finished versions of eight eukaryotic genomes including: *Ostreococcus* CCE9901, *Pichia stipitis, Thalassiosira pseudonana*, *Phaeodactylum tricornutum*, *Nectria haematococca*, *Aspergillus niger*, *Mycosphaerella graminicola* and *Micromonas pusilla* NOUM17. For this calendar year we plan to release improved versions of three difficult genomes: *Populus trichocarpa*, *Chlamydomonas reinhardtii* and *Laccaria bicolor*. We are also currently working on improvement strategies for *Postia placenta*, *Trichoplax adhaerens, Phycomyces blakesleeanus* and *Aureococcus anophagefferens*. Over time we have developed and improved strategies for working with difficult to sequence genomes, polymorphic genomes and genomes that present with poor long-range contiguity in order to greatly benefit the analysis and interpretation of the genomic sequence of these organisms. At the JGI Users Meeting, we will present an overview of our improvement strategies for WGS sequenced genomes, present statuses of the genomes under our care and an overview of our other activities as part of the JGI organization.

DNA Polymerases from Thermophilic Viruses for Improved Amplification and Sequencing

Thomas Schoenfeld,[1] Vinay Dhodda,[1] Robert DiFrancesco,[1] Melodee Patterson,[1] Ronald Godiska,[1] Paul Richardson,[2] and **David Mead**[1]* (dmead@lucigen.com)

[1]Lucigen Corporation, Middleton, WI and [2]DOE Joint Genome Institute, Walnut Creek, CA

Discovery of new enzymes has relied almost exclusively on samples isolated by culturing. In the case of thermophilic phage, this approach is extremely inefficient and has limited the availability of new DNA polymerases, despite their potential utility in nucleic acid amplification and sequencing. Viruses directly isolated from two hot springs, Bear Paw at 74ºC and Octopus at 93ºC, in Yellowstone National Park were explored as sources of new enzymes. Metagenomic libraries were constructed and nearly 30 Mb of sequence was determined. These sequences were screened by similarity to known *pol* genes, resulting in the discovery of over 200 *pol* genes, 58 of which were full length. The encoded

thermostable viral DNAPs are distinct and show considerable diversity compared to known thermostable DNA polymerases.The fact that no *pol* gene was re-isolated suggests many additional genes remain to be discovered. We have expressed 10 of these genes to produce thermostable DNA polymerases. The first of these has been extensively characterized. Its thermostability allows it to survive thermocycling in PCR and DNA sequencing. Its inherent proofreading capability allows some of the highest fidelity DNA amplification. Its potent strand displacement activity and its efficient initiation at nicks allows a new approach to isothermal whole genome amplification that eliminates the need for exogenous primers, improves specificity and lowers background compared to conventional methods. Its inherent reverse transcriptase activity allows efficient single tube, single enzyme RT PCR. This enzyme has also shown improved peak resolution compared to conventional enzymes when used in DNA sequencing.

The Complete Genome Sequence of *Bifidobacterium longum* bv. *infantis*

D. A. Sela[1]* (dasela@ucdavis.edu), S. L. Freeman,[1] P. M. Richardson,[2] J. B. German,[1] and D. A. Mills[1] (damills@ucdavis.edu)

[1]University of California, Davis, CA and [2]DOE Joint Genome Institute Production Genomics Facility, Walnut Creek, CA

Bifidobacterium longum bv. *infantis* is one of the early colonizers of the breast-fed infant colon and subsequently dominates the lower gastrointestinal tract prior to weaning. We have characterized the *B. longum* bv. *infantis* isolate UCD272 which exhibits strong growth on human milk oligosaccharides (HMO) as a sole carbon source. UCD272 also exhibits fucosidase and sialidase activities which are believed to contribute to degradation of the complex HMO polymer. In collaboration with the Department of Energy's Joint Genome Institute, we have completely sequenced the genome of this microorganism. UCD272 represents the largest bifidobacterial genome reported to date (2.8 Mbp), and is predicted to contain >500 additional ORFs compared to previously sequenced genomes. Accordingly, the UCD272 genome possesses a large number of two-component regulatory systems, transport proteins, and selfish elements, by comparison to other sequenced bifidobacteria. Furthermore, the UCD272 genome sequence reveals a number of genes associated with utilization of mammalian-derived oligosaccharides and glycoconjugates which are unique to this strain. Characterization of the *B. longum* bv. *infantis* genome helps define the role for this important member of the infant gut consortium.

Metagenomic Analysis of the Mobile Gene Pool of the Activated Sludge Microcosms from Municipal Wastewater Treatment Plants

Vladimir Sentchilo* (vladimir.sentchilo@unil.ch) and Jan Roelof van der Meer (janroelof.vandermeer@unil.ch)

Department of Fundamental Microbiology, University of Lausanne, Bâtiment Biophore, 1015 Lausanne, Switzerland

The task of wastewater treatment plants (WWTPs) is to protect the aquatic environment from the otherwise harmful discharges of modern societies. Removal of chemical contaminants, nutrients and particulates relies on composition, integrity, stability and

adaptability of complex microbial communities inhabiting the different WWTP compartments.

The goal of our project is to get a concise picture of the capacity of the microbial communities in WWTPs to exchange genetic material via plasmids (i. e., the *mobilome*) and to understand the variability of the mobilome as a function of WWTP operations. Our current work is focused on the development of a robust method for the isolation of a representative pool of microbial community plasmid DNA suitable for the construction of sequencing libraries.

Aeration basins of two similarly operated WWTPs in the Lausanne region, Switzerland, were chosen as source of the activated sludge. Mechanical disintegration of sludge particles and detachment of bacterial cells was done in a kitchen cocktail blender in 12.5 mM Na-phosphate buffer (pH7.0) supplemented with 1 mM EDTA, and the reducing agents 1.4 mM 2-mercapthoethanol and 10 mM Na-ascorbate. Use of 0.1% Na-pyrophosphate, 0.05% Triton X-100 or higher than 1 mM concentrations of EDTA was found to reduce microbial diversity in the cell preparations, as judged from microscopy observations. Separation of bacterial cells from the bulk of organic matter and eukaryotic cells, and washing were done in a series of centrifugation steps. In order to minimize polysaccharide contamination, the bacterial cell pellets were further treated with a mixture of polysaccharide degrading enzymes (beta-galactosidase, alpha-amylase and pectinase) and collected by centrifugation. After lysozyme treatment a modified NaOH/SDS alkaline lysis in the presence of 2% (w/v) polyvinylpolypyrrolidone was implemented to obtain cleared cell lysates. Isopropanol precipitated plasmid DNA fractions still contained significant amounts of non DNA material and produced brightly fluorescing smears in agarose gel, even after phenol-chloroform and/or cetyltrimethylammoniumbromid treatment. Therefore, further purification of the crude plasmid DNA preparations was done by cesium chloride-ethidium bromide density gradient ultracentrifugation. Finally, plasmid DNA of acceptable purity was obtained. Agarose gel electrophoresis revealed the presence of multiple circular plasmid DNA molecules ranging from ca. 2 kb to more than 60 kb in size.

Further improvements are necessary in cell lysis efficiency and plasmid DNA yield, as well as removal of chromosomal DNA contamination in the final preparations. Finally we plan to pool together several mobilome samples from each WWTP to produce representative DNA samples which are to be fragmented, cloned, sequenced and analyzed.

High-Throughput Metagenome Assembly

Harris Shapiro (hjshapiro@lbl.gov), Serge Dusheyko, **Craig Furman*** (cfurman@lbl.gov), Jasmyn Pangilinan, and Hank Tu

DOE Joint Genome Institute, Walnut Creek, CA

Metagenome data sets present a qualitatively different assembly problem than traditional single-organism whole-genome shotgun (WGS) assembly. The unique aspects of such projects include the presence of a potentially large number of distinct organisms and their representation in the data set at widely different fractions. In addition, multiple closely related strains could be present, which would be difficult to assemble separately. Failure to take these issues into account can result in poor assemblies that either jumble together different strains or which fail to yield useful results.

The DOE Joint Genome Institute has sequenced a number of metagenomic projects and plans to considerably increase this number in the coming year. As a result, the JGI has a need for high-throughput tools and techniques for handling metagenome projects.

We present the techniques developed to handle metagenome assemblies in a high-throughput environment. This includes a streamlined assembly wrapper, based on the JGI's in-house WGS assembler, Jazz. It also includes the selection of sensible defaults targeted for metagenome data sets, as well as quality control automation for cleaning up the raw results. While analysis is ongoing, we will discuss preliminary assessments of the quality of the assembly results (http://fames.jgi-psf.org).

Eukaryotic Library Quality Control at the JGI

Harris Shapiro (hjshapiro@lbl.gov), Jeff Froula, Jasmyn Pangilinan, Hank Tu, and **Michael Zhang*** (myzhang@lbl.gov)

DOE Joint Genome Institute, Walnut Creek, CA

The DOE Joint Genome Institute presently sequences 20 – 30 eukaryotic genomes per year, ranging in size from 15 MB to more than 1 GB. Being able to produce high-quality assemblies for this number of organisms requires high-quality shotgun libraries as input. The situation is complicated by the fact that many of the organisms are relatively novel. There has been little prior characterization of their genomes, and even basic measures such as the genome size have large uncertainties associated with them. Finally, a high fraction of the JGI's eukaryotes have "bonus" prokaryotes present in their source DNA, making the early assessment of contamination levels critical.

