First Annual
DOE Joint Genome Institute
User Meeting

Sponsored By

U.S. Department of Energy
Office of Science

March 29–April 1, 2006

Embassy Suites Hotel and DOE Joint Genome Institute

Walnut Creek, California

Contents

2006 DOE JGI User Meeting
March 29-April 1, 2006
Embassy Suites Hotel, Walnut Creek, CA

AGENDA

(All events will take place in the **Embassy Suites Contra Costa Room** *unless otherwise noted.)*

WEDNESDAY – March 29, 2006

Start Time	*End Time*	*Subject*	*Session Chair/Speaker*
5:30 PM	**6:30 PM**	*Registration,* **Welcome Reception** *(**Location:** Embassy Suites, Atrium)*	
6:30 PM	**8:00 PM**	**Keynote Addresses:**	
6:30 PM	7:15 PM	***Prochlorococcus*: A model for systems biology** **Penny Chisholm** Lee and Geraldine Martin Professor of Environmental Studies, Dept. of Biology, MIT	
7:15 PM	8:00 PM	**Post-Synthesis Genomics** **Drew Endy** Assistant Professor, Biological Engineering Division, MIT	

THURSDAY – March 30, 2006

8:15 AM	8:45 AM	*Arrival, Registration*	
8:45 AM	**9:00 AM**	**Welcome, Introduction**	Eddy Rubin
9:00 AM	**12:00 PM**	**Session I: Large Eukaryotic Genomics**	Dan Rokhsar
9:00 AM	9:30 AM	The Xenopus tropicalis genome and its uses	Richard Harland
9:30 AM	10:00 AM	Plant genomics	Joe Ecker
10:00 AM	10:30 AM	Poplar genome	Jerry Tuskan
10:30 AM	11:00 AM	*Break*	
11:00 AM	11:30 AM	Mimulus guttatus	John Willis
11:30 AM	12:00 AM	Chlamydomonas genome	Arthur Grossman
12:00 PM	**2:00 PM**	*Lunch,* **Poster Session 1** (***Location:*** Embassy Suites, Alameda Room)	
2:00 PM	**5:00 PM**	**Session II: Microbial Genomics**	Paul Richardson
2:00 PM	2:30 PM	In situ gene expression and genomic analysis of the Geobacteraceae	Dawn Holmes
2:30 PM	3:00 PM	Multiple genomes of Shewanella strains: what are we learning?	Ken Nealson
3:00 PM	3:30 PM	Post Genomic Studies in the Facultative Photosynthetic Bacterium, *Rhodobacter sphaeroides* 2.4.1	Sam Kaplan
3:30 PM	3:45 PM	*Break*	
3:45 PM	4:00 PM	The role of mismatch repair in the evolution of Oenococcus oeni	David Mills
4:00 PM	4:30 PM	Ethanol producing E. coli	Lonnie Ingram
4:30 PM	5:00 PM	The white rot fungus Phanerochaete chrysosporium	Dan Cullen
5:00 PM	5:30 PM	*Transportation to JGI* (Buses leave from the Embassy Suites Hotel Main Entrance)	
5:30 PM	**8:00 PM**	**Reception, Poster Session 2, Tour of JGI** (***Location:*** JGI Conference Center)	
8:00 PM	8:30 PM	*Transportation from JGI to Embassy Suites*	

2006 DOE JGI User Meeting
March 29-April 1, 2006
Embassy Suites Hotel, Walnut Creek, CA

(All events will take place in the **Embassy Suites Contra Costa Room** *unless otherwise noted.)*

FRIDAY – March 31, 2006

Start Time	*End Time*	*Subject*	*Session Chair/Speaker*
8:30 AM	**12:30 PM**	**Session III: Microbial Communities**	Jim Bristow
8:30 AM	9:00 AM	Phylogenetic and other approaches to binning of metagenomic sequences: theoretical and applied issues	Jonathan Eisen
9:00 AM	9:30 AM	A tale of two sludges	Phil Hugenholtz
9:30 AM	10:00 AM	Microbial community genomics in the ocean: from gene ecology to systems biology	Ed DeLong
10:00 AM	10:30 AM	Horizontal gene transfer in recent evolutionary history of Archaea	Rachel Whitaker
10:30 AM	11:00 AM	*Break*	
11:00 AM	11:30 AM	Sequencing the hindgut bacterial community from a wood-feeding termite species from Costa Rica	Jared Leadbetter
11:30 AM	12:00 AM	Bacteriophage DNA Polymerases and Other Surprises from Superheated Thermal Aquifers	David Mead
12:00 AM	12:15 PM	Deciphering symbiosis through metagenomic analysis of the bacterial consortium in a gutless worm	Nicole Dubilier
12:15 PM	12:30 PM	Complete Genome Sequences and Functional Genomics of Green Bacteria	Don Bryant
12:30 PM	**2:00 PM**	*Lunch,* **Users Business Meeting**	
2:00 PM	**4:45 PM**	**Session IV: Technology Development**	Sue Lucas
2:00 PM	2:30 PM	Solexa technology	Gary Schroth
2:30 PM	3:00 PM	454 technology	Michael Egholm
3:00 PM	3:30 PM	Single Molecule Sequencing of DNA and RNA	Stan Lapidus
3:30 PM	3:45 PM	*Break*	
3:45 PM	4:15 PM	Integrating New Technologies into The JGI Microbial Program	Paul Richardson
4:15 PM	4:45 PM	Informatics: Surviving the data deluge	Darren Platt

SATURDAY – April 1, 2006

9:00 AM	**1:00 PM**	**Session V: IMG Workshop**	Nikos Kyrpides
9:00 AM	9:30 AM	JGI Microbial Genome Analysis pipeline	Athanasios Lykidis
9:30 AM	10:00 AM	Using IMG	Iain Anderson
10:00 AM	10:30 AM	Using IMG/M (IMG/metagenomes)	Natalia Ivanova
10:30 AM	10:45 AM	*Break*	
10:45 AM	11:00 AM	Future Developments of IMG	Nikos Kyrpides
11:00AM	11:45 AM	Hands-on IMG	
11:45 AM	12:15 PM	Discussion/Feedback	
12:15 PM	1:00 PM	Hands-on IMG, continued (optional)	
	1:00 PM	**End of 2006 JGI User Meeting**	

Speaker Presentations

Abtracts in order of presentation according to agenda (p. iv).

Prochlorococcus: A Model for Systems Biology

Penny Chisholm (chisholm@mit.edu)

Massachusetts Institute of Technology, Cambridge, MA

Prochlorococcus is the numerically dominant photosynthetic microbe in the tropical and sub-tropical oceans, often reaching concentrations of 10^8 cells per liter, and contributing a significant fraction of the total photosynthesis in ocean ecosystems. This group is comprised of closely related (less than 3% different in their rRNA sequence), phylogenetically and physiologically distinct 'ecotypes' that display different relative abundances along the light, temperature, nutrient, and predator (including viruses) gradients in the oceans. Genome sequences from a number of cultured *Prochlorococcus* strains, phages that infect them, and fragments of DNA from wild *Prochlorococcus* cells are providing exciting new insights into the origins and dynamics of micro-diversity within the *Prochlorococcus* meta-population, and the self-organization of microbial ecosystems.

Prochlorococcus is thus proving to be a useful model for developing cross-scale systems biology, the goal of which is to understand life processes at all scales of organization, from the cell to the biosphere.

In Situ Gene Expression and Genomic Analysis of *Geobacteraceae*

Dawn Holmes (dholmes@microbio.umass.edu)

University of Massachusetts, Amherst, MA

To be able to optimize and predict the activity of *Geobacteraceae* in subsurface environments, it is necessary to first have an understanding of which *Geobacteraceae* genes are present in the environment and whether they are expressed *in situ*. Three different approaches were used to obtain genomic data from a uranium-contaminated aquifer in Rifle, CO in which the growth of *Geobacteraceae* was stimulated by the addition of acetate:1) multiple displacement amplification (MDA) of genomic DNA from single cells isolated from the environment; 2) small insert clone libraries constructed from genomic DNA extracted directly from the environment; and 3) targeted amplification of *Geobacteraceae* genes with degenerate primers. The genomic information obtained from these studies was then used to construct a number of quantitative RT-PCR primers targeting genes that are diagnostic of a range of metabolic states and highly conserved throughout the *Geobacter*aceae. Rates of metabolism, stress

levels, chemotaxis, motility, and nutrient requirements of *Geobacteraceae* in groundwater samples collected from the U(VI) contaminated site during bioremediation were monitored with these primers. A direct correlation was observed between expression of a central metabolism gene in *Geobacteraceae*, malate dehydrogenase (*mdh*) and acetate concentrations detected in the groundwater. Expression of *ideR*, an iron dependent repressor remained constant throughout the experiment once *Geobacteraceae* became dominant members of the community, indicating that cells were limited for Fe(II) during the bioremediation process. In addition, the number of transcripts for *feoB* (iron uptake protein), the twitching motility protein (*pilT*), and the heavy metal resistance protein (*cusA*) increased significantly during U(VI) bioremediation. These results demonstrate that monitoring the *in situ* transcript levels of key genes can provide insight into the rates of metabolism and nutrient requirements of *Geobacteraceae* during *in situ* bioremediation.

Post Genomic Studies in the Facultative Photosynthetic Bacterium, *Rhodobacter sphaeroides* 2.4.1

Samuel Kaplan (samuel.kaplan@uth.tmc.edu)

University of Texas Health Science Center, Houston, TX

www.rhodobacter.org

Rhodobacter sphaeroides is a member of the ά-3 proteobacteria belonging to the non-sulfur purple photosynthetic group. *R. sphaeroides* possesses as extensive repertoire of reaction pathways enabling it to obtain energy, reducing power, carbon and nitrogen through a display of physiological diversity unmatched by most living systems. *R. sphaeroides* possesses a complex genome consisting of two chromosomes whose relationship to one another appears to date to the origin of the species. Of particular note is the ability of *R. sphaeroides* to undergo significant morphogenetic change when it is shifted from chemoheterotrophic growth in the presence of O_2 to photoheterotrophic growth in the absence of O_2. Under these conditions, the cell membrane undergoes a series of invaginations to form the intracellular, vesicular membrane system or ICM, which comprises the photosynthetic apparatus whose abundance varies inversely to light intensity. The ICM is comprised of approximately 200-250 unique proteins of which only approximately 40 have been unequivocally identified, and which correspond to approximately 40-60% of the ICM protein. We now have the ability, for the first time, to identify all of the ICM proteins and their possible role(s) in photosynthesis and ICM structure through a combination of proteomic and post-genomic studies. Regulating the induction and synthesis of the ICM are: the PpsR repressor/AppA antirepressor unique to PS gene control, the Prr two component regulatory system which in addition to regulating ICM formation is a truly global regulatory element controlling either directly or indirectly over 800 genes/orfs, FnrL which is highly selective in PS gene

control and tetrapyrrole regulation. The major pathway through which O_2 exercises control over PS gene expression is through the terminal cytochrome *c* oxidase, cbb_3 which signals the membrane bound PrrB histidine kinase to increase its phosphatase activity relative to its kinase activity which maintains the response regulator PrrA in the phosphate negative mode. The cbb_3 terminal oxidase, in addition to its role in respiratory electron transport, is also shown to conduct electron flow under anaerobic conditions. In conjunction with this anaerobic electron flow the cbb_3 plays a role in the regulation of over 230 genes/orfs and the ultimate electron acceptor for this anaerobic electron flow is thought to be an intermediate involved in carotenoid biosynthesis. The Prr system is shown to regulate either directly or indirectly over 21% of the *R. sphaeroides* genome. Given the complete DNA sequence of the *R. sphaeroides* 2.4.1 genome, its annotation and an array of tools, e.g., facile genetics, DNA chip, proteome profiles etc., we are in the unique position of being able to assess the complete composition and function of the ICM and the regulatory processes which accompany the transition of one growth state to another, as environmental conditions change.