We describe the library quality control (QC) checks that we have implemented, to try to identify potential problems as soon as possible. We also discuss the improvements to this process that have been put in place over the past year, including:

- A greatly simplified web-based interface for launching and analyzing library QCs;
- The merger of the eukaryotic, prokaryotic, and EST QC pipelines;
- The initial steps in switching from batch-based QCs to the Rolling QC system; and
- The design and implementation of a novel technique for identifying and removing vector, adapter, and linker sequences in EST libraries.

Life-Cycle of a JGI Microbial Finishing Project – LANL

David Sims* (dsims@lanl.gov), Olga Chertkov, Chris Munk, Hajnalka Kiss, Liz Saunders, Sue Thompson, Linda Meincke, and Cliff Han

Los Alamos National Laboratory, Los Alamos, NM

A broad-brush overview of the process utilized to finish a microbial genome at JGI-LANL will be presented. The presentation will include activities of informatics, wet lab and finishing team members as they interact and coordinate efforts to complete the sequencing of the microbe. Some of the accomplishments and highlights of the finishing processes at LANL will be presented. Additionally, the interaction of the LANL efforts with those of the other portions of the JGI will be described.

Functional Studies in 79 Bacterial Genomes Define Barriers to Horizontal Gene Transfer

Rotem Sorek* (RSorek@lbl.gov), Yiwen Zhu, and Eddy Rubin

DOE Joint Genome Institute, Walnut Creek, CA

Horizontal gene transfer (HGT) enables prokaryotes to acquire and use genetic material of other, sometime distantly related, microorganisms. The chronicled abundance of HGT has led to a controversy concerning whether accurate tree-like phylogenies of prokaryotes is even possible. Here we experimentally studied 264,827 genes from 79 different bacterial and archaeal genomes for their ability to be transferred to an *E. coli* recipient. We show that specific categories of genes, particularly those involved in transcription, translation and membrane biogenesis, tend to markedly inhibit the growth of the recipient when transferred into it from organisms spanning wide phylogenetic range. We further show that expression of the protein product, rather than the mere presence of the foreign DNA, is the cause for the recipient growth inhibition. For a subset of these genes, the toxic effect stems from dosage intolerance rather than incompatibility, as *E. coli* could not tolerate additional copies of such genes coming from its own genome. Our functional studies define phyla independent barriers for horizontal gene transfer.

Sequencing the Genome of the Forest Pathogen *Heterobasidion annosum*

Jan Stenlid[1]* (jan.stenlid@mykopat.slu.se), Matteo Garbelotto,[2] Ursula Kües,[3] James B. Anderson,[4] Francis Martin,[5] and Halvor Soilheim[6]

[1]Department of Forest Mycology and Pathology, Swedish University of Agrcultural Sciences, Uppsala, Sweden; [2]University of California, Berkeley; [3]Georg August University of Göttingen, Germany; [4]University of Toronto, Canada; [5]INRA-Nancy, France; and [6]Norwegian University of Life Sciences, Ääs Norway

Heterobasidion annosum causes a devastating root rot in conifer plantations and natural forests throughout the northern hemisphere. In a collaboration with the Joint Genome Institute, the genome of *H. annosum* will be the first plant pathogenic homobasidiomycete to be sequenced allowing for new insights into plant-microbe interactions. Comparisons with plant pathogens with a gradient of taxonomic relatedness to *H. annosum* will help understanding the evolution of pathogenicity factors. Response of the model tree *Populus* to various types of trophic interactions can be studied including rust pathogen fungi and mycorrhizal mutualists. Furthermore, comparisons with the model white rotter *Phanerochaete crysosporium,* will deepening our understanding of wood degradation including ligninolytic and polysaccharide degradation pathways and several bioremediation applications. Moreover, this project will also gain insights into fungal evolutionary history and biology including development, non-self recognition, mating, and secondary metabolism.

Single-Cell Genomics of Marine Bacterioplankton

R. Stepanauskas* (rstepanauskas@bigelow.org) and M. E. Sieracki

Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, Maine

We report a novel methodology for the analysis of multiple DNA loci in single marine bacteria cells, enabling high-throughput biogeochemical function assignment of yet-uncultured taxa. The protocol consists of high-speed fluorescence-activated cell sorting, whole genome multiple displacement amplification (MDA) and subsequent PCR screening. Methods were optimized to reduce DNA contamination and to maximize MDA-PCR success rate. A pilot library of eight single amplified genomes (SAGs) was constructed from Gulf of Maine bacterioplankton. The taxonomic composition of the library, determined from SSU rRNA gene sequences, implied a higher proportion of Flavobacteria in the bacterioplankton sample than that found by environmental PCR. Two out of five Flavobacteria in the SAG library contained proteorhodopsin genes, implying that Flavobacteria are among the major carriers of this novel photoautotrophy system. Alphaproteobacteria-related *puf*M and *nas*A were detected in some 100-cell aliquots but not in SAGs, demonstrating that organisms carrying bacteriochlorophyll and assimilative nitrate reductase genes constituted <1% of the sampled bacterioplankton. Compared to the metagenomics approach, the power of SAG PCR lies in the ability to assign functions to uncultured microorganisms directly, even when the functional genes are located far from taxonomic markers (e.g. SSU rRNA gene). Selected SAGs are currently undergoing whole genome sequencing

The Genome of the Arsenite Oxidizing Gammaproteobacterium *Alkalilimnicola ehrlichii*

J. F. Stolz[1]* (stolz@duq.edu), The Joint Genome Institute, R. Tabita,[2] B. Witte,[2] G. King,[3] J. M. Santini,[4] J. S. Hoeft,[5] C. Richie,[1] P. Basu,[1] and R. S. Oremland[5]

[1]Duquesne University, Pittsburgh, PA; [2]Ohio State University, Columbus, OH; [3]University of Marine, Walpole, ME; [4]University College London, London, UK; and [5]U.S. Geological Survey, Menlo Park, CA

Arsenic speciation, mobility, and toxicity can be greatly influenced by microbial activity (e.g., arsenate reduction, arsenite oxidation, methylation, demethylation). The arsenic genome project was undertaken to better understand the role of arsenic in microbial metabolism. Four organisms were proposed as candidates for sequencing: three arsenate respiring bacteria (Bacillus selenitireducens strain MLS-10, Clostridium sp. strain OhILAs, and Mono Lake strain MLMS-1), and an arsenite oxidizing bacterium (*Alkalilimnicola ehrlichii* strain MLHE-1^T). *A. ehrlichii* strain MLHE-1^T is a haloalkaliphilic gammaproteobacterium isolated from Mono Lake CA. It has the unique ability to grow both aerobically as a heterotroph and anaerobically as a chemolithoautotroph coupling arsenite oxidation to nitrate reduction. The latter ability has implications for arsenic and nitrogen cycling as well as carbon fixation in anoxic environments. The genome, recently sequenced to closure, is 3.2 MB with 2869 ORFs and a 67.5% mole GC content. Direct comparison with *Nitrococcus mobilis* and *Halorhodospira halosphila*, two other members of the Ectothiorhodospiraceae for which complete genome data is available, indicate only 57% of their genomes are shared. Analysis of the *A. ehrlichii* strain MLHE-1^T genome has identified the operons for carbon fixation (RuBisco- *cbbRLSQO*), CO oxidation (*coxFEDLSM)*, nitrate reduction

(*narLXK$_2$GHJI*), and arsenic resistance (*arsBADCR*). Remarkably, no homolog of arsenite oxidase was found, however, two arsenate reductases (*arrCAB*, *arrCBAD*) were. Arsenate reductase activity was detected by activity assays and in zymograms of cell fractions from chemoautotrophically grown cultures, however, attempts to grow the organism with arsenate as a terminal electron acceptor have been unsuccessful. Aspects of nitrate respiration are also unique. In addition to the dissimilatory nitrate reductase (Nar), homologs for nitric oxide reductase (*norDQBC*) and nitrous oxide reductase (*nosLYDZR*) were found. Nevertheless, nitrite is the end product of nitrate reduction. This result can be explained by the absence of a nitrite reductase (*nirK* or *nirS*) in the genome. Thus the genomic data has already provided valuable insight into the physiology of this unique organism and will allow for more in depth proteomic and genomic analysis.