This work is supported by grants from the NIH and the DOE.

The White Rot Fungus *Phanerochaete chrysosporium*

Dan Cullen (dcullen@facstaff.wisc.edu)
University of Wisconsin, Madison, WI

The most abundant source of global carbon is plant biomass, composed primarily of cellulose, hemicellulose and lignin. Many microorganisms are capable of degrading and utilizing cellulose and hemicellulose. However, a much smaller group of filamentous fungi have the ability to breakdown lignin, the most recalcitrant component of plant cell walls. Collectively referred to as white rot fungi, they possess the unique ability to efficiently degrade lignin to CO_2. Common inhabitants of forest litter and fallen trees, these fungi have attracted considerable attention in a wide range of applications including fiber bleaching and hazardous waste remediation. All white rot fungi are members of the Basidiomycotina, a large and diverse phylum.

In 2002, the DOE Joint Genome Institute released the initial genome assembly (v1.0) for *Phanerochaete chrysosporium*, the most widely studied white rot fungus. Analysis of the ~30 Mb haploid genome revealed an impressive array of genes encoding extracellular proteins (Martinez et al 2004 *Nat. Biotechnol.* 22:695-700), many of which are likely involved in lignocellulose degradation. Among the secreted oxidative enzymes, complex families of peroxidases, copper radical oxidases, and FAD-dependent oxidases were identified. The genome also harbors an amazing number of putative carbohydrate-active enzymes, including at least 160 glycoside hydrolases. In addition to extended families of structurally

related genes, an assortment of class I and II transposable elements (TEs) were found in the *P. chrysosporium* genome. Several TEs have disrupted and inactivated genes potentially involved in lignocellulose degradation.

The JGI's February 2005 release (v2.1) of improved gene models has paved the way for high throughput protein identification by LC-MS/MS. Recent results will be presented.

A Tale of Two Sludges

Phil Hugenholtz (phugenholtz@lbl.gov),[1] Héctor García Martín,[1] Victor Kunin,[1] S. Brook Peterson, [3] Natalia Ivanova,[1] Falk Warnecke,[1] Linda L. Blackall,[2] and Katherine D. McMahon[3]

[1]DOE Joint Genome Institute, Walnut Creek, CA; [2]Advanced Wastewater Management Centre, University of Queensland, Queensland, Australia; and [3]Civil and Environmental Engineering Department, University of Wisconsin, Madison, WI

Metagenomic datasets hold the allure of providing genome-wide insights into microbial population structure without the potential bias of cultivation. We obtained ca. 150 Mbp of shotgun sequence data from two sludge samples taken from geographically remote lab-scale bioreactors to investigate the structure and biogeography of the dominant and as-yet uncultivated bacterial species, Candidatus Accumulibacter phosphatis (CAP). Each sludge was dominated by a single nearly clonal CAP population that was flanked by a number of other low abundance CAP strains. The dominant strains were closely related sharing >95% sequence identity at the nucleotide level over most of their genomes. Differences were mostly localized in genomic regions encoding cassettes of exopolysaccharide synthesis genes and CRISPR elements that we speculate are highly variable in response to phage predation. We conclude that CAP can be globally dispersed possibly via aquatic reservoirs and that CAP populations in sludge may be inherently vulnerable to catastrophic collapse by virtue of their clonality and absence of detectable homologous recombination.

Microbial Community Genomics in the Ocean: From Gene Ecology to Systems Biology

Edward F. DeLong (delong@mit.edu)

Massachusetts Institute of Technology, Cambridge, MA

Microbial life predominates in the ocean, yet much remains to be learned about its genomic variability, especially along the depth continuum. We recently conducted genomic analyses of planktonic microbial communities in the Pacific Ocean, from the surface waters to 4000 m deep. Sequence variation of microbial community genes reflected vertical zonation of

taxonomic groups, as well as differential distributions of functional genes. These distributional patterns of microbial genes suggested depth-variable community trends in carbon and energy metabolism, attachment and motility, gene mobility, and host-viral interactions. In addition, lateral gene transfer events that tracked environmental forcing factors and selective pressure were also observed. Our Pacific Ocean genomic depth profiles have revealed several ecological trends that apparently reflect higher order community organization and dynamics.

Despite the dizzying progress in microbial metagenomic sequencing, its important to recognize that a number of critical challenges lay ahead. The fostering of broad-based informatics efforts is a critical, including the integration of genomic datasets with environmental and ecological data. Measurement of microbial diversity and genomic variability, as well an microbial function process, and distributions, are increasing in scope, resolution and frequency. The various "snapshots" of gene content in ocean microbial populations may soon develop into motion pictures, that better describe genomic diversity, function and evolution in complex natural communities. Genome-derived functional hypotheses need to be tested, verified and quantified in the field. Similarly, process-oriented studies of individual "functional groups" need to strive towards simultaneous analyses of many different groups, to understand their dynamics and interactions. Genomic and post genomic tools may soon help bring the resolution and frequency of many biological measurements, on par with the physical and chemical measurements of environmental processes and phenomena.

Sequencing the Hindgut Bacterial Community from a Wood-Feeding Termite Species from Costa Rica

Jared Leadbetter (jleadbetter@caltech.edu)

California Institute of Technology, Pasadena, CA

Termites play critical roles in plant lignocellulose turnover and the biological emissions of CO_2 and CH_4 globally. They are dependent in this endeavor on the activities of a unique microbial community housed within their hindguts. This community can contain on the order of 100 microbial species found no where else in nature. These microbial ecosystems are simple in an unusual sense of the word, because they are tiny, on the order of one microliter in volume, and have a distinct boundary—the gut lining of the insect. Yet because of the huge reach and great abundances of their hosts, these microbial ecosystems become major sites of catalysis that ultimately exert a huge impact on global C & N biogeochemistry, especially in tropical rainforests and similar biomes. Despite their mutually obligate interactions with their host, 80 years of study have revealed only a meager understanding of the roles, functions, and even the identities of the majority of termite gut microbes. This CSP has focused on making major, novel inroads into achieving a better understanding of the

termite gut microbiota and its lignocellulose processing activities. A metagenomic study on the hindgut microbial community of a representative "higher" wood-feeding termite species, *Nasutitermes corniger*, is now well underway. Ca. 80 Mb of termite gut metagenome has been sequenced and is now being analyzed and annotated. This program involves key collaborations between laboratories at the California Institute of Technology in Pasadena; Diversa Corp. of San Diego; INBio, Costa Rica's National Biodiversity Institute; and the DOE Joint Genome Institute in Walnut Creek.

Bacteriophage DNA Polymerases and Other Surprises from Superheated Thermal Aquifers

David Mead (dmead@lucigen.com),[1] Vinay Dhodda,[1] Robert DeFrancesco,[1] Melodee Patterson,[1] Paul Richardson,[2] Mark Young,[3] Phillip Brumm,[1] and Thomas Schoenfeld[1]

[1]Lucigen, Middleton, WI; [2]DOE Joint Genome Institute, Walnut Creek, CA; and [3]Montana State University, Bozeman, MT

Boiling thermal pools present an attractive target for community genomic analysis. They are physically isolated from other ecosystems, are among the most hostile environments known, and little is known about their constituents. The abundance of microbes and phage in thermal aquifers is significantly lower than fresh or salt water habitats, requiring new methods for extracting genetic information. Water column samples were collected from Bath hot spring, a boiling thermal pool in Yellowstone National Park, and adherent cells surrounding the pool. Planktonic organisms in the water column potentially originate from regions of 150-200°C found at depths of 100-300 meters throughout YNP. The microbial community of the water column was distinct from that found in the outflow sediment. Unusual and novel morphologies, some resembling multicellular assemblies, were observed in enrichment cultures. A pure culture demonstrating some of these characteristics has 16S rDNA similar to Thermus.

Waterborne viruses (phages) in thermal springs have been largely unexamined, despite their potential importance in the biosphere. Phages promote microbial diversity by predation of the most abundant microbes and by transfer of genes through transduction and lysogeny. Sequence analysis (>37,000 reads) of thermal aquifer phage libraries provides insight into viral complexity, lifestyles, molecular diversity, and access to coding sequences for useful proteins. Homology to a handful of cultivated viruses from high temperature springs establishes an ecological framework for these environmental samples. Numerous similarities to enzymes associated with temperate phage, such as integrases, suggest that lysogeny is common in thermal environments. Unexpectedly, a *Thermosynechococcus*-like photosynthesis regulatory gene was found next to an integrase gene. Over 200 DNA polymerase genes have been identified with approximately

58 being full length. Ten active enzymes have been expressed, one of which allows isothermal amplification at elevated temperatures with greater specificity than conventional polymerases. This subset of thermostable phage DNA polymerases appears much more diverse than known microbial or phage enzymes. The fact that no *pol* genes were re-isolated suggests the level of diversity seen so far is the tip of a very large iceberg.

Poster Presentations

Posters alphabetical by first author. *Presenting author.

1. Industrial, Clinical and Environmental Significance of *Ralstonia pickettii*

Catherine C. Adley* (catherine.adley@ul.ie)
University of Limerick, Limerick, Ireland

2. Sequence-Based Identification and Phylogenetic Analysis of Filamentous Fungi Isolated from Extreme Environments

G. L. Andersen, E. L. Brodie, T. DeSantis, P. Hu, Y. Piceno, and **T. Torok*** (ttorok@lbl.gov)
Center for Environmental Biotechnology, Lawrence Berkeley National Laboratory, Berkeley, CA

Extremophillic microorganisms have become a valuable bioprospecting target because they represent a challenging scientific opportunity not only for those interested in microbial diversity and the evolution of life but for researchers searching for clues to extraterrestrial life. Extremophiles also produce unique biocatalysts that function under extreme conditions comparable to those prevailing in various industrial processes. Therefore, taxonomic identification and detailed characterization of these microorganisms provide a greater understanding of their diversity, unique metabolism, and ecological function.

Fungi are of great importance because (1) they are the primary decomposers and biosorbents in all terrestrial ecosystems, (2) they are important symbiotic associates of vascular plants both in mutualistic and parasitic relationships, (3) they constitute the overwhelming majority of plant pathogens and as such have a tremendous economic impact, and thereby, represent the next wave of potential bioterrorism threats, (4) several human pathogens are fungi, (5) they offer several well-developed genetic systems for molecular biologists as eukaryotic model organisms, (6) they are of significant potential for bioremediation, and finally (7) they are crucial to the fermentation and biotechnology industries.

Over a period of almost two decades, large numbers of filamentous fungi were isolated from extreme environments in the USA and the territory of the former Soviet Union. These environmental field sites were considered extreme in terms of their water activity, xenobiotic contamination, temperature, geology or chemistry. Published media and isolation techniques were modified and applied to simulate the eco-physiological conditions at the environmental habitat. The pure cultures of filamentous

fungi were preserved and stored at ultra-low temperature for long-term maintenance.

Currently ongoing work involves a collection of filamentous fungi that were isolated from environmental samples collected at and around the failed nuclear reactor in Chernobyl, Ukraine. The collection contains some 2,500 strains representing 98 genera and over 200 species. Characterization of these fungal organisms was carried out using classical colony- and cell morphology–based techniques, as well as molecular-level protocols. Here we compare the former with genomic DNA sequencing of specific target genes valuable for phylogenetic differentiation: preliminary data obtained by sequencing the D1/D2 domains of the large ribosomal subunit RNA-genes in almost 200 strains allowed us to confirm the classical identification results in over 85% of the cases. Based on that, we discuss our objectives of (1) developing a repository of high quality biomarker gene sequences from this valuable collection of extremophilic filamentous fungi and thereby providing a nucleus for an online database of fungal biomarkers, (2) generating a cohesive fungal biomarker sequence database available online, and (3) using the compiled sequences to select DNA sequences unique to fungal families, genera and species leading to the design of a state-of-the-art Affymetrix-style DNA microarray for fungal detection and identification.

3. The Genomes of Five Polycylic Aromatic Hydrocarbon Degrading Mycobacteria

Anne J. Anderson* (anderson@biology.usu.edu)
Utah State University, Logan, UT

4. Community Sequencing of Novel Ultra-Small *Euryarchaea* from an Extremely Acidic Subsurface Mine, Iron Mountain

Brett J. Baker* (bbaker@eps.berkeley.edu),[1] Gene W. Tyson,[2] Richard Webb,[3] Phil Hugenholtz,[4] Eric E. Allen,[2] and Jillian F. Banfield[1, 2]

[1]Dept. of Earth and Planetary Sciences, University of California, Berkeley, CA; [2]Environmental Sciences, Policy, and Management, University of California, Berkeley, CA; [3]Dept. of Microbiology and Parasitology, University of Queensland, Australia; and [4]DOE Joint Genome Institute, Walnut Creek, CA

Culture-independent PCR-based surveys have greatly expanded our knowledge of microbial diversity. However these studies are blind to the subset of microorganisms whose genes (e.g. 16S rRNA) do not contain conserved primer motifs. Random shotgun sequencing of an extremely acidic (pH <0.9) chemolithotrophic subsurface biofilm community

recovered a genome fragment from a novel archaeon, named WTF-1. This novel lineage's 16S rRNA gene has several mismatches with all commonly used broad-specificity PCR primers. The group is deeply branched within a region of the *Euryarchaeota* that has few cultivated representatives. WTF specific rRNA gene primers revealed two new groups in mine samples, WTF-2 and WTF-3, which are highly divergent from WTF-1 sequences (8% and 17% respectively). Fluorescence *in situ* hybridization and PCR-based detection of these groups has shown that these organisms are ubiquitous in microbial communities involved in acid mine drainage (AMD) generation at the Richmond Mine field site, but they are relatively low in abundance. Analysis of community sequence (100 Mb) from another site in the mine has revealed another ~38kb of WTF-2 and less of WTF-1. Using GC content, coverage (4X in this new library), and phylogenetic analyses of unassigned archaeal genomic fragments we have been able to identify >70kb of WTF in this new dataset. Comparison of a syntenic region between these two groups (WTF-1 and 2) shows that there are roughly 89% similar at the DNA level. It is difficult to get adequate coverage of organisms that occur in relatively low abundance via community genomic analysis without greatly over sampling the dominant members. Using filtration (0.45 μm), and PCR to confirm it, we have been able to enrich for these WTF groups. It is this filtrate that we are sequencing in order learn more about these novel *Archaea*.

5. Flagellar Genes of *Selaginella moellendorffii*

Jody Banks* (banksj@purdue.edu)
Purdue University, West Lafayette, IN

6. The *Acidothermus cellulolyticus* Genome: A Preliminary Analysis

Ravi D. Barabote* (ravib@lanl.gov),[1] Gary Xie,[1] William S. Adney,[2] Philippe Normand,[3] Pierre Pujic,[3] and **Alison M. Berry*** (amberry@ucdavis.edu)[4]
[1]Biosciences Division, Los Alamos National Laboratory, Los Alamos, NM; [2]National Renewable Energy Laboratory, National Bioenergy Center, Golden, CO; [3]Ecologie Microbienne, UMR CNRS, Universite Claude Bernard Lyon, Villeurbanne, France; and [4]Dept. of Plant Sciences, University of California, Davis, CA

Acidothermus cellulolyticus is a high GC actinobacteria that dwells in acid hot springs such as Yellowstone National Park. It was isolated in a screen for efficient and complete cellulose degradation, and the glycosyl hydrolases are highly thermotolerant. *A. cellulolyticus* is also of interest evolutionarily, because its closest phylogenetic neighbor is *Frankia*, a

nitrogen-fixing plant root symbiont. Recently, the genome of *A. cellulolyticus* has been sequenced by the JGI. Preliminary genomic characterization of *A. cellulolyticus* is presented. The 2.4 MB genome is approximately 67% GC-rich and is predicted to encode 2151 proteins. A significant proportion of the *A. cellulolyticus* proteins show closest homology to proteins from *Streptomyces*, *Frankia*, and *Thermobifida fusca*. Approximately 2% of its proteins do not have any recognizable homologs in other organisms and appear to be unique to *A. cellulolyticus*. Analyses of the predicted transporters revealed that the genome encodes very few recognizable membrane channel proteins, and approximately equal numbers of permeases and ABC-type transporters. The genome encodes numerous glycosyl transferases and other enzyme activities involved in the complex degradation pathway of cellulose. Although previous studies have shown a bias in relative abundance of certain amino acids in thermophiles, no significant differences were observed in a comparative analysis of the amino acid distributions in *A. cellulolyticus* and six mesophilic high GC actinobacteria. Analyzing the *A. cellulolyticus* proteins whose closest homologs are proteins from other thermophiles, such as *T. fusca*, may reveal insights into common evolutionary adaptations to thermophilic lifestyles of these organisms. The underlying basis for thermotolerance could be exploited to engineer proteins for industrial and environmental applications.

7. *Thellungiella halophila*

Hans J. Bohnert* (bohnerth@life.uiuce.edu)
University of Illinois, Urbana-Champaign, IL

8. *Phaeodactylum tricornutum*: A Model Species for Exploring the Cell Biology of Diatoms

Chris Bowler* (cbowler@biologie.ens.fr),[1,2] Andrew Allen,[1] Sacha Coesel,[2] Alessandra De Martino,[1] Angela Falciatore,[2] Marc Heijde,[1] Kamel Jabbari,[1] Uma Maheswari,[1] Manuela Mangogna,[2] Florian Maumus,[1,2] Anton Montsant,[1,2] Edda Rayko,[1] Magali Siaut,[2] and Assaf Vardi[1]
[1]Dept. of Biology, Ecole Normale Supérieure, Paris, France and [2]Cell Signalling Laboratory, Stazione Zoologica, Naples, Italy

Diatoms are eukaryotic, photosynthetic microorganisms found throughout marine and freshwater ecosystems that are responsible for around 20% of global primary productivity. A defining feature of diatoms is their ornately patterned silicified cell wall (known as frustule), which display species-specific nanoscale-structures. These organisms therefore play major roles in global carbon and silicon cycles. The marine pennate diatom *Phaeodactylum tricornutum* is the second diatom for which a whole

genome sequence has been generated. It was chosen primarily because of the superior genetic resources available for this diatom (e.g., genetic transformation, 80,000 ESTs), and because it has been used in laboratory-based studies of diatom physiology for several decades. Although not considered to be of great ecological significance, it has been found in several locations around the world, typically in coastal areas with wide fluctuations in salinity. Unlike other diatoms it can exist in different morphotypes, and changes in cell shape can be stimulated by environmental conditions. This feature can be used to explore the molecular basis of cell shape control and morphogenesis. Furthermore the species can grow in the absence of silicon, and the biogenesis of silicified frustules is facultative, thereby providing opportunities for experimental exploration of silicon-based nanofabrication in diatoms. The sequence is 30 mega base pairs and, together with the sequence from the centric diatom *Thalassiosira pseudonana* (34 Mbp; the first diatom whole genome sequence), it provides the basis for comparative genomics studies of diatoms with other eukaryotes and will provide a foundation for interpreting the ecological success of these organisms.

9. Microbial Systems Division: Metabolism of Microbial Communities

Elbert Branscomb, **Patrik D'Haeseleer*** (patrikd@llnl.gov), Eivind Almaas, Jason Raymond, Tomer Altman, Jennifer Pett-Ridge, Steven Singer, Harry Beller, and Michael Thelen

Biosciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA

The Microbial Systems Division at LLNL takes a Systems Biology approach towards understanding the energy metabolism of microbial communities. This includes integrating biochemistry, thermodynamics, metabolite transport and utilization, meta-genomic sequencing, regulatory and metabolic network analysis, and comparative and evolutionary genomics. Our objective is to predict and rationalize the dominant microbially mediated metabolic activities that develop in a given type of environment, and to predict a microbe's behavior and lifestyle directly from its genomic sequence.

10. The JGI EST Pipeline

Peter Brokstein* (pbbrokstein@lbl.gov), Jarrod Chapman, and Erika Lindquist

DOE Joint Genome Institute, Walnut Creek, CA

Expressed Sequence Tags (ESTs) provides a critical counterpart to whole genome shotgun sequence data. ESTs provide a direct evidence of

transcribed regions in the genome, greatly increasing the accuracy of JGI genome annotations. EST projects can also provide a large amount of information about a genome at a relatively low cost when compared to whole genome shotgun projects. ESTs are also used as part of the quality control process for genome assembly.

Over the past 2 years, the JGI EST Pipeline has processed 3,000,000 ESTs from 75 organisms. The pipeline is largely automated, and executes a workflow comprised of 44 distinct states. Tasks executed by the EST workflow include: vector trimming, quality trimming, contaminant screening, ab-initio clustering, cluster consensus sequence generation, and cluster consensus sequence annotation. The JGI EST Pipeline provides important feedback to the EST program laboratory process in the form of library quality and diversity statistics.

EST data is highly error prone and redundant, which necessitates a filtering process whereby ESTs from the same gene are grouped together into clusters. The process of Ab-initio clustering involves the comparison of the ESTs with one another, and the generation of clusters based on similarity between the ESTs. The JGI EST Pipeline uses a simple transitive closure graph based clustering approach derived from technology developed for the JAZZ genome assembler.