Direct Sequencing of Large Insert Size Clones Using Templates Generated by Rolling Circle Amplification

Damon Tighe* (tighe2@llnl.gov), Nancy Hammon, Susan Lucas, and Jan-Fang Cheng

U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA

Rolling circle amplification (RCA) has been widely used in production sequencing facilities for preparing high quality sequencing templates from small insert size (3 and 8 Kb) clones. This approach, however, has not been successful in preparing sequencing templates from large insert size clones (fosmids and BACs) due to the inconsistency of generating high quality reads. In the attempt to optimize this process, we have tested several conditions including heat lysis of cells, lysis buffer, addition of DMSO, premix to cell lysate volume ratio, and cycle sequencing. We will describe in detail how the various conditions influence the quality of the reads. The results show that a 15 second heat lysis at 95C, with $MgCl_2$ and TE, 10% DMSO, 2 to 1 ratio of premix to cell lysate, and 38 cycles of sequencing reaction give the best sequencing quality. We have applied this condition to sequence 42 plates of fosmid clones derived from two libraries including the soybean and an environmental sample. We have obtained an average read length of 649 bp and a pass rate of 89%. In the initial testing, using clones from 12 different fosmid libraries, we have found that different libraries could generate very different sequence quality under the same condition. We also found that the induction of the fosmid copy number may actually lead to a decrease of sequence quality in particular fosmid libraries. The differences in sequencing quality resulted from different libraries are being investigated. We have begun to test conditions that are suitable for BAC end sequencing using the RCA templates. The preliminary data shows that it is possible to sequence up to 700 bp directly from BACs using the RCA products.

This work was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.
UCRL-ABS-228494
LBNL-62526 Abs.

Investigating the Biological Adaptations of Cord-Forming Wood Decay Fungi Utilizing Discrete Coarse Woody Litter: *Phanerochaete velutina* and *Serpula lacrymans*

Monika Tlalka,[1] Daniel Eastwood[2] (daniel.eastwood@warwick.ac.uk), Mark Fricker,[1] and **Sarah Watkinson**[1]* (sarah.watkinson@plants.ox.ac.uk)

[1]University of Oxford, Oxford, UK and [2]University of Warwick, UK

A 'foraging' mode of growth is common to cord-forming wood decay fungi (CFWDF), adapted to life as persistent mycelial networks exploiting the coarse woody debris of the forest floor as their sole carbon and energy source, while scavenging mineral nutrients opportunistically from wood, soil and exhausted mycelium. Internal source/sink-responsive long-distance nutrient translocation through cords contributes to the metabolic homeostasis required for protein synthesis at the growing mycelial margin. Cords are linear aggregates of hyphae differentiated for bidirectional mass flow, which form as nutrient channels to provision the advancing mycelial front, and retrieve resources from exhausted regions. The entire system continually remodels itself as it colonises successive stochastically-distributed woody resources, generating a mycelial network which can be several metres in extent.

Key to the biology of CFWDF is the ability to regulate metabolism, gene expression and development in response to local extracellular and intracellular nutrients.

We currently use realistic, spatially heterogeneous microcosms to image nutrient flows, and to investigate metabolic profile and gene expression in the CFWDF *Phanerochaete velutina*.

Serpula lacrymans (Coniophoraceae; Boletales; Basidiomycota; Fungi) is a destructive brown rot wood-decaying fungus, pernicious in buildings because it destroys the strength of wood by uniquely aggressive cellulolysis, and because it can spread across non-nutrient surfaces by translocating nutrients from the colonized wood base to the advancing hyphal tips.

Preliminary imaging experiments in *S. lacrymans* show rapid translocation and accumulation of nutrients into the N-starved parts of the colonies. The *S. lacrymans* whole genome sequence - to be undertaken by JGI - will make the biology of CFWDF accessible by molecular technologies, including transcriptomics and metabolomics, gene substitutions, live cell imaging with reporter constructs, and bioinformatics. These approaches will elucidate the enzymology and regulation of the processes of brown rot decay, and the spatial regulation of metabolism and development underlying the colonization of particulate cellulosic substrates. The knowledge gained might be exploited in developing controlled energy release processes for bioenergy from wood chips; for predictive models of the role of brown rot CFWDF in boreal forest carbon sequestration; and for targeted control of timber degradation.

References

1. Tlalka, M. et al., (2002). Continuous imaging of amino-acid translocation in intact mycelia of *Phanerochaete velutina* reveals rapid, pulsatile fluxes. *New Phytologist* **153**, 173–184: {http://www.blackwell-synergy.com/links/doi/10.1046/j.0028-646X.2001.00288.x/abs/}
2. Videos:{http://www.blackwellpublishing.com/products/journals/suppmat/NPH/NPH288/NPH288sm.htm}
3. Kauserud, H. et al., (2004). Molecular phylogenetics suggest a North American link between the anthropogenic dry rot fungus *Serpula lacrymans* and its wild relative

S.himantioides. *Molec. Ecol.* **13**, 3137-46. {http://www.blackwell-synergy.com/links/doi/10.1111/j.1365-294X.2004.02307.x/pdf}

Creation of a Linkage Map for Marker-Trait Association in Switchgrass Using EST-Sequence Data

Christian Tobias[1]* (ctobias@pw.usda.gov), John Vogel,[1] Paul Twigg,[3] Olin Anderson,[1] Ken Vogel,[2] and Gautam Sarath[2]

[1]Genomics and Gene Discovery Reasearch Unit, Western Regional Research Center, USDA Agricultural Research Service, Albany, CA; [2]Grain, Forage and Bioenergy Research Unit, USDA Agricultural Research Service, Lincoln, NE; and [3]Department of Biology, University of Nebraska, Kearney, NE

Availability of a linkage map is fundamental for advanced breeding strategies and molecular markers are indispensable tools for modern agriculture. Limited mapping efforts utilizing RFLPs have led to the publication of a partial genetic linkage map (Missaoui et al. 2005), but this map is insufficient for marker assisted selection and the identification of loci associated with agronomic traits (QTLs) was not attempted. We are currently creating linkage maps in switchgrass and are in the process of marker development. Many switchgrass-derived sequences available in the near future will aid in this endeavor. The DOE-Joint Genome Institute is currently undertaking a large-scale EST-sequencing project in switchgrass that is being led by the USDA, ARS. This sequencing project, existing resources from related grass species, and resources we plan on developing will provide ample sequences from gene regions and intervening sequences that will form the basis for marker development.

In order to develop a saturated genetic map two tetraploid mapping populations designated ALB and UGA are available. We are currently focusing our efforts on the ALB population that was produced from a randomly selected “Kanlow” individual (K5) as the female parent and an “Alamo” individual (A4) with good response to tissue culture and transformability. These parental genotypes arise originally from Oklahoma (35ºN lat) and Texas (28ºN lat) respectively and from different USDA hardiness zones. Seed was collected from the Kanlow parent of which 321 individual F1 were germinated. From these, 260 have been partially genotyped and verified as true crosses while 10 have been identified as self pollinated based on genotyping results with EST-SSRs. The entire population is maintained in greenhouse facilities at Albany and has been propagated to a perennial nursery at the ARS Coastal Plains Experiment Station in Tifton, GA.

Several basic questions that have not been adequately addressed in this complex polyploid are being answered definitively by this research. Foremost, this research is determining if switchgrass behaves as an autopolyploid, allopolyploid, or mixed polyploid by examining the degree of preferential pairing and possibility of double reduction at many different loci using co-dominant markers containing different alleles in each parent. Once answered, the best genetic models for construction of a linkage map and QTL analysis can be determined. Analysis of marker segregation data will provide several ways of estimating genome length, degree of recombination in each parent and reveal patterns of segregation distortion and marker distribution. Major genome rearrangements between switchgrass, and other members of the subfamily *Panicoideae* will be revealed.

"Cloneview": An Assembly Viewer for Microbial Genome Finishing

S. Trong[1]* (trong1@llnl.gov), E. Goltsman,[1] S. Malfatti,[2] P. Chain,[2] and A. L. Lapidus[1]

[1]DOE Joint Genome Institute Production Genomics Facility, Walnut Creek, CA and [2]Lawrence Livermore National Laboratory, Livermore, CA

The process of improving and completing (Finishing) draft genome assemblies using whole genome shotgun approach involves many complex iterative steps. The effectiveness of this process depends on many factors, such as the complexity of the genome, the quality of the DNA sequence and the accuracy of the assembled data. Although many efforts have been attempted to automate Finishing, it is still a relatively manual process. One such effort involves the identification and correction of regions that are incorrectly pieced together by the assembly program. To aid the finishers in visualizing and identifying these problematic areas, we have developed a visualization tool called Cloneview for displaying mate pair information in an assembly.

Our goal was to provide a graphical tool for displaying assembly information in a way that could be used to easily identify misassembled regions, repeat regions and possible biases in library creation.

Cloneview's ability to highlight read pairs that are inconsistently placed in the assembly allows the finishers to easily detect regions that are misassembled. By extracting the names of the reads to a file, further processing to correct the misjoined areas can be achieved with minimal effort. Other features of Cloneview include displaying read and clone depth, read pairs that span gaps, library-specific reads and repeat regions. In addition to the viewer, we have developed an accompanying software program to detect and report misassembled regions by looking for areas where violations in mate pairing predominately outweigh valid ones. This tool along with the viewer further enhances the finishers' ability to resolve misassemblies more efficiently.

The Microbial Finishing groups at the JGI /PGF and JGI/LLNL have incorporated these tools into their pipeline as part of an effort to rapidly finish microbial genomes.

This work was performed under the auspices of the U.S. DOE of Science, Biological and Environmental Research Program, and by the University of California, LLNL under Contract No. W-7405-Eng-48, LBNL under Contract No. DE-AC02-05CH11231 and LANL under Contract No. W-7405-ENG-36.