11. Complete Genome Sequences and Functional Genomics of Green Bacteria

Donald A. Bryant* (dab14@psu.edu),[1] Tao Li,[1] Julia A. Maresca,[1] Aline Gomez Maqueo Chew,[1] Jörg Overmann,[2] Alex Copeland,[3] Susan Lucas,[3] Alla Lapidus,[3] Kerrie Barry,[3] J. Chris Detter,[3] Tijana Glavina,[3] Nancy M. Hammon,[3] Sanjay Israni,[3] Sam Pitluck,[3] Miriam L. Land,[4] Frank W. Larimer,[4] and Paul M. Richardson[3]

[1]Dept. of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA; [2]Institut für Genetik und Mikrobiologie, Ludwig-Maximilians-Universität München, München, Germany; [3]DOE Joint Genome Institute, Walnut Creek, CA; and [4]Oak Ridge National Laboratory, Oak Ridge, TN

Green sulfur bacteria (GSB; or *Chlorobi*) and filamentous anoxygenic phototrophs (FAPs; *Chloroflexi*; formerly the green non-sulfur bacteria) are widespread but poorly characterized phototrophic eubacteria. In anoxic environments GSB can account for more than 80% of the carbon fixation and much of the nitrogen fixation and sulfur oxidation. The genome sequences of 8 *Chlorobi* and 7 *Chloroflexi* are being determined as part of an effort to define the metabolic diversity of anoxygenic phototrophic bacteria. Additionally, the two genomes of the phototrophic consortium, *"Chlorochromatium aggregatum,"* are being determined. Each consortium consists of ~20 GSB cells attached to one β-proteobacterial cell, the "central rod," which has one polar flagellum. The central rod

provides motility to the entire consortium, which is phototactic due to light signals received by the epibiont GSB cells. It is anticipated that a novel mechanism of cell-to-cell communication may exist between these organisms, and the metabolic basis for their co-dependency is unknown.

Ten sequenced GSB genomes range from 2 to 3 Mb and encode about 1700 to 2800 genes. The FAP genomes are larger, about 5.5 to 6 Mb in size, and they encode more than 5000 genes. Moderately thermophilic *Chlorobium tepidum* [Eisen et al., PNAS, 2002] has become an important model GSB for functional genomics studies because it is efficiently and naturally transformable. Using various comparative bioinformatics approaches, we have identified and verified all steps in the carotenoid biosynthetic pathway and all but one step in the bacteriochlorophyll (BChl) *c* biosynthetic pathway. This work has identified more than 10 novel genes and the enzymes they encode. For example, two novel lycopene cyclases have been characterized, which are the first members of a large family of previously unknown lycopene cyclases found in GSB and cyanobacteria. By comparing the genomes of brown-colored GSB, which synthesize BChl *e* and the important biomarker carotenoid isorenieratene, and green-colored GSB, which produce BChl *c* and the carotenoid chlorobactene, we have identified a 6-kb gene cluster encoding 4 genes. This cluster includes a gene required for isorenieratene biosynthesis and seems likely to confer the physiologically and ecologically important trait of "brownness," since a radical SAM enzyme encoded in this cluster is probably responsible for the introduction of the C7 formyl group into BChl *e*. BChl *e* and isorenieratene synthesis are co-distributed, polyphyletic traits, which almost certainly spread among GSB by lateral gene transfer. These pigments confer the ability to grow much deeper in stratified aquatic environments where sulfide concentrations are typically higher. We are currently testing whether these four genes can transform the green-colored *C. tepidum* to "brownness" by introducing this 6-kb gene cassette by transformation into a non-essential chromosomal gene. Finally, *Chlorobium ferrooxidans* is the only known GSB that cannot oxidize sulfide but instead oxidizes ferrous iron to provide the electrons for carbon fixation via the reverse TCA cycle. Comparative analyses confirm that *C. ferrooxidans* lacks all known genes required for thiosulfate, sulfide, polysulfide, and sulfite oxidation except for sulfide quinone reductase, and candidate genes for iron oxidation have been identified.

12. Comparative Genomics and the California Purple Sea Urchin

R. Andrew Cameron* (acameron@caltech.edu) and Eric Davidson
Biology Division and Center for Computational Regulatory Genomics, California Institute of Technology, Pasadena, CA

The extensive knowledge base in cell and developmental biology of the California purple sea urchin plus the remarkably complete echinoderm fossil record afford an excellent opportunity to inform the annotation of the gene set predicted from the assembled genome. Two unique echinoderm features are the water vascular system and calcium carbonate skeleton. It is to be expected that the differentiation gene batteries used in these structures are likely to be de novo constructions from ancestral genes. Other unique biological characteristics of the echinoderms such as catch collagen, odorant receptors, the digestive enzymes used in marine herbivory, echinochrome pigments, body wall components and teeth ought to likewise involve distinctive sets of genes. A number of genome-wide comparisons to the sea urchin sequences using the several well annotated bilaterian genomes currently available further highlight some of these unique features. About 10 of the most prevalent domains in the sea urchin gene models are not as abundant in the other genomes and thus constitute expansions that are, at least, specific to sea urchins if not to the complex of echinoderms and hemichordates that form the sister group to the chordate deuterostomes. Obvious among these are the motifs that correspond to genes encoding innate immunity proteins. These analyses show that the specific complement of genes in the sea urchin genome result largely from expansion and contraction of existing families from the common bilaterian "toolkit" of genes

13. The Genome of Obligate Methylotroph *Methylobacillus flagellatus*

Ludmila Chistoserdova* (milachis@u.washington.edu),[1] Alla Lapidus,[2] Cliff Han,[3] Lynne Goodwin,[3] Liz Saunders,[3] Tom Brettin,[3] Roxanne Tapia,[3] Paul Gilna,[3] Susan Lucas,[2] Paul M. Richardson,[2] and Mary E. Lidstrom[1]
[1]University of Washington, Seattle, WA; [2]DOE Joint Genome Institute, Production Genomics Facility, Walnut Creek, CA; and [3]DOE Joint Genome Institute, Los Alamos National Laboratory, Los Alamos, NM

The complete genome of a model obligate methylotroph, *M. flagellatus* (strain KT) was sequenced. The genome is represented by a single circular chromosome of approximately 3 Mbp potentially encoding a total of 2,766 proteins. Based on genome analysis as well as the results from previous genetic and mutational analyses, methylotrophy is enabled by methanol- and methylamine dehydrogenases, the tetrahydromethanopterin-linked formaldehyde oxidation pathway, the assimilatory and dissimilatory

branches of the ribulose monophosphate cycle, and a formate dehydrogenase. Some of the methylotrophy genes are present in more than one (identical or non-identical) copy. The obligate dependence on single carbon compounds is likely due to the incomplete tricarboxylic acid cycle, as no genes potentially encoding alphaketoglutarate-, malate- or succinate dehydrogenases are identifiable. The genome of *M. flagellatus* was compared, in terms of methylotrophy functions, to the previously sequenced genomes of three methylotrophs: *Methylobacterium extorquens* (Alphaproteobacterium, 7 Mbp), *Methylibium petroleophilum* (Betaproteobacterium, 4 Mbp), and *Methylococcus capsulatus* (Gammaproteobacterium, 3.3 Mbp). Strikingly, metabolically and/or phylogenetically, methylotrophy functions in *M. flagellatus* are more similar to the ones in *M. capsulatus* and *M. extorquens* than to the ones in the more closely related *M. petroleophilum*, providing the first genomic evidence of polyphyletic origin of methylotrophy in Betaproteobacteria.

14. Environmental Genomic Characterization of a Deep Subsurface Microorganism

Dylan Chivian* (dcchivian@lbl.gov),[1,2] Eric J. Alm,[2,3] Fred J. Brockman,[4] Eoin L. Brodie,[1] David E. Culley,[4] Thomas Gihring,[5] Alla Lapidus,[6] Li-Hung Lin,[7] Duane P. Moser,[8] Paul Richardson,[6] Adam P. Arkin,[1,2,9,10] Terry C. Hazen,[1,2] and Tullis C. Onstott[11]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Virtual Institute for Microbial Stress and Survival, Berkeley, CA; [3]Massachusetts Institute of Technology, Cambridge, MA; [4]Pacific Northwest National Laboratory, Richland, WA; [5]Florida State University, Tallahassee, FL; [6]DOE Joint Genome Institute, Walnut Creek, CA; [7]National Taiwan University, Taipei, Taiwan; [8]Desert Research Institute, Las Vegas, NV; [9]University of California, Berkeley, CA; [10]Howard Hughes Medical Institute, Berkeley, CA; and [11]Princeton University, Princeton, NJ

The deep subsurface planktonic community found at depths greater than 1.5 km across central South Africa is dominated by a species of *Firmicutes* that has never been successfully cultivated. A metagenome was assembled from DNA obtained from fracture water emanating from a borehole at 2.8 kilometers depth in a South African Gold mine. The draft assembly is consistent with a single microorganism, and implies that this species is capable of living in isolation in this deep subsurface environment. Genes for dissimilatory sulfate reduction were present in the genome. Based the 16S rRNA gene, the most similar described organism is *Desulfotomaculum kuznetsovii* (~91% identity), making this one of the first sequenced genomes of a dissimilatory sulfate reducing Gram-positive bacterium. In addition to sulfate reduction, the genome indicates that the organism is capable of formate oxidation. The genome also contains: 1) the acetogenic acetyl-CoA (Wood-Ljungdahl) pathway, 2) a partial TCA cycle, 3) a N_2 fixation pathway, 4) genes for sporulation and germination,

5) heat shock proteins, 6) genes for pilus formation, and 7) genes for flagellum formation and chemotaxis. The variety of metabolic pathways and chemotactic capability is not suggestive of a streamlined genome for a sulfate reducer in an energy depleted environment, but rather more consistent with a motile sulfate reducer in an energy rich environment that actively seeks a specific subsurface niche when present and is capable of surviving long periods of time when that niche is absent. The fracture zone geochemistry is consistent with these inferences. The apparent absence of O_2 tolerance genes indicates the organism is an obligate anaerobe consistent with an indigenous origin.

15. Sequencing *Arabidopsis lyrata* and *Capsella rubella*: What We Can Learn from the Genomes of *A. thaliana*'s Close Relatives

Richard M. Clark* (richard.clark@tuebingen.mpg.de),[1] Yalong Guo,[1] Joy M. Bergelson,[2] Justin O. Borevitz,[2] Brandon S. Gaut,[3] Anne E. Hall,[4] Charles H. Langley,[5] Barbara Neuffer,[6] Klaus F. X. Mayer,[7] Magnus Nordborg,[8] Outi Savolainen,[9] Yves van de Peer,[10] Stephen I. Wright,[11] and Detlef Weigel[1]

[1]Dept. of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany; [2]Dept. of Ecology and Evolution, University of Chicago, Chicago, IL; [3]Dept. of Ecology and Evolutionary Biology, University of California, Irvine, CA; [4]Dept. of Molecular Genetics and Cell Biology, The University of Chicago, Chicago, IL; [5]Section of Evolution and Ecology, University of California, Davis, CA; [6]Dept. of Systematic Botany, Universität Osnabrück, Osnabrück, German; [7]Munich Information Center for Protein Sequences, Institute for Bioinformatics, Gesellschaft fur Strahlenforschung Research Center for Environment and Health, Neuherberg, Germany; [8]Molecular and Computational Biology, University of Southern California, Los Angeles, CA; [9]Dept. of Biology, University of Oulu, Oulu, Finland; [10]Dept. of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology (VIB), Ghent University, Ghent, Belgium; and [11]Dept. of Biology, York University, Toronto, Canada

Arabidopsis thaliana has many advantages that make it a powerful system for genetic and genomic studies. These include a short generation time, small genome size (~125 Mb), easy DNA-mediated transformation, and extensive intraspecific polymorphism datasets. While these resources rival those available in other multicellular reference organisms such as *C. elegans*, *D. melanogaster* and *M. musculus*, reference genome sequences from close relatives of *A. thaliana* have been lacking. In order to leverage genomic information available from *A. thaliana*, as well as to better understand plant genome evolution, we have proposed two close relatives for genome sequencing (*A. lyrata* and *Capsella rubella*). Both genomes are now scheduled for ~8-fold shotgun genomic sequencing by

the Department of Energy Joint Genome Institute in 2006. From a large number of close relatives, *A. lyrata* and *C. rubella* were selected for sequencing because of their comparatively small genome sizes (~230 and 250 Mb, respectively), interesting life histories (*e.g.*, selfing versus outcrossing), and existing genetic and genomic resources (*e.g.*, genetic maps and BAC libraries are available for both species). Comparing the *A. thaliana*, *A. lyrata*, and *C. rubella* genome sequences will provide an unprecedented opportunity to understand dynamic mutational processes, such as genome rearrangements and gene duplication/deletions, that appear to occur in plants at much higher rates than in animals. Given the recent divergence times for these species (~5-10 million years), a particular interest will be to understand the basis for the approximately 2-fold difference in genome size between *A. thaliana* and both *A. lyrata* and *C. rubella*. An overview of the project and planned analyses will be presented.

16. The *Daphnia* Genomics Consortium Partnership with the DOE Joint Genome Institute: Towards a Model System for Ecological Genomics

John Colbourne* (jcolbour@cgb.indiana.edu),[1] Michael Pfrender,[2] Kelley Thomas,[3] Justen Andrews,[1] Michael Lynch,[1] and Jeffrey Boore[4]
[1]Indiana University, Bloomington, IN; [2]Utah State University, Logan, UT; [3]University of New Hampshire, Durham, NH; and [4]DOE Joint Genome Institute, Walnut Creek, CA

The freshwater crustacean *Daphnia* has attracted attention from biologists for over two centuries because of its sentinel role within nearly all freshwater ecosystems and its unique habits when responding to environmental challenges. For instance, within just a short time (~4-6 million years) members of the *D. pulex* species complex have adapted to the numerous physical and biological factors that impact life in ephemeral pools, eutrophic lakes, oligotrophic reservoirs, UV radiated ponds and saline habitats. Within yet shorter periods (~100 years), many populations have faced novel stressors from industrial activities. In many cases, the response of *Daphnia* populations involves phenotypic plastic traits that rapidly modify their morphology, physiology and even their mode of reproduction, which are triggered by cues from the environment. Although the *Daphnia* system is ideal for genetic studies on the potential and limits of adaptive evolution, only recently have we produced the necessary tools that can link the structure and expression of the genome to individual fitness and population-level responses of *Daphnia* to environmental stress.

In partnership with the Joint Genome Institute, the *Daphnia* Genomics Consortium has created (1) a microsatellite-based genetic map to identify ecologically important genes by quantitative trait locus (QTL) analyses, (2) microarrays for the functional annotation of *Daphnia* genes and (3) a genome sequence. Expressed sequence tags (ESTs) were also created from

Daphnia exposed to 12 distinct ecological stressors ranging from toxic metals and UV radiation to hypoxia, starvation and predation. So far, 12,600 unique genes from 71,000 high quality ESTs were recovered. Clear patterns exist among the different libraries that highlight classes of loci transcribed under various ecological conditions. Yet overall, the majority of sequenced cDNA from *Daphnia* (>55%) show no similarity to genes from other organisms. Although the candidate gene approach will prove to be useful for ecological studies – by selecting loci that share sequence conservation with functionally well characterized genes of other model species – microarrays that probe for all gene transcripts will undoubtedly be a more valuable tool for finding ecologically important loci.

Acknowledgements: Elizabeth Bohuski, Jeong-Hyeon Choi, Melania Cristescu, Brian Eads, Jelena Radivojac, (Indiana University), Erika Lindquist, Peter Brokstein, Harris Shapiro (JGI), Darren Bauer (University of New Hampshire), Joseph Shaw, Jennifer Davey, Joshua Hamilton, Carol Folt, Celia Chen (Dartmouth College).

This project is supported by the Department of Energy's Office of Science and the National Science Foundation.

17. An Overview of the Capillary Electrophoresis Process at the DOE Joint Genome Institute Production Genomics Facility: The Dual Operation of the AB 3730xl and GE MegaBACE 4500 DNA Sequence Analyzers

Christopher Daum* (daum1@llnl.gov), **Lena Philip***, Danielle Mihalkanin, Cailyn Spurrell, Don Miller, Susan M. Lucas, Alex Copeland, Damon Tighe, Eric Abbott, Marlon Arcaina, Melanie Lafrades, Albert Linkowski, Adrienne Loero-Pequignot, Andy Yuen, and Mathew Zane

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

At the center of the Department of Energy's (DOE) Joint Genome Institute (JGI) Production Genomics Facility (PGF), lies a highly efficient and automated production line devoted to the generation of high-quality genomic DNA sequence. The JGI utilizes a dual platform of DNA sequence analyzers within its Production Capillaries group: seventy Applied Biosystems 3730xl and thirty-six GE Healthcare MegaBACE 4500 instruments. The Capillaries group is comprised of eleven employees that are responsible for operating and maintaining both platforms; the group is also involved in the daily monitoring of performance stats and the troubleshooting of DNA sample and instrument related issues. The operation of these high-throughput fluorescence-based DNA sequence analyzers will be assessed on the strengths and benefits of each platform, including instrument overviews of operational parameters and mechanical/component specifications. In addition, instrument setups for production operation, operation schedules, loading, and maintenance strategies as well as the various sequencing strategies for each platform

will be compared. Throughput numbers and sequencing quality results will 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. W-7405-ENG-36.

18. Phylogenomic Analyses of Cytoskeletal Genes in the Free-Living Basal Eukaryotes *Spironucleus vortens* and *Naegleria gruberi*

Scott Dawson* (sdawson@nature.berkeley.edu), **Lillian Fritz-Laylin*** (fritz-laylin@berkeley.edu), and W. Z. Cande

Molecular and Cell Biology, University of California, Berkeley, CA

Molecular phylogenetic analyses indicate that the majority of the over seventy proposed "kingdom-level" lineages of eukaryotes are actually microbial, and understanding the origins and evolution of the eukaryotes requires genomic data from diverse eukaryotic microbes. Most of our knowledge of biological processes, such as the cell cycle and cell division, and their functional diversity derive from studies of more recently evolved, and phylogenetically-limited, macroscopic eukaryotic model organisms such as fungi and metazoans. For obvious medical reasons, most basal microbial eukaryotes currently studied are parasites (e.g. *Trypanosoma brucei, Giardia intestinalis*). This is our rationale for genome sequencing of a free-living diplomonad, *Spironucleus vortens,* and a heterolobosean, *Naegleria gruberi*. Using pulsed-field gel electrophoresis, we estimated the genome size of *Spironucleus vortens* to be 18Mb, and *Naegleria gruberi* to be about 40Mb. Additionally, as compared to *Giardia*, an intestinal parasite, we have identified more complete biosynthetic pathways of vitamins and amino acids in *Spironucleus* – supporting the notion that *Giardia*'s gene losses in the adaptation to parasitism are mainly metabolic. In terms of the cytoskeleton, the microtubule system in *Spironucleus* appears to contain a complete array of component proteins, including many all tubulin isoforms, and many kinesins and dyneins. Yet in stark contrast, the actin or microfilament system is missing the majority of known proteins other than actin itself. Specifically, we find no evidence for myosin, arps 1, 2, 3, profilin, cofilin, or numerous proteins involved in actin assembly, disassembly and stability. Thus our analyses of two diplomonads – the free-living *Spironucleus* and the parasitic *Giardia* – indicates a lack of actin-associated proteins, implying a later evolution of an elaborated actin cytoskeleton. Further, although *Naegleria* has some cytoskeletal components, there is far less than in other more recently evolved amoebae like *Dictyostelium*. Additional analyses will allow us to identify core features of the eukaryotic cytoskeleton in *Spironucleus* and *Naegleria*.

19. Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate Genome

Paramvir S. Dehal* (psdehal@lbl.gov)[1] and Jeffrey L. Boore[1,2]
[1]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, CA and [2]Dept. of Integrative Biology, University of California, Berkeley, CA

The hypothesis that the relatively large and complex vertebrate genome was created by two ancient, whole genome duplications has been hotly debated, but remains unresolved. We reconstructed the evolutionary relationships of all gene families from the complete gene sets of a tunicate, fish, mouse, and human, then determined when each gene duplicated relative to the evolutionary tree of the organisms. We confirmed the results of earlier studies that there remains little signal of these events in numbers of duplicated genes, gene tree topology, or the number of genes per multigene family. However, when we plotted the genomic map positions of only the subset of paralogous genes that were duplicated prior to the fish-tetrapod split, their global physical organization provides unmistakable evidence of two distinct genome duplication events early in vertebrate evolution indicated by clear patterns of 4-way paralogous regions covering a large part of the human genome. Our results highlight the potential for these large-scale genomic events to have driven the evolutionary success of the vertebrate lineage.

20. "PhIGs" (Phylogenetically Inferred Groups): A Tool for Whole Genome Phylogenetic Analysis

Paramvir Dehal[1] and **Jeffrey L. Boore***[1,2] (jlboore@berkeley.edu)
[1]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, CA, and [1,2]Dept. of Integrative Biology, University of California, Berkeley, CA

We present the PhIGs database, a phylogenomic resource for sequenced genomes. Although many methods exist for clustering gene families, very few attempt to create truly orthologous clusters sharing descent from a single ancestral gene across a range of evolutionary depths. Although these non-phylogenetic gene family clusters have been used broadly for gene annotation, errors are known to be introduced by the artifactual association of slowly evolving paralogs and lack of annotation for those more rapidly evolving. PhIGS automates the clustering of genes into families with a method that respects the known evolutionary relationships among the organisms, creation of evolutionary trees for each cluster using maximum likelihood, the most sophisticated method of evolutionary reconstruction, determination of orthologous and paralogous relationships, and the correlation of this with information on gene function and relative position. We plan further work to link other information, such as gene expression patterns.