Comparative Community Genomics of Infant and Mother Gut Microbiota

Parag A. Vaishampayan[1]* (pavaishampayan@lbl.gov), Jeff L. Froula,[1] Jenna Morgan,[1] Howard Ochman,[2] and M. Pilar Francino[1]

[1]Evolutionary Genomics Program, DOE Joint Genome Institute, Walnut Creek, CA and [2]Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, AZ

We are using a comprehensive metagenomic approach to investigate the comparative genomics of the gastrointestinal microbiota of a human infant and mother pair. Fosmid libraries were constructed with bacterial genomic DNA isolated directly from fecal samples of infant and mother at two different time periods, one month and 11 months after delivery. The four fosmid libraries in total cover 8000 Mbp of community DNA. In a

fosmid end sequencing approach including 12288 sequence tags, we have shown parallel shifts in the gut microbiota of infant and mother. The microbiota in the one-month old infant is dominated by *Bacteroides* sp. and *Escherichia coli*, while the eleven-month old infant sample exhibits a dominating *Bifidobacterium* sp. population. The maternal gut flora also exhibits a *Bacteroides* to *Bifidobacterium* shift between one and 11 months after delivery. We have also examined 3072 bacterial 16S ribosomal RNA sequences from fecal samples of infant and mother at the two time intervals, as well as *Bacteroides* and *Bifidobacterium* genus-specific 16S sequences. The 16S studies demonstrate that the maternal and infant gut microbiota at one month after delivery exhibit identical phylotypes of the dominating *Bacteroides* and *E. coli* populations, clearly indicating maternal transmission. Fosmid sequencing is underway to confirm strain identity between mother and infant. On the other hand, eleven months after delivery maternal and infant gut microbiota contain different strains of *Bifidobacterium,* suggesting independent acquisition and/or development of *Bifidobacterium* populations in the two individuals.

Construction of Normalized cDNA Libraries at JGI

Mei Wang* (mwang@lbl.gov), Dean Ng, Edward Kirton, Erika Lindquist (EALindquist@lbl.gov), and Paul Richardson

U.S. DOE Joint Genome Institute, Walnut Creek, CA

The JGI makes cDNA libraries from mRNA for all Eukaryotic whole genome shotgun projects to aid in gene discovery and annotation. cDNA and Expressed Sequence Tag (EST) sequencing are important components in identifying gene sequences. EST sequences are used in genome annotation, specifically for training of gene calling algorithms and validation of gene models. Furthermore, JGI also generates ESTs for organisms without sequenced genomes to provide transcriptome information to the user community. Additionally, the cDNA library clones are a valuable resource. We are continually working to improve our cDNA process and have recently instituted normalized library production for some projects.

Given that transcript abundance varies from 1 to 200,000 copies, direct sequencing of cDNA clones from standard libraries may result in the repeated sequencing of abundant transcripts. Normalization dramatically decreases the prevalence of such abundant transcripts and increases the number of unique, low-abundance sequences.

To construct libraries with reduced representation of highly abundant clones, we adapted the normalization method of Zhulidov *et al.* (2004), based on a duplex-specific nuclease (DSN). The first step in normalization involves the denaturation and subsequent renaturation of cDNA. Due to hybridization kinetics, highly represented sequences reassociate more efficiently into a double-stranded form and are degraded by DSN. The normalized single-stranded DNA fraction is amplified, ligated and transformed to generate a normalized cDNA library.

In our first experiment, we evaluated the normalization method using a human placenta RNA sample and observed a high degree of diversity in sequenced clones with few cDNA sequences repeated. Next, we applied this methodology to construct a normalized cDNA library from porcelain crab RNA. Clones were randomly picked and sequenced. Sequence analysis revealed that only 0.26% of them were insertless and 95% of them were unique sequences. This is in contrast to a standard cDNA library initially generated from the same RNA where 35% of the clones represented the same gene. Because the normalized library showed a marked improvement in diversity over the standard cDNA library, more clones

are being sequenced from the normalized library which should generate many more unique sequences than our standard process.

1. Zhulidov P.A., Bogdanova E.A., Shcheglov A.S., Vagner L.L., Khaspekov G.L., Kozhemyako V.B., Matz M.V., Meleshkevitch E., Moroz L.L., Lukyanov S.A., Shagin D.A. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acid Res., 2004, 32: e37.

This work was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. DE-AC52-06NA25396. UCRL-ABS-228595

Metagenomics Implicates Termite Hindgut Bacteria in Lignocellulose Hydrolysis

Falk Warnecke[1]* (FWarnecke@lbl.gov), Peter Luginbühl,[2] Natalia Ivanova,[1] Majid Ghassemian,[2] Toby Richardson,[2] Rotem Sorek,[1] Susannah G. Tringe,[1] Justin Stege,[2] Gordana Djordjevic,[2] Mircea Podar,[3] Hector Garcia Martin,[1] Victor Kunin,[1] Alice C. McHardy,[5] Kerrie Barry,[1] Asaf Salamov,[1] Daniel Dalevi,[1] Julita Madejska,[1] Edward Kirton,[1] Nikos Kyrpides,[1] Eric Matson,[4] Elizabeth Ottesen,[5] Xinning Zhang,[4] Mirian Hernandez,[6] Catalina Murillo Cruz,[6] Luis Guillermo,[6] Isidore Rigoutsos,[5] Giselle Tamayo,[6] Brian Green,[2] Cathy Chang,[2] Eddy Rubin,[1] Dan Robertson,[2] Eric Mathur,[2] Philip Hugenholtz,[1] and Jared R. Leadbetter[4]

[1]DOE Joint Genome Institute, Walnut Creek, CA; [2]Diversa Corporation, San Diego, CA; [3]Biology Department, Portland State University, Portland, Oregon; [4]Environmental Science and Engineering and [5]Biology, California Institute of Technology, Pasadena, CA; [6]IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY; [7]INBio, Instituto Nacional de Biodiversidad, Santo Domingo de Heredia, Costa Rica

The gut system of termites is a tiny though highly efficient bioreactor for the breakdown of recalcitrant plant material and has attracted the attention of microbiologists for about 150 years. Most research has centered on lower termites which harbor protozoans thought to be primarily responsible for lignocellulose conversion, while bacteria are generally considered not to be involved. Therefore, it is a matter of some contention as to who is responsible for lignocellulose hydrolysis in higher termite species which lack protozoan symbionts. We performed a meta-genomic analysis of a higher termite hindgut community to determine the genes and pathways responsible for lignocelluose hydrolysis and to identify the species responsible for this process. We obtained ~100 Mbp of sequence information which were assembled and annotated using IMG/M, an experimental system for the comparative analysis of metagenomic data. We discovered an over-representation of bacterial enzymes involved in lignocellulose hydrolysis. This shows for the first time the importance of bacteria in this critical step. Moreover and through an accompanying 16S rRNA survey we could confirm the presence of a group of bacteria distantly related (~80% similarity on 16S rRNA sequence) to *Fibrobacter succinogenes,* a species responsible for cellulose hydrolysis in ruminants. We present the first genomic data for this group and hypothesize that they may at least partially fulfill the functional role of the protozoans present in lower termites. Knowledge gained through this analysis may improve industrial processes for the conversion of cellulosic biomass to biofuels such as ethanol and hydrogen.

An Overview of the Joint Genome Institute Production Sequencing Line

Chris Wong*, Miranda Harmon-Smith, Kathleen Lail, Chris Daum, and Lena Philip

DOE Joint Genome Institute, Walnut Creek, CA

The Department of Energy's (DOE) Joint Genome Institute (JGI) Production Genomics Facility (PGF) performs high-throughput Sanger sequencing using whole genome shotgun sequencing methods. The goal of the Production Line is to produce high quality sequence data in order to allow the downstream assembly of whole genome, BAC and metagenomic projects. The Production Line generates end sequence from three different size DNA fragments: 3, 8, and 40 kb, to a depth of 8.5x coverage.

The production line is comprised of three subgroups: Library Support, Sequencing Prep and Capillary Electrophoresis, which consist of about 35 people and perform its work within roughly 12,000 square feet of laboratory space. The production line uses 70 Applied Biosystem 3730xl DNA sequencers and 36 MegaBACE 4500 DNA sequencers to generate approximately 90 Megabases a day. The Applied Biosystem 3730xl sequencers are operated 24 hours a day, 365 days a year, while the MegaBACE 4500 sequencers are run approximately five times a day, five days a week.

For fiscal year 2006, the Production line produced 49.4 million lanes and 32.6 gigabases of sequence with an average read length of 647 bases per lane, sequencing approximately 200 projects. This poster will describe the infrastructure, process steps and quality control methods to ensure the production of high quality sequence.

This work was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.
UCRL-ABS-228492
LBNL-62525 Abs.