The PhIGs database (http://phigs.org) currently contains 23 completely sequenced eukaryotic genomes of fungi and metazoans, containing 409,653 genes that have been grouped into 42,645 gene clusters. Clusters can be searched using keywords on annotations, such as GO and InterPro assignments, and by sequence similarity using BLAST and HMM. Results include graphics that show the evolutionary relationships of all genes within each gene family, color-coded multiple sequence alignments, comparisons of exon-intron structures, links to databases on gene function, and multi-species maps showing the relative positions of sets of large numbers of orthologous genes.

21. The Genome of *Halorhodospira halophila*

Ratnakar H. Deole* and **Wouter D. Hoff***

Oklahoma State University, Stillwater, OK

22. High-Density Universal 16S rRNA Microarray Analysis Reveals Broader Diversity than Typical Clone Library When Sampling the Environment

Todd DeSantis, **Eoin Brodie*** (elbrodie@lbl.gov), Jordan Moberg, Ingrid Zubietta, Yvette Piceno, and Gary Andersen

Molecular Microbial Ecology Group, Ecology Department, Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA

Molecular approaches aimed at broad prokaryotic environmental detection routinely rely upon classifying heterogeneous 16S rRNA genes amplified by universal PCR. The general method of sampling DNA types has been to clone and sequence the PCR products. However, the number of clones required to adequately catalogue the majority of taxa in a sample is unwieldy. Alternatively, hybridizing target sequences to a universal 16S rRNA microarray may provide a more rapid and comprehensive view of prokaryotic community composition. This study investigated the breadth and accuracy of a microarray in detecting diverse 16S rRNA gene sequence types compared to clone-and-sequencing using three environmental samples: urban aerosol, subsurface soil and subsurface water. PCR products generated from universal 16S primers were classified using either the clone-and-sequence method or by hybridization to a novel high-density microarray of 297,851 probes complementary to 842 prokaryotic sub-families. The three clone libraries comprised 1,391 high-quality sequences. Approximately 8% of the clones could not be placed into a known sub-family and were considered novel. The microarray results confirmed the majority of clone-detected sub-families and additionally demonstrated greater amplicon diversity extending into phyla missed by the cloning method. Sequences matching OTUs within the phyla *Nitrospira*, *Plantomycetes,* and TM7, which were uniquely detected

by the array, were verified with specific primers and subsequent amplicon sequencing. Clone-observed richness appeared to be an underestimate when compared to the non-parametric richness predictions, however array-observed richness corresponded well. It was concluded that although the microarray is unreliable in classifying novel prokaryotic taxa, it is more adept at identifying greater diversity in environmental samples than sequencing a typically-sized clone library. Furthermore, the microarray allowed samples to be rapidly evaluated with replication, a significant advantage in studies of microbial ecology.

23. A Multiple Sequence Alignment Server for Comparative Analysis of 16S rRNA

Todd DeSantis,[1,4] Phil Hugenholtz,[2] Keith Keller,[5,4] Eoin Brodie,[1] Niels Larsen,[3] **Yvette Piceno*** (ympiceno@lbl.gov),[1] Richard Phan,[1,4] and Gary Andersen[1,4]

[1]Center for Environmental Biotechnology, Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Microbial Ecology Program, DOE Joint Genome Institute, Walnut Creek, CA; [3]Danish Genome Institute, Aarhus, Denmark; [4]Virtual Institute for Microbial Stress and Survival, Lawrence Berkeley National Laboratory, Berkeley, CA; and [5]Quantitative Biomedical Research, University of California, Berkeley, CA

Microbiologists conducting surveys of bacterial and archaeal diversity often require comparative alignments of thousands of 16S rRNA genes collected from a sample. The computational resources and bioinformatics expertise required to construct such an alignment has inhibited high-throughput analysis. It was hypothesized that an online tool could be developed to efficiently align thousands of 16S rRNA genes via the NAST (Nearest Alignment Space Termination) algorithm for creating multiple sequence alignments (MSA). The tool was implemented with a web-interface at http://greengenes.lbl.gov/cgi-bin/nph-NAST_align.cgi. Each user-submitted sequence is compared to Greengenes' "Core Set", comprising approximately 10,000 aligned non-chimeric sequences representative of the currently recognized diversity among bacteria and archaea. User sequences are oriented and paired with their closest match in the Core Set to serve as a template for inserting gap characters. Non-16S data (sequence from vector or surrounding genomic regions) is conveniently removed in the returned alignment. From the resulting MSA, distance matrices can be calculated for diversity estimates and organisms can be classified by taxonomy. The ability to align and categorize large sequence sets using a simple interface has enabled researchers with various experience levels to obtain prokaryotic community profiles.

24. Coral Reef Genomics: A Genome-Wide Approach to the Study of Coral Symbiosis

Michael DeSalvo* (mdesalvo@ucmerced.edu),[1] Jodi Schwarz,[2] Peter Brokstein,[2] Cindy Lewis,[3] Shruti Lal,[1] Constance Rogers-Lowery,[4] Chitra Manohar,[5] Carol Tang,[6] Alina Szmant,[4] Mary Alice Coffroth,[3] and Mónica Medina[1]

[1]University of California, Merced, CA; [2]DOE Joint Genome Institute, Walnut Creek, CA; [3]State University of New York, Buffalo, NY; [4]University of North Carolina, Wilmington, NC; [5]Lawrence Livermore National Laboratory, Livermore, CA; and [6]California Academy of Sciences, San Francisco, CA

Coral reefs are among the most beautiful and diverse ecosystems on Earth, upon which 500 million people depend for food, coastal protection, and other resources. Yet coral reefs are threatened by the declining health of the marine environment: coral bleaching and coral 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 reef-building corals and reef nutrient and carbon cycles 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 have been used to print cDNA microarrays. cDNA libraries and gene expression microarrays are also being developed for two *Symbiodinium* strains identified to infect the two coral hosts. By experimentally infecting aposymbiotic coral larvae with dinoflagellates, we will see gene expression changes that are correlated with the onset and maintenance of the symbiosis. These observations will provide new knowledge and avenues of research to help protect these endangered organisms. We have also started to design 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 *Montastraea* and *Acropora* Caribbean species and their hybrids; and 5) differences in gene expression within and between populations.

25. The Genome Sequence of *Alkaliphilus metallireducens*, an Alkaliphilic, Metal-Reducing Bacterium

M. W. Fields* (fieldsmw@muohio.edu),[1] J. Zhou,[2] A. Lapidus,[3] C. S. Han,[4] O. Chertkov,[4] and P. M. Richardson[3]

[1]Dept. of Microbiology, Miami University, Oxford, OH; [2]Institute for Environmental Genomics, University of Oklahoma, Norman, OK; [3]Genomics Facility, DOE Joint Genome Institute, Walnut Creek, CA; [4]DOE-Joint Genome Institute, Los Alamos National Laboratory, Los Alamos, NM

Alkaliphilus metalliredigens is an alkaliphilic, iron- and metal-reducing bacterium that was isolated from borate leachate ponds, and the draft genome sequence is now available. *A. metalliredigens* is a low G+C Gram-positive bacterium and is distantly related to other metal-reducing bacteria. The genome is approximately 4.4 Mb in size, and has a 36.6% G+C content. The genome is predicted to contain 4,090 protein-coding ORFs, and approximately 30% do not have function prediction. In the laboratory, growth is supported only in the presence of a metal as a terminal electron acceptor; however, the genome contained few ORFs annotated as cytochromes. Only four flavocytochromes were identified as possible fumarate reductase subunits. An ammonia-forming nitrite reductase and a nitric oxide reductase were predicted, and attempts to cultivate the organism on nitrate or nitrite were not successful. Multiple F-type, V-type, and P-type ATPases were predicted, and the organism is moderately halophilic and was isolated from a site that contained multiple heavy metals. Genes annotated as pyruvate dehydrogenase, pyruvate-ferredoxin oxidoreductase, and pyruvate-formate lyase were observed. In relation to intermediary metabolism, an incomplete glycolysis, an incomplete reductive TCA, and a complete pentose phosphate pathway were detected. Annotated genes involved in butanoate production and possibly propanoate production were observed. Annotated genes for mannitol and fructose-specific PTS genes were observed, and mechanisms for the utilization of mannitol, fructose, sorbose, sorbitol, melibiose, and mannose could be predicted. In addition, the genome was predicted to contain two polysaccharide-degrading enzymes: arabinanase and mannanase. Almost 50 genes were predicted to be involved in sporulation, and spore-formation was detected. A ferrichrome system (enterochelin) was predicted, both the binding protein and permease. In accordance with different metabolisms and dynamic environment, *A. metallireducens* was predicted to have approximately 25 ORFs with PAS domains, 7 of which also contained HTH motifs.

26. Phylogenomics with Single Copy Universal Genes (SCUGs)

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

[1]Evolutionary Genomics Department, DOE Joint Genome Institute, Walnut Creek, CA, and [2]School of Natural Sciences, University of California, Merced, CA

The duplicability, or tendency of a gene family to proliferate within a genome, is very variable. Some genes, mostly those involved in information processing and other complex mechanisms requiring many molecular interactions, are nearly universally found in single copy across all genomes, others tend to form small gene families, and a few appear greatly duplicated in many organisms. We suggest that the set of Single Copy Universal Genes (SCUGs), or genes present in all organisms of a given clade in single copy, is likely to be the most reliable set of genes for phylogenetic purposes.

The utilization of SCUGS will avoid several of the basic problems commonly confounding phylogenetic reconstruction. In addition to the obvious advantage of avoiding hidden paralogy, SCUGs generally evolve at slower rates and are less prone to horizontal gene transfer. Also, genes whose histories involve bursts of duplication in different lineages would be more variable in evolutionary rate through time and across species, because gene duplication is most often followed by accelerations of evolutionary rate in some of the resulting duplicates.

Here we present phylogenetic analyses based on SCUGs for two deep clades of the tree of life: the Bacteria and the Opisthokonta, the eukaryotic lineage that contains animals and fungi. We show that substantial levels of phylogenetic resolution can be obtained for each clade by concatenating alignments of their respective SCUGs. We also present a comparative analysis of the gene constituency of the two SCUG sets, and a study of phylogenetic congruence among different genes.

27. Over and Under-Representation of Sigma-70 Binding Sites in Different Bacterial Genomic Regions

Jeff Froula* (jlfroula@lbl.gov) and M. P. Francino

Evolutionary Genomics, DOE Joint Genome Institute, Walnut Creek, CA

The σ^{70} subunit of the bacterial RNA polymerase holoenzyme is responsible for specific binding to the –10 and –35 promoter motifs. It has been shown that selection acts to remove –10 and –35 consensus sequences in both coding and noncoding regions, implying that it is disadvantageous to maintain misplaced sites that can strongly bind σ^{70} and interfere with proper gene expression. Here we analyze 47 bacterial

genomes and show that the numbers of potential σ^{70} binding sites in noncoding regions deviate significantly from the random expectations based on base composition, di- and tri-nucleotide contents in a majority of eubacteria. This implies that not only is there selection for maintaining high densities of potential σ^{70} binding sites signals in *regulatory* DNA, there is also selection against these sites in *non-regulatory* DNA. This result suggests that these often overlapping binding sites confer some subtle survival advantage although experimental evidence has shown only one or two of these sites are actually transcription initiation sites. Additional results showing that faster growing bacteria have more selection against potential σ^{70} binding sites in non-regulatory DNA is evidence that the efficiency needed for faster growing bacteria can only be achieved by reducing spurious RNA polymerase binding to false sites.

28. Cloning the *Alvinella pompejana* XPD Homolog

Jill O. Fuss* (jfuss@lbl.gov),[1] David S. Shin,[2] Quen J. Cheng,[1] S. Craig Cary,[3] and John A. Tainer[1,2]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Scripps Research Institute, La Jolla, CA; and [3]University of Delaware, Lewes, DE

Alvinella pompejana are one of the earth's most thermophilic multi-cellular eukaryotes. These polychaete annelids, or segmented bristle worms, live in hydrothermal vents 2500 meters below the surface of the Pacific Ocean, where they thrive in an environment of high temperature (20-80°C), low pH, and rich in heavy metals, carbon dioxide, and hydrogen sulfide. Despite living in an environment that is vastly different from our own, our initial analyses of EST sequences generated by the JGI suggest that *A. pompejana* and human genes share high homology. One such example is the *A. pompejana* homolog of the human XPD gene.

XPD is a 5′-3′ helicase that is required for the efficient repair of damaged DNA. Mutations in the XPD gene result in three distinct hereditary diseases: Xeroderma pigmentosum (XP), Cockayne's syndrome (CS) and Trichothiodystrophy (TTD). XP is characterized by extreme sun-sensitivity and susceptibility to skin cancer due to inability to repair UV and chemical DNA damage. CS is a post-natal developmental disorder characterized by severe progressive neurological deterioration. TTD patients have severe neurological deterioration and developmental abnormalities that are distinct from similar symptoms in CS. The spectrum of XPD mutations sheds little light as to the mechanistic cause of XP, CS, or TTD. Disease causing point mutations are spread throughout the length of the protein and even adjacent mutations can give rise to different phenotypes. Efforts to understand XPD structural biology have been hindered by difficulties in expressing recombinant human XPD in sufficient quantities for structural analysis. In addition, XPD has been difficult to model in thermophilic archaea due to poor sequence conservation.

Using EST sequences generated by our FY2005 JGI collaboration, we have cloned the *A. pompejana* homolog of human XPD. A BLAST search of *A. pompejana* ESTs revealed a 700 bp fragment with high homology to the human XPD protein sequence. Using primers complementary to this fragment, we amplified specific 5′ and 3′ RACE products that spanned the length of the *A. pompejana* XPD gene. The protein sequence of ApXPD is highly conserved to human XPD: 75% identity and 89% similarity. Most importantly for understanding human disease, residues that are mutated in human disease are largely conserved in ApXPD.

Understanding the biology of *A. pompejana* will be important not only for understanding human disease, but will also provide insight into how eukaryotes survive in a toxic environment of high temperature, carbon dioxide, and heavy metal stresses.

29. An Overview of the Sequencing Prep Process at the DOE Joint Genome Institute's Production Genomics Facility

Tijana Glavina del Rio* (glavinadelrio1@llnl.gov), Arshi Khan, Megan Kennedy, Zalak Trivedi, Michael Philips, David Robinson, Terri Jackson, and Michael Pintor

DOE Joint Genome Institute, Walnut Creek, CA

The sequencing prep process is the second step of the production process at the DOE Joint Genome Institute (JGI) Production Genomics Facility (PGF). The goal of the process is to prepare labeled fragments in 384-well plates for loading onto the capillary sequencers. The process begins by alliquoting 2µl of glycerol stock from the Library Support Process using Matrix PlateMate Plus. The 384-well plates, containing glycerol stock and buffer, are placed on the PE 9700 thermocyclers and heated to 95°C for 5 minutes in order to lyse the cells and release the plasmids. The amplification of plasmid DNA is performed using Templiphi, a kit made by GE Healthcare in order to generate large quantities of template. TempliPhi is added with a multidrop micro instrument and samples are incubated at 30°C for 20 hours to allow the amplification to occur. The amplified template is then heated to 95°C to inactivate the enzyme. Small amounts of amplified DNA are transferred using a Robbins Hydra Twister robot into two 384-well plates to set up dual sequencing reactions. The Chemistry cocktail is added by the Cavro Dispensing instrument and then cycled. For high GC templates we utilize a 5% final concentration of DMSO that is added to our TempliPhi and chemistry reactions. Before DNA can be loaded onto the capillary sequencer, the leftover reagents, cell debris, buffers, and salts must be removed from the sample. The last step in the Sequencing prep process is the cleanup step. The process uses a modified magnetic bead protocol to purify DNA fragments from the sequencing reaction and get it ready for sequencing. This step is performed using the Beckman Coulter Biomek FX robots. The reactions

are then ready to load on either the ABI 3730xl or the MegaBACE 4500 for capillary electrophoresis. This poster will present a detailed overview of the sequencing prep process reviewing each step and the instruments that are used.

30. Annotation of Eukaryotic Genomes at JGI

Igor V. Grigoriev* (ivgrigoriev@lbl.gov)
DOE Joint Genome Institute, Walnut Creek, CA

31. Automated Microbial Genome Annotation

Loren J. Hauser* (hauserlj@ornl.gov)
Oak Ridge National Laboratory, Oak Ridge, TN

32. The Genome Sequences of the Ethanol Producing Bacteria *Thermoanaerobacter ethanolicus* 39E and *Thermoanaerobacter ethanolicus* X514

C. L. Hemme* (hemmecl@ou.edu),[1,2] M. W. Fields,[3] E. H. Saunders,[4] M. L. Land,[1] L. J. Hauser,[1] A. L. Lapidus,[5] C. S. Han,[4] D. C. Bruce,[4] P. Richardson,[5] and J. Zhou[2]
[1]Oak Ridge National Laboratory, Oak Ridge, TN; [2]The Institute for Environmental Genomics, University of Oklahoma, Norman, OK; [3]Miami University, Oxford, OH; [4]Los Alamos National Laboratory, Los Alamos, NM; and [5]DOE Joint Genome Institute, Walnut Creek, CA

Recent volatility in the global energy markets has sparked renewed interest in the development of alternative fuel technologies to reduce American reliance on fossil fuels. One promising technology that has generated much interest is the generation of bio-produced ethanol from cellulosic materials such as switch grass. To this end, the genomes of two strains of the thermophilic ethanol-producing bacteria *Thermoanaerobacter ethanolicus* have been sequenced. Strain 39E was originally isolated from thermal springs in Yellowstone National Park while the metal-reducing strain X514 was obtained from a deep subsurface location geographically isolated for ~200 mya. The preliminary annotations for these two strains were compared to the annotation for the completed genome of *T. tengcongensis*. All species studied encode systems for the breakdown of cellulose and glycogen. The two *T. ethanolicus* strains also encode systems for the uptake and fermentation of xylose and other pentose sugars derived from xylan and hemicellulose. A β-xylosidase gene implicated in degradation of xylan and hemicellulose was previously cloned from *T. ethanolicus* JW200 and this gene was found to be highly conserved in strain 39E but significantly diverged in

strain X514. The *T. ethanolicus* strains also encode components of the Entner-Doudroff pathway which are lacking in *T. tengcongensis*. Interestingly, strain X514 lacks a methylglyoxal shunt that is present in the other two *Thermoanaerobacter* species. In many fermenting bacteria, the methylglyoxal shunt is employed to limit energy production when carbohydrates are plentiful. Thus, the loss of the shunt in X514 may represent an adaptation of the strain to the presumably energy-poor conditions of the deep subsurface. The preliminary results suggest that all three organisms utilize similar carbon metabolism pathways but that differences in the carbon metabolism and energy flux pathways may have evolved in response to particular environmental pressures.

33. Fish *Hox* Cluster and the Evolution of Regulatory Elements

Simone Hoegg,[1] Jeffrey L. Boore,[2] and **Axel Meyer***[1] (axel.meyer@uni-konstanz.de)

[1]Dept. of Biology, LS Evolutionary Biology, University of Konstanz, Konstanz, Germany and [2]Evolutionary Genomics Program, DOE Joint Genome Institute, Walnut Creek, CA

While tetrapods, as human, mouse and frog (*Xenopus tropicalis*), have a constant number of four Hox clusters, teleost fish have undergone an lineage-specific genome duplication event and, subsequently, experienced gene and cluster losses. Based on current genomic projects, we know the Hox cluster composition of four teleost fish, *Danio rerio* (zebrafish), *Oryzias latipes* (medaka), *Takifugu rubripes* (torafugu) and *Tetraodon nigroviridis* (spotted green pufferfish). All these species possess seven clusters; however, zebrafish has lost the *HoxDb* cluster, while the other species described have lost the *HoxCb* cluster. From this we can conclude that the loss of these clusters happened independently in the Ostariophysii, to which *Danio rerio* belongs and in the Neoteleostei, which are represented by the other three species. Since Hox gene setup in ray-finned fish show much higher variability than in tetrapods further studies for a denser sampling are required.

With ongoing sequencing projects and the development of large insert size libraries (BAC, PAC), research on Hox clusters has moved from counting genes to comparative analyses of the non-coding proportions of these clusters. Hox cluster provide an ideal genomic model region for research on conserved non-coding elements, since the gene clusters are highly condensed, contain relatively few repetitive elements (e.g. SINEs, LINEs, LTRs [long terminal repeats] and DNA transposons) and are conserved over a large evolutionary scale. Therefore, homologous DNA stretches are easy to identify between species and gain and loss of conserved non-coding elements can be traced easily on a phylogeny. Also the content of conserved non-coding sequences seems to be especially high in this genomic region.

Ongoing research projects on e.g. bichir (*Polypterus senegalus*) and coelacanth (*Latimeria menadoensis*) are providing data for basal lineages for which data usually is difficult to obtain. Another sampling gap so far is the order Perciformes, the largest order of ray-finned fish with approximately 10,000 species in 150 families and morphologically highly diverse. So far, coding sequences for *Hox* genes have been described and the complete *HoxAa* cluster for *Oreochromis niloticus* (tilapia). A BAC (Bacterial Artificial Chromosome) library of *Astatotilapia burtoni*, a haplochromine cichlid with a basal position relative to the Lake Victoria radiation, has been constructed in our lab. This enables us to sequence the complete *Hox* clusters. East African cichlids with their high morphological diversity and low genetic distance provide a fascinating study field for possible influence of regulatory evolution on morphology.