Whole Genome Sequencing of *Micromonas pusilla*, a Ubiquitous Picoeukaryotic Phytoplankter

A. Z. Worden[1]* (aworden@rsmas.miami.edu), S. McDonald,[1] P. von Dassow,[2] F. Not,[3] and M. L. Cuvelier[1]

[1]Rosenstiel School of Marine and Atmospheric Science, University of Miami, Coral Gables, Florida; [2]University of Washington, Seattle, WA; and [3]Institut de Ciències del Mar, CMIMA, Barcelona, Spain

Micromonas pusilla is a marine photosynthetic picoeukaryote (<2-3 μm) first reported to be highly abundant in the 1950s. Nevertheless, only recently has follow-up work addressed the ubiquity and abundance of this prasinophyte in the world's oceans. Whole genome shotgun sequencing helped reveal uncultured clades present in the Sargasso Sea, a region where *Micromonas* dynamics have not previously been investigated or considered important. More recently we have found that coexisting populations of *Micromonas* and its relative *Ostreococcus*, the smallest free-living eukaryote known, can be responsible for a large fraction of picoplanktonic carbon production in the Sargasso Sea. Furthermore, a recently published study (2007) demonstrated that carbon from this size fraction is efficiently sequestered to the deep ocean. This contrasts with the 20 year-standing

paradigm, which accepted the hypothesis that picoplankton did not play a major role in the flux of carbon to the deep ocean. Because of the recognized importance of *Micromonas* two phylogenetically distinct strains, isolated from different ocean basins, were selected for genome sequencing. *M. pusilla* belongs to the order Mamiellales, within the Prasinophyceae. These are early diverging green algae and, as such, are a key to studying the "green" lineage in addition to their ecological roles in modern oceans.

Micromonas NOUM17, an equatorial Pacific isolate, has a 24 Mb genome organized in 17 chromosomes, with 8007 predicted genes. Many aspects of the genome are being investigated in the manual annotation phase, which has just been initiated. Preliminary work has begun to reveal a variety of evolutionary developments in this organism. For example, prasinophytes have long been thought to harbor a unique light harvesting complex servicing both PSI and PSII, which was thought to represent the ancestral state. This is not the case for *Micromonas*, nor *Ostreococcus*, and has led to the refutation of previous work on the topic. *Micromonas* has a greater number of light harvesting proteins than *Ostreococcus* and the expression of these genes is being explored. Of the 97 flagellar proteins identified in *Chlamydomonas reinhardtii* by mass spectrometry (Pazour et al 2005), 49 have been identified thus far within NOUM17. The outer arm flagellar proteins are composed of 4 dynein heavy chains, 2 intermediate chains and 8 light chains, a putative blue-light receptor (potentially involved in the phototactic motion of this organism) has also been identified. Both the *Ostreococcus* and *Micromonas* genomes encode essential machinery for meiotic recombination, including Spo11, DMC1, Rad50, and Mre11. This strongly suggests that these picoeukaryotic cells have retained a sexual phase in their life cycle, even though this has not yet been observed in nature or in culture. Lastly, duplicate versions of a number of genes, with one being bacterially derived and the other more akin to the green lineage, have been identified. Draft sequencing of the second strain, isolated from the North Atlantic is currently underway. Comparative genomics work will be used to explore the functional diversification of particular genes and/or gene families. Furthermore, phylogenomic approaches will be used to aid understanding of how genomes are shaped by ecological and environmental divergence.

Random-Shear BAC Library Construction and Efficient Genome Gap Closing

Chengcang Wu, Sarah Vande Zande, Rebecaa Hochstein, **David Mead***
(dmead@lucigen.com), and Ronald Godiska

Lucigen Corporation, Middleton, WI

Bacterial artificial chromosome (BAC) libraries are indispensable for physical mapping, positional cloning, genetic analysis, and sequencing of large genomes. BAC libraries have been created from many species, including *Arabidopsis*, *Drosophila*, rice, mouse, and human. A significant limitation of the current methods is the use of partial restriction digestion to generate genomic DNA fragments of 100-300 kb. The inherently skewed genome distribution of restriction sites causes at least 10-fold under- or over-representation of particular sequences, with some regions being entirely absent from the BAC libraries. Another drawback is the instability of inserts in current cloning vectors due to transcription and secondary structure formation. As a result, existing BAC libraries built with conventional methods and vectors are biased, and numerous gaps exist in all of the physical and sequencing maps of eukaryotic multi-cellulargenomes. To circumvent these problems we have successfully developed techniques to construct unbiased, randomly-sheared BAC libraries (>100 kb inserts). We have demonstrated that a single 5X random

shear BAC library covers various genome regions uniformly and closes several gaps in the *Arabidopsis thaliana* genome. We believe it will be possible to finish the physical mapping and sequencing of *Drosophila*, *Arabidopsis*, rice, mouse, and human with this approach, closing all of the existing genomic gaps, including centromeres. We have also developed transcription-free BAC vectors. These vectors show much higher stability of inserts containing AT-rich sequences, direct and inverted repeats, and other deleterious DNAs. It is thus possible for the first time to construct unbiased BAC libraries to achieve complete closure of a large complex genome.

Tree-Based Small Subunit rRNA Taxonomy Assigning Pipeline (STAP)

Dongying Wu* (dongyingwu1@gmail.com) and Jonathan Eisen

Genome Center, University of California, Davis, California

The comparative analysis of ribosomal RNA sequences is one of the major and most powerful approaches to study phylogenetic relations among living organisms. In order to adept phylogenetic methods for the classification of Small Subunit rRNA (ss-rRNA) sequence to study the diversity of microorganism, we've developed a pipeline with the capacity of automatically identify the taxonomies of environmental ss-rRNAs based on phylogenetic trees. Small subunit rRNAs from PCR of environmental samples or metagenomic shotgun sequences are searched against our ss-rRNA database by BLASTN. The ss-rRNA database is a non-redundant subset the *Greengenes Project with JGI's taxonomy annotations*. The pipeline subsequently builds phylogenetic trees from the BLASTN results. The taxonomy assignments are based upon either neighbor-join or maximum likelihood trees rather than top BLASTN hits, thus are more reliable and accurate. The pipeline combines publicly available packages such as PHYLIP, PHYML, BLASTN and CLUSTALW with our in-house automatic alignment masking and tree parsing programs; the automation stays faithful to the manual phylogenetic tree analysis processes with the speed and capacity that are inapproachable for any manual efforts.

Community-Involved Genome Annotation and Analysis at JGI-LANL: Facilitating Publication of Genome Papers through Bioinformatics Support and Training

Gary Xie* (xie@lanl.gov), Jean Challacombe, Diego Martinez, Ravi Barabote, Monica Misra, and Thomas Brettin

JGI-Los Alamos National Laboratory, Los Alamos, NM

The role of the genome annotation and analysis team at JGI-LANL is to facilitate publication of JGI genome papers and provide bioinformatics support and training to promote community-involved genome annotation and analysis. Since 2005, we have conducted on-site week long bioinformatic training sessions for 8 JGI collaborators and hosted 23 JGI collaborators as part of our genome explorer seminar series. We also received 16 special requests from JGI collaborator labs for conducting specialize analysis. In projects where JGI-LANL team members played a leading role in the analysis and preparation of genome papers, 2 genome papers have been published, 2 papers have been accepted and 4 are in preparation. In addition to our microbial genome effort, our eukaryotic genome annotation team is working with the annotation team at JGI-PGF and

fungal research community to provide high quality manually curated annotations of fungal genomic sequences. We have hosted 2 off-site annotation Jamborees to promote community involved analysis for publishing a full analysis and annotation of *Trichoderma reesei* and *Aspergillus niger* chromosomes. Work is under way to publish its genome papers.

Predicting Structure and Function for Novel Proteins of the Iron Oxidizing Bacterium, *Leptospirillum* Group II, from a Natural Acidophilic Community

Adam Zemla,[1] Steven Singer,[1] Jason Raymond,[1] Korin Wheeler,[1] Daniela Goltsman,[2] Vincent Denef,[2] Jill Banfield,[2] and **Michael P. Thelen**[1]* (mthelen@llnl.gov)

[1]Biosciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA and [2]Department of Earth and Planetary Sciences, University of California, Berkeley

Proteins isolated from uncultivated microbial populations represent critical components of microbial processes and community fitness under natural conditions. However, a formidable obstacle to investigations enabled by genome sequence information is the high proportion of proteins of unknown function. We have identified hundreds of these novel (hypothetical) proteins produced by a natural microbial biofilm[1]. This low-diversity, chemoautotrophic community actively oxidizes iron and acidifies the local environment, generating toxic acid mine drainage. An extensive genomic and proteomic foundation has been established for this community, and we have focused directly on the problem of novel proteins within this well-defined system.

The biofilm's dominant member, *Leptospirillum* group II, produces many proteins of unknown function at detectable levels. In an extracellular fraction, up to 85% of the abundant proteins are novel. A combination of chromatographic techniques, N-terminal sequencing, mass spectrometry and computational analyses has been used to isolate, characterize, and assign functions to these novel proteins.

Structure predictions are greatly facilitated by identifying an evolutionary related protein with a known structure. Homology searches were carried out for a selected set of 421 novel *Leptospirillum* proteins using an initial version of our protein structure modeling system (AS2TS)[2]; we then assigned structural templates to 260 proteins. Both the accuracy of calculated alignments and the sensitivity in detection of potential structural homologs were subsequently enhanced by incorporating three additional modules to PSI-BLAST distant homology searches: an intermediate sequence search, automated construction of local libraries, and automated constructions of local 3D models created for identified intermediate sequences. This approach resulted in increasing the number of modeled proteins to 360. The number of PDB templates assigned to each target protein also increased, resulting in a greater chance that some of assigned PDB template structures contain sufficient functional information to predict biochemical properties and propose experiments for validation.

References

1. Ram et al, 2005, *Science* 308:1915-20, "Community Proteomics of a Natural Microbial Biofilm"
2. Zemla et al, 2005, *Nucleic Acids Res* 33 (Web Server issue):W111-5, "AS2TS system for protein structure modeling and analysis"

Attendees

Current as of March 9, 2007

William S. Adney
National Renewable Energy Laboratory
bill_adney@nrel.gov

Andrea Aerts
DOE Joint Genome Institute
alaerts@lbl.gov

Kristin K. Ahrens
University of California, Davis
kkahrens@ucdavis.edu

Martin Allgaier
University of California, San Francisco
mar.allgaier@web.de

Eric Altermann
AgResearch Limited
eric.altermann@agresearch.co.nz

Andrew J. Alverson
Indiana University
alversoa@indiana.edu

Olin D. Anderson
USDA Agricultural Research Service
oandersn@pw.usda.gov

Carl W. Anderson
Brookhaven National Laboratory
cwa@bnl.gov

Olin D. Anderson
USDA Agricultural Research Service
oandersn@pw.usda.gov

Iain J. Anderson
DOE Joint Genome Institute
IJAnderson@lbl.gov

John M. Archibald
Dalhousie University
jmarchib@dal.ca

Christopher E. Bagwell
Savannah River National Laboratory
christopher.bagwell@srnl.doe.gov

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

Ryan J. Baker
Researcher
ntety@yahoo.com

Brett Baker
University of California, Berkeley
bbaker@eps.berkeley.edu

Miriam Barlow
University of California, Merced
mbarlow@ucmerced.edu

Kerrie Barry
DOE Joint Genome Institute
kwbarry@lbl.gov

Diane M. Bauer
DOE Joint Genome Institute
bauer20@llnl.gov

Sebastian F. Behrens
Stanford University
sbehrens@stanford.edu

Harry R. Beller
Lawrence Livermore National Laboratory
beller2@llnl.gov

Randy M. Berka
Novozymes, Inc.
ramb@novozymes.com

Stephanie M. Bernard
Lawrence Berkeley National Laboratory
SMBernard@lbl.gov

Dianna L. Berry
Stony Brook University, Southampton
diberry@optonline.net

Matthew J. Blow
Lawrence Berkeley National Laboratory
mjblow@lbl.gov

Harvey Bolton, Jr.
Pacific Northwest National Laboratory
harvey.bolton@pnl.gov

Jeffrey L. Boore
SymBio
JLBoore@Berkeley.edu

Jennifer Bownas
Oak Ridge National Laboratory
bownasjl@ornl.gov

Lambert Brau
Murdoch University
lbrau@murdoch.edu.au

Jim Bristow
DOE Joint Genome Institute
JBristow@lbl.gov

Eoin L. Brodie
Lawrence Berkeley National Laboratory
elbrodie@lbl.gov

David C. Bruce
DOE Joint Genome Institute
dbruce@lanl.gov

Kathy G. Byrne-Bailey
University of California, Berkeley
kbyrne@nature.berkeley.edu

W. Zacheus Cande
University of California, Berkeley
zcande@berkeley.edu

Charles H. Cannon
Texas Tech University
chuck.cannon@ttu.edu

John E. Carlson
Penn State University
jec16@psu.edu

Grace D. Catino
DOE Joint Genome Institute
gdcatino@lbl.gov

Patrick Chain
Lawrence Livermore National Laboratory
chain2@llnl.gov

Anu Chakicherla
Lawrence Livermore National Laboratory
chakicherla1@llnl.gov

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

Feng Chen
DOE Joint Genome Institute
fchen@lbl.gov

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

Chi-Hing C. Cheng
University of Illinois, Urbana-Champaign
c-cheng@uiuc.edu

Olga Chertkov
Los Alamos National Laboratory
ochrtkv@lanl.gov

Ludmila Chistoserdova
University of Washington
milachis@u.washington.edu

Dylan Chivian
Lawrence Berkeley National Laboratory
DCChivian@lbl.gov

Cindy Y. Choi
DOE Joint Genome Institute
cchoi@lbl.gov

Attendees

Alice C. Churchill
Cornell University
acc7@cornell.edu

Collin J. Closek
University of California, Merced
cclosek@ucmerced.edu

Alicia Clum
DOE Joint Genome Institute
aclum@lbl.gov

Matthew Coleman
Lawrence Livermore National Laboratory
coleman16@llnl.gov

Maureen L. Coleman
Massachusetts Institute of Technology
colemanm@mit.edu

Chad A. Crain
University of California, Davis
cacrain@ucdavis.edu

Patrik D'haeseleer
Lawrence Livermore National Laboratory
patrikd@llnl.gov

Eileen Dalin
DOE Joint Genome Institute
e_dalin@lbl.gov

Chris Daum
DOE Joint Genome Institute
daum1@llnl.gov

Scott C. Dawson
University of California, Davis
scdawson@ucdavis.edu

Paramvir S. Dehal
Lawrence Berkeley National Laboratory
psdehal@lbl.gov

Vincent J. Denef
University of California, Berkeley
vdenef@berkeley.edu

Michael K. DeSalvo
University of California, Merced
mdesalvo@ucmerced.edu

Chris Detter
JGI – Los Alamos National Laboratory
cdetter@lanl.gov

Gregory J. Dick
University of California, Berkeley
gdick@berkeley.edu

Norman A. Doggett
Los Alamos National Laboratory
doggett@lanl.gov

Daniel W. Drell
OBER, U.S. Department of Energy
daniel.drell@science.doe.gov

Inna Dubchak
Lawrence Berkeley National Laboratory
ildubchak@lbl.gov

Garry A. Duncan
Nebraska Wesleyan University
gduncan@nebrwesleyan.edu

Peter Dunfield
Institute of Geological and Nuclear Sciences
p.dunfield@gns.cri.nz

David D. Dunigan
University of Nebraska, Lincoln
ddunigan2@unl.edu

Serge Dusheyko
DOE Joint Genome Institute
sdusheyko@lbl.gov

Elizabeth A. Edwards
University of Toronto
edwards@chem-eng.utoronto.ca

Jonathan A. Eisen
University of California, Davis
and DOE Joint Genome Institute
jaeisen@ucdavis.edu

David Emerson
American Type Culture Collection
demerson@gmu.edu

Joni B. Fazo
JGI – Lawrence Livermore National Laboratory
jbfazo@lbl.gov

Helene Feil
University of California, Berkeley
bhfeil@nature.berkeley.edu

William S. Feil
University of California, Berkeley
bhfeil@uclink4.berkeley.edu

Marsha Fenner
DOE Joint Genome Institute
MWFenner@lbl.gov

Judith Flanagan
University of California, San Francisco
flanaganj@anesthesia.ucsf.edu

Gilberto E. Flores
Portland State University
floresg@pdx.edu

Brian Foster
DOE Joint Genome Institute
bfoster@lbl.gov

Samuel E. Fox
Oregon State University
foxsa@onid.orst.edu

Jim K. Fredrickson
Pacific Northwest National Laboratory
jim.fredrickson@pnl.gov

Lillian K. Fritz-Laylin
University of California, Berkeley
fritz-laylin@berkeley.edu

Jeff L. Froula
DOE Joint Genome Institute
jlfroula@lbl.gov

Susan Fuerstenberg
Genome Project Solutions
sifuerst@genomeprojects
olutions.com

Craig Furman
DOE Joint Genome Institute
cfurman@lbl.gov

Scott Geib
Penn State University
smg283@psu.edu

Cheol-Min Ghim
Lawrence Livermore National Laboratory
cmghim@llnl.gov

David E. Gilbert
DOE Joint Genome Institute
gilbert21@llnl.gov

Tijana Glavina del Rio
Joint Genome Institute
glavinadelrio1@llnl.gov

Ivan Godinez
University of California, Davis
igodinez@ucdavis.edu

Eugene Goltsman
DOE Joint Genome Institute
egoltsman@lbl.gov

Elaine Gong
DOE Joint Genome Institute
ELGong@lbl.gov

Stephen B. Goodwin
USDA Agricultural Research Service
Purdue University
sgoodwin@purdue.edu

Lynne A. Goodwin
Los Alamos National Laboratory
lynneg@lanl.gov

Laurie Gordon
University of California
and Lawrence Livermore National Laboratory
gordon2@llnl.gov

Charles L. Greenblatt
Hebrew University
greenbl@cc.huji.ac.il

Annette M. Greiner
DOE Joint Genome Institute
amgreiner@lbl.gov

Igor V. Grigoriev
DOE Joint Genome Institute
ivgrigoriev@lbl.gov

Jane Grimwood
Stanford Human Genome Center
jane@shgc.stanford.edu

Yalong Guo
Max Planck Institute for Developmental Biology
ya-long.guo@tuebingen.mpg.de

Romey Haberle
University of California, Davis
rchaberle@ucdavis.edu

Christopher A. Hack
DOE Joint Genome Institute
cahack@lbl.gov

Masood Hadi
Sandia National Laboratories
mzhadi@sandia.gov

Natsuko Hamamura
Portland State University
nhama@pdx.edu

Nancy M. Hammon
DOE Joint Genome Institute
nmhammon@lbl.gov

James K. Han
DOE Joint Genome Institute
jkhan@lbl.gov

Cliff S. Han
Los Alamos National Laboratory
han_cliff@lanl.gov

Miranda L. Harmon-Smith
DOE Joint Genome Institute
harmonsmith2@llnl.gov

Kirk Harris
University of Colorado
jjharris@colorado.edu

Amber L. Hartman
University of California, Davis
alhartman@ucdavis.edu

Caroline S. Harwood
University of Washington
csh5@u.washington.edu

Samuel P. Hazen
Scripps Research Institute
hazen@scripps.edu

Dennis Hedgecock
University of Southern California
dhedge@usc.edu

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

Chris L. Hemme
University of Oklahoma
hemmecl@ou.edu

Daniel Herlemann
Max Planck Institute for Terrestrial Microbiology
herleman@mpi-marburg.mpg.de

David W. Hillman
DOE Joint Genome Institute
hillman3@llnl.gov

Ann M. Hirsch
University of California, Los Angeles
ahirsch@ucla.edu

Isaac Ho
DOE Joint Genome Institute
iyho@lbl.gov

Kelli Hoover
Penn State University
kxh25@psu.edu

Leila A. Hornick
DOE Joint Genome Institute
lahornick@lbl.gov

Michael J. Hornsby
University of California, Davis
mjhornsby@ucdavis.edu

Ping Hu
Lawrence Berkeley National Laboratory
phu@lbl.gov

Philip Hugenholtz
DOE Joint Genome Institute
phugenholtz@gmail.com

Karla M. Ikeda
DOE Joint Genome Institute
kmikeda@lbl.gov

Wakako Ikeda-Ohtsubo
Max Planck Institute for Terrestrial Microbiology
ikeda@staff.uni-marburg.de

William P. Inskeep
Montana State University
binskeep@montana.edu

Chavonda J. Jacobs-Young
U.S. Department of Agriculture
cjacobs@csrees.usda.gov

Janet K. Jansson
Swedish University of Agricultural Sciences
janet.jansson@mikrob.slu.se

Paul R. Jensen
Scripps Institution of Oceanography, University of California, San Diego
pjensen@ucsd.edu

Yongqin Jiao
Lawrence Livermore National Laboratory
jiao1@llnl.gov

David R. Johnson
University of California
daverj@ce.berkeley.edu

Kou-San Ju
University of California
kju@ucdavis.edu

Patricia I. Kale
DOE Joint Genome Institute
kale1@llnl.gov

Marina Kalyuzhnaya
University of Washington
mkalyuzh@u.washington.edu

Jay Keasling
University of California, Berkeley
JDKeasling@lbl.gov

Gerrit H. Kema
Plant Research International
gert.kema@wur.nl

Jaehan Kim
University of California, Davis
jaykim@ucdavis.edu

Stephen F. Kingsmore
National Center for Genome Resources
sfk@ncgr.org

Edward Kirton
DOE Joint Genome Institute
eskirton@lbl.gov

Hajnalka E. Kiss
Los Alamos National Laboratory
hajkis@lanl.gov

Carey L. Kopay
Edward Teller Education Center
ckopay@ucdavis.edu

Chai-Shian Kua
Xishuangbanna Tropical Botanical Garden
kuacs@yahoo.com

Jennifer V. Kuehl
DOE Joint Genome Institute
jvkuehl@lbl.gov

Victor Kunin
DOE Joint Genome Institute
vkunin@lbl.gov

Alan Kuo
DOE Joint Genome Institute
akuo@lbl.gov

Nikos Kyrpides
DOE Joint Genome Institute
NCKyrpides@lbl.gov

Attendees

Jesus Lacal Romero
University of California, Davis
jesus.lacal@eez.csic.es

Kathleen M. Lail
DOE Joint Genome Institute
klail@lbl.gov

Regina Lamendella
University of Cincinnati
reginalamendella@hotmail.com

Dorothy M. Lang
Lawrence Livermore National Laboratory
lang21@llnl.gov

Alla L. Lapidus
DOE Joint Genome Institute
alapidus@lbl.gov

Warren C. Lathe
OpenHelix
wlathe@openhelix.com

Debbie Laudencia-Chingcuanco
USDA Agricultural Research Service
Western Regional Research Center
dlc@pw.usda.gov

Gerard R. Lazo
USDA Agricultural Research Service
Western Regional Research Center
lazo@pw.usda.gov

Patrick Lee
University of California, Berkeley
leep@berkeley.edu

Sara K. Light
Lawrence Livermore National Laboratory
light6@llnl.gov

Erika A. Lindquist
DOE Joint Genome Institute
ealindquist@lbl.gov

Jinyuan Liu
Gillian Turgeon Laboratory
jl454@cornell.edu

Siqing Liu
USDA Agricultural Research Service
National Center for Agricultural Utilization Research
hughesk@ncaur.usda.gov

Riccardo LoCascio
University of California, Davis
rglocascio@ucdavis.edu

Todd M. Lowe
University of California, Santa Cruz
lowe@soe.ucsc.edu

Stephen R. Lowry
DOE Joint Genome Institute
slowry@lbl.gov

Susan M. Lucas
DOE Joint Genome Institute
lucas11@llnl.gov

Athanasios Lykidis
DOE Joint Genome Institute
alykidis@lbl.gov

Susan V. Lynch
University of California, San Francisco
lynchs@anesthesia.ucsf.edu

Rachel Mackelprang
DOE Joint Genome Institute
rmackelprang@lbl.gov

Julita E. Madejska
DOE Joint Genome Institute
MADEJSKA2@LLNL.GOV

Tim Magnuson
Idaho State University
magntimo@isu.edu

Kristina M. Mahan
University of California, Davis
kmmahan@ucdavis.edu

Stephanie A. Malfatti
JGI-LLNL
malfatti3@llnl.gov

Angela M. Marcobal
University of California, Davis
amarcobal@ucdavis.edu

Konstantinos Mavrommatis
DOE Joint Genome Institute
KMavrommatis@lbl.gov

Sarah M. McDonald
Rosenstiel School of Marine and Atmospheric Science
mcdsarah@googlemail.com

Kevin McLoughlin
Lawrence Livermore National Laboratory
mcloughlin2@llnl.gov

David A. Mead
Lucigen
dmead@lucigen.com

Monica Medina
University of California, Merced
mmedina@ucmerced.edu

Linda Meincke
JGI Los Alamos National Laboratory
meincke@lanl.gov

Todd P. Michael
Salk Institute for Biological Studies
tmichael@salk.edu

Todd R. Miller
Johns Hopkins University
trmiller@jhsph.edu

Charles D. Miller
Utah State University
cdmiller@biology.usu.edu

David A. Mills
University of California, Davis
damills@ucdavis.edu

Bradley Moore
Scripps Institution of Oceanography
bsmoore@ucsd.edu

Nancy A. Moran
University of Arizona
nmoran@email.arizona.edu

Jochen A. Mueller
Morgan State University
jmueller@jewel.morgan.edu

A. Christine Munk
JGI-LANL
cmunk@lanl.gov

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

Ali Navid
Lawrence Livermore National Laboratory
navid1@llnl.gov

Beth A. Nelson
Novozymes, Inc.
bane@novozymes.com

Audra Nemir
University of California, Berkeley
audra.nemir@gmail.com

Dianne Newman
California Institute of Technology
dkn@gps.caltech.edu

Dean Ng
DOE Joint Genome Institute
dng@lbl.gov

Wing Chi Ngau
DOE Joint Genome Institute
wcngau@lbl.gov

Chris E. Noriega
University of California, Davis
noriegce@yahoo.com

Donald L. Nuss
University of Maryland Biotechnology Institute
nuss@umbi.umd.edu

Maria Olmedo
Universidad de Sevilla
mariaolmedo@us.es

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

Ronald S. Oremland
U.S. Geological Survey
roremlan@usgs.gov

Robert Otillar
DOE Joint Genome Institute
RPOtillar@lbl.gov

Anu Padki
DOE Joint Genome Institute
ampadki@lbl.gov

Brian Palenik
University of California, San Diego
bpalenik@ucsd.edu

Jasmyn Pangilinan
DOE Joint Genome Institute
jlpangilinan@lbl.gov

Juan Parales
University of California, Davis
jvparales@ucdavis.edu

Rebecca E. Parales
University of California
reparales@ucdavis.edu

Andrew H. Paterson
University of Georgia
paterson@uga.edu

Jayant Patil
DOE Joint Genome Institute
jmpatil@lbl.gov

Len Pennacchio
DOE Joint Genome Institute
LAPennacchio@lbl.gov

Kory M. Pennebaker
University California, Santa Cruz
pennebak@pmc.ucsc.edu

Rene Perrier
DOE Joint Genome Institute
RAPerrier@lbl.gov

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

Gaston M. Pfluegl
University of California, Los Angeles
gaston@lsic.ucla.edu

Jonathan K. Pham
University of California, Davis
jkpham@ucdavis.edu

Lena Philip
DOE Joint Genome Institute
lphilip@lbl.gov

Michael Philips
DOE Joint Genome Institute
mphilips@lbl.gov

Michael R. Pintor
DOE Joint Genome Institute
mrpintor@lbl.gov

Antonio G. Pisabarro
Public University of Navarre
gpisabarro@unavarra.es

Samuel Pitluck
DOE Joint Genome Institute
s_pitluck@lbl.gov

Darren M. Platt
DOE Joint Genome Institute
platt5@llnl.gov

Soshanna Potter
California Institute of Technology
szp@caltech.edu

James F. Presston
University of Florida
jpreston@ufl.edu

Stephen Quake
Stanford University Medical Center
quake@stanford.edu

Christine A. Rabinovitch
University of California, Davis
carabino@ucdavis.edu

Brian A. Rabkin
DOE Joint Genome Institute
barabkin@lbl.gov

Preethi Ramaiya
Novozymes Inc.
pira@novozymes.com

Jason Raymond
Lawrence Livermore National Laboratory
raymond20@llnl.gov

Wayne G. Reeve
Murdoch University
reeve@murdoch.edu.au

David A. Relman
Stanford University
relman@stanford.edu

Paul M. Richardson
DOE Joint Genome Institute
pmrichardson@lbl.gov

Monica Riley
Marine Biological Laboratory
mriley@mbl.edu

Monica Riley
Marine Biological Laboratory
mriley@mbl.edu

Simon R. Roberts
DOE Joint Genome Institute
srroberts@lbl.gov

David S. Robinson
DOE Joint Genome Institute
dsrobinson@lbl.gov

Julio L. Rodriguez-Romero
Universidad de Sevilla
jlrodriguez@us.es

Dan Rokhsar
DOE Joint Genome Institute
DSRokhsar@lbl.gov

Bill Romine
Lawrence Livermore National Laboratory
romine1@llnl.gov

Nina Rosenberg
Lawrence Livermore National Laboratory
rosenberg4@llnl.gov

Eddy Rubin
DOE Joint Genome Institute
EMRubin@lbl.gov

Asaf A. Salamov
DOE Joint Genome Institute
aasalamov@lbl.gov

Kennan Salinero
University of California, Berkeley
kellarkv@nature.berkeley.edu

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

John Sanseverino
University of Tennessee
jsansev@utk.edu

Elizabeth Saunders
Los Alamos National Laboratory
ehs@lanl.gov

Wendy S. Schackwitz
DOE Joint Genome Institute
wsschackwitz@lbl.gov

Stephan Schmitz-Esser
University of Vienna
schmitz-esser@microbial-ecology.net

Jeremy J. Schmutz
Stanford Human Genome Center
jeremy@shgc.stanford.edu

David Sela
University of California
dasela@ucdavis.edu

Vladimir S. Sentchilo
University of Lausanne
vladimir.sentchilo@unil.ch

Keelnatham T. Shanmugam
University of Florida
shan@ufl.edu

Attendees

Harris Shapiro
DOE Joint Genome Institute
hjshapiro@lbl.gov

Nicole R. Shapiro
DOE Joint Genome Institute
nrshapiro@lbl.gov

Vandana Sharma
University of Dayton
sharmava@notes.udayton.edu

Henry F. Shaw
Lawrence Livermore National Laboratory
shaw4@llnl.gov

Sheri Simmons
University of California, Berkeley
sheris@eps.berkeley.edu

Robert W. Simons
University of California
bobs@microbio.ucla.edu

David R. Sims
Los Alamos National Laboratory
dsims@lanl.gov

Steven W. Singer
Lawrence Livermore National Laboratory
singer2@llnl.gov

Doug G. Smith
DOE Joint Genome Institute
dgsmith@lbl.gov

Jay V. Solnick
University of California, Davis
jvsolnick@ucdavis.edu

Rotem Sorek
DOE Joint Genome Institute
rsorek@lbl.gov

Jim Spain
Georgia Institute of Technology
jspain@ce.gatech.edu

John R. Spear
Colorado School of Mines
jspear@mines.edu

Cailyn Spurrell
DOE Joint Genome Institute
chspurrell@lbl.gov

David A. Stahl
University of Washington
dastahl@u.washington.edu

Jan Stenlid
Swedish University of Agricultural Sciences
Jan.Stenlid@mykopat.slu.se

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

Craig Stephens
Santa Clara University
cstephens@scu.edu

John F. Stolz
Duquesne University
stolz@duq.edu

Richard Sucgang
Baylor College of Medicine
rsucgang@bcm.edu

Hui Sun
DOE Joint Genome Institute
hsun@lbl.gov

Shini Sunagawa
University of California, Merced
ssunagawa@gmx.de

Lukasz Szajkowski
Lawrence Livermore National Laboratory
szajkowski2@llnl.gov

Ernest Szeto
JGI – Lawrence Berkeley National Laboratory
eszeto@lbl.gov

Makio Tamura
Lawrence Livermore National Laboratory
tamura2@llnl.gov

Astrid Terry
DOE Joint Genome Institute
ayterry@lbl.gov

Michael P. Thelen
Lawrence Livermore National Laboratory
mthelen@llnl.gov

Linda S. Thompson
Los Alamos National Security, LLC
thompson_l_sue@lanl.gov

Hope N. Tice
DOE Joint Genome Institute
tice1@llnl.gov

James M. Tiedje
Michigan State University
tiedjej@msu.edu

Damon Tighe
DOE Joint Genome Institute
tighe2@llnl.gov

Sara Ting
DOE Joint Genome Institute
sting@lbl.gov

Christian M. Tobias
USDA Agricultural Research Service
Western Regional Research Center
ctobias@pw.usda.gov

Tamas Torok
Lawrence Berkeley National Laboratory
ttorok@lbl.gov

Susannah G. Tringe
DOE Joint Genome Institute
sgtringe@lbl.gov

Stephan Trong
DOE Joint Genome Institute
trong1@llnl.gov

Hank Tu
DOE Joint Genome Institute
hctu@lbl.gov

Gillian B. Turgeon
Cornell University
bgt1@cornell.edu

Jerry Tuskan
DOE Joint Genome Institute
GATuskan@lbl.gov

Parag Vaishampayan
DOE Joint Genome Institute
PAVaishampayan@lbl.gov

Fernando Valle
Genencor
fvalle@genencor.com

John P. Vogel
USDA Agricultural Research Service
Western Regional Research Center
jvogel@pw.usda.gov

Christian C. Voolstra
Postdoctoral Scholar
chris.voolstra@gmail.com

Larry Wackett
University of Minnesota
wackett@cbs.umn.edu

Irene I. Wagner-Döbler
Helmholtz-Center for Infection Research
Irene.Wagner-Doebler@Helmholtz-hzi.de

Jeffrey J. Walker
University of Colorado
jeffrey.walker@colorado.edu

David A. Walsh
University of British Columbia
dawalsh@interchange.ubc.ca

Mei Wang
DOE Joint Genome Institute
mwang@lbl.gov

Falk Warnecke
DOE Joint Genome Institute
FWarnecke@lbl.gov

Sarah C. Watkinson
University of Oxford, UK
sarah.watkinson@plants.ox.ac.uk

Greg Werner
DOE Joint Genome Institute
gmwerner@lbl.gov

Kimberlee West
University of California, Berkeley
kimberleew@gmail.com

Korin E. Wheeler
Lawrence Livermore National Laboratory
korin@llnl.gov

Bryan A. White
University of Illinois
bwhite44@uiuc.edu

Paul Wilmes
University of California, Berkeley
pwilmes@berkeley.edu

Reinhard Wilting
Novozymes A S
rwil@novozymes.com

Paul Winward
DOE Joint Genome Institute
pwinward@lbl.gov

Christopher T. Wong
DOE Joint Genome Institute
CTWong@lbl.gov

Valerie Wong
University of California, Berkeley
vwong@berkeley.edu

Alexandra Z. Worden
University of Miami
aworden@rsmas.miami.edu

Tanja Woyke
DOE Joint Genome Institute
twoyke@lbl.gov

John C. Wright
Applied Biosystems
john.wright@appliedbiosystems.com

Dongying Wu
University of California, Davis
dongyingwu1@gmail.com

Martin Wu
University of California, Davis
Genome Center
martinwu@ucdavis.edu

Cindy H. Wu
Lawrence Berkeley National Laboratory
CHWu@lbl.gov

Gary Xie
JGI-Los Alamos National Laboratory
xie@lanl.gov

Debbie S. Yaver
Novozymes Inc
dsy@novozymes.com

Suzan Yilmaz
DOE Joint Genome Institute
syilmaz@lbl.gov

John I. Yoder
University of California, Davis
jiyoder@ucdavis.edu

Andy C. Yuen
DOE Joint Genome Institute
acyuen@lbl.gov

Tao Zhang
DOE Joint Genome Institute
tzhang3@lbl.ov

Michael Y. Zhang
Joint Genome Institute
myzhang@lbl.gov

Jean Zhao
DOE Joint Genome Institute
zyzhao@lbl.gov

Jizhong Zhou
University of Oklahoma
jzhou@ou.edu

Kemin Zhou
DOE Joint Genome Institute
kzhou@lbl.gov

Carol E. Zhou
Lawrence Livermore National Laboratory
zhou4@llnl.gov

Author Index

Authors

Authors

Notes