34. Genomic Sequencing of the Wood-Rotting Mushroom *Schizophyllum commune*

Stephen Horton* (hortons@union.edu),[1] **Tom Fowler*** (tfowler@siue.edu),[2] Vanessa Bailey,[3] Scott Baker,[3] Allen Gathman,[4] Erika Kothe,[5] Walt Lilly,[4] Jon Magnuson,[3] and Han Wosten[6]

[1]Union College, Schenectady, NY; [2]Southern Illinois University, Edwardsville, IL; [3]Pacific Northwest National Laboratory, Richland, WA; [4]Southeast Missouri State University, Cape Girardeau, MO; [5]Friedrich-Schiller University, Jena, Germany; and [6]Utrecht University, Untrecht, The Netherlands

The filamentous fungus *Schizophyllum commune* belongs to the group of gilled mushrooms that includes the commercially valuable species *Agaricus bisporus* (white button mushroom) and *Pleurotus ostreatus* (oyster mushroom). The worldwide market for these and other mushrooms for use as food, as dietary supplements and nutriceuticals, and in medicine was estimated to be over U.S. $13 billion per year in 1995. Mushrooms from many of these fungi are also the source of important medicinal compounds with activities against tumors, viruses, and other microbes.

S. commune has been utilized extensively as a model system for studying mating-type gene function and mushroom development. It is a genetically tractable system that has well-developed methodologies for both targeted and random gene transformation. The *S. commune* genome is about 36-38 Mb in size, with a G+C content of 57%. 14 chromosomes have been discerned from CHEF gel/hybridization experiments, ranging in size from 1.6 Mbp to 4.7 Mbp. Taken together, these molecular characteristics make *S. commune* an excellent sequencing target. The ability of *S. commune* to degrade complex plant biomass, including lignin, also makes the organism attractive for the DOE-JGI genomic sequencing program. We expect that its' genome will yield an interesting catalog of secreted enzymes that play a role in ligno-cellulose degradation. The organism has demonstrated bioremediation potential with respect to heavy metal uptake. A genome

sequence, combined with already established molecular techniques, would make *S. commune* an ideal model organism for additional studies in metal uptake by fungi.

We expect that the availability of the *S. commune* genome will attract considerable interest in the fungal research community. Genomic comparisons between wood-rotters *S. commune* and *Phanerochaete chrysosporium*, and the dung-fungus *Coprinopsis cinerea* would potentially allow identification of specializations for wood-rotting, as distinct from general saprophytic mechanisms. The comparison of two distinct model organisms for the study of mushroom development (*S. commune* and *C. cinerea*) should distinguish elements common to this process apart from those elements that are species-specific. The *S. commune* genome sequence, in conjunction with these other fungal genomes and the wealth of genetic information about *S. commune*, will be a valuable resource to promote continued and vigorous research in mushroom biology.

35. *Pichia stipitis*: A Native Xylose Fermenting Yeast that Inhabits Beetle Gut

Thomas W. Jeffries* (twjeffri@wisc.edu)[1], Igor Grigoriev,[2] Jane Grimwood,[3] Andrea Aerts,[2] Asaf Salamov,[2] Erika Lindquist,[2] Yong-Su Jin,[4] Volkmar Passoth,[5] José M. Laplaza,[6] and Paul Richardson[2]

[1]USDA, Forest Service, Forest Products Laboratory and University of Wisconsin, Madison, WI; [2]DOE Joint Genome Institute, Walnut Creek, CA; [3]Stanford Human Genome Center, Palo Alto, CA; [4]Dept. of Food Science and Biotechnology, Sungkyunkwan Univeristy, Suwon, Korea; [5]Dept. of Microbiology, Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden; and [6]BioTechnology Development Center, Cargill, Minneapolis, MN

Pichia stipitis Pignal (1967) is an predominantly haploid, heterothallic yeast related to *Candida shehatae* and other pentose metabolizing ascomycetous species. It belongs to a clade that uses an alternative nuclear genetic code in which CUG codes for serine rather than leucine. Its closest relatives are found in the intestines of passalid beetles, which in turn are found grazing on white-rotted hardwood. *P. stipitis* has the highest native capacity for xylose fermentation of any known microbe. Strains of *P. stipitis* are among the best xylose-fermenting yeasts in type culture collections. Fed batch cultures of *P. stipitis* produce up to 47 g/L of ethanol from xylose at 30°C under low aeration conditions. In addition to fermenting D-xylose to ethanol, *P. stipitis* can assimilate cellobiose and will oxidize the lignin-related compounds veratraldehyde and vanillin to their respective alcohols and acids. As such it is an ideal organism for lignocellulose bioconversion.

The entire 15.44 Mbp of the *P. stipitis* genome was shotgun sequenced, finished to high quality, and annotated. *P. stipitis* has 8 chromosomes ranging from 3.5 to 0.97 mbp. The genome assembly was annotated using the JGI Annotation Pipeline. *P. stipitis* release v2.0 includes 5841 gene models supported by EST, cDNA and protein evidence. Average gene**,** transcript and CDS lengths are 1.6 kb, 1.5 kb and 493 amino acids, respectively. Average gene density is 56% with 4204 single exon genes. The genome size, number of genes and CDS lengths are comparable to the numbers found in other sequenced yeast genomes. A set of 19,635 ESTs was sequenced and clustered: 3839 (94%) of the clusters mapped to the genome; 2252 (40%) genes were supported by ESTs. An absolute majority of predicted genes were supported by protein homology including 4879 (84%) with strong homology in other fungi. 4083 (70%) of all predicted genes have a predicted protein domain.

More than 2600 genes have been manually curated using the JGI Genome Portal Tools, and there appear to be major differences between *P. stipitis* and other yeasts in oxidative phosphorylation, fatty acid metabolism and fatty acid synthesis. The sequenced genome (CBS 6054) includes seven family 3 β-glucosidases, a family 10 endo xylanase, one β-mannanase, two exo-1,3 β-glucanases, five cinnamyl alcohol dehydrogenases and 57 transporters in the major facilitator superfamily – including 5 putative xylose transporters. Aside from genes for assimilating a wide variety of lignocellulosic polymers, *P. stipitis* codes for several proteins that enable it to ferment xylose with the production of very little xylitol. These include four NADP-dependent alcohol dehyrogenases, two primary alcohol dehydrogenases and three pyruvate decarboxylases.

In preparing ESTs for annotation, cells were grown on glucose or xylose under aerobic or oxygen limited conditions. Each of these four pools was clustered separately and the resulting hit frequencies provided estimates of gene expression.

36. Comparative Genomics of Two Biomedically Important Marine Actinomycetes

Paul R. Jensen* (pjensen@ucsd.edu) and Bradley S. Moore

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

Actinomycetes are the single most significant source of microbial natural products, accounting for more than one half of all known antibiotics. We have recently discovered a new actinomycete genus (*Salinispora*) whose members are proving to be an important source of novel secondary metabolites including a new treatment for cancer. These actinomycetes reside in marine sediments and display fundamental physiological differences from those that occur on land including an obligate requirement of seawater for growth. JGI is in the process of sequencing

the type strains for the species *S. tropica* and *S. arenicola*. These will be the first marine actinomycetes and, more generally, the first marine Gram-positive bacteria to be sequenced thus providing important new information about an environmentally significant yet poorly studied group of prokaryotes. These two species produce completely different suites of secondary metabolites with a preliminary evaluation of the *S. tropica* draft sequence indicating that as much as 10% of the genome is devoted to secondary metabolite biosynthesis. This includes pathways for the clinically relevant enediyne antitumor agents and salinosporamide A, a potent proteasome inhibitor. The genome sequences will advance our understanding of how secondary metabolites are produced and provide new information about the evolutionary significance of secondary metabolite production.

37. Functional Metagenomics of Methylotrophy

Marina G. Kalyuzhnaya* (mkalyuzh@u.washington.edu),[1] David C. Bruce,[2] Kerrie W. Barry,[3] Paul M. Richardson,[3] Mary E. Lidstrom,[1] and Ludmila Chistoserdova[1]

[1]University of Washington, Seattle, WA; [2]DOE Joint Genome Institute, Los Alamos National Laboratory, Los Alamos, NM; and [3]DOE Joint Genome Institute, Production Genomics Facility, Walnut Creek, CA

Microorganisms form a foundation of Earth's biosphere and are key drivers of global biogeochemical processes. Therefore understanding, monitoring and prediction of environmental changes require comprehensive studies of microbial communities that lay beyond the reach of tradition microbiology. Rapid advances in modern molecular methods and high-throughput genomics, such as the whole genome shotgun (WGS) community sequencing approach, open new ways to study microbial ecology. However, WGS sequencing often results in low sequence coverage, not allowing for sequence assembly, due the high complexity and/or dynamics of natural communities. Here we present a modified WGS sequencing approach (functional metagenomics) that should allow for higher sequence coverage, and potentially for assembling complete genomes from environmental samples. In this approach, a chosen functional group is specifically enriched for, to reduce the community complexity. To test this approach, we focused on the methylotroph community in Lake Washington sediment and employed Stable Isotope Probing (SIP) for enrichment, followed by WGS sequencing. Sediment samples were exposed separately to five different ^{13}C-labeled C_1 compounds: methane, methanol, methylamine, formate and formaldehyde. Total DNA was extracted from each microcosm, and the ^{13}C-labeled fractions were separated from unlabeled DNA by isopycnic centrifugation. ^{13}C-labeled DNA was used to construct five separate shotgun libraries, and these were sequenced at the JGI facility. 42 to 89 Mb of sequence was produced from each microcosm, totaling 372 Mb. These were assembled, and preliminary analysis of community richness was conducted. First

genomic insights into a broad variety of bacterial groups actively metabolizing C_1 compounds in Lake Washington were obtained, and specific shifts in populations in response to specific C_1 compounds were detected. In each microcosm, a finite number of phylotypes were detected, and some of these were present in more than one type of enrichment. Those enrichment patterns potentially reflect the substrate repertoires of methylotroph populations in Lake Washington and their adaptation to specific C_1 compound fluxes in the environment. The results of this project also pointed toward the role in local C_1 cycling for species previously not implicated in this role, such as Betaproteobacteria belonging to *Rhodocyclales* and *Burkholderiales*, or for species with poorly understood ecological function, such as Planctomycetes. In conclusion, the functional metagenomics approach we present is designed to link the phylogenetic diversity of a chosen functional group to its genetic and metabolic potential, and ultimately to its ecological function.

38. Whole Genome Analysis of MTBE-Degrading Beta-Proteobacterium *Methylibium petroleiphilum* Strain PM1

Staci Kane* (kane11@llnl.gov),[1] Anu Chakicherla,[1] Patrick Chain,[1,4] Radomir Schmidt,[2] Maria Shin,[1] Kate Scow,[2] Frank Larimer,[3,4] Susan Lucas,[4] Paul Richardson,[4] and Krassimira Hristova[2]

[1]Lawrence Livermore National Laboratory, Livermore, CA; [2]Dept. of Land Air and Water Resources, University of California, Davis, CA; [3]Genome Analysis Group, Oak Ridge National Laboratory, Oak Ridge, TN; and [4]DOE Joint Genome Institute Production Genomics Facility, Walnut Creek, CA

Methylibium petroleiphilum strain PM1 was one of the first pure culture strains isolated for its ability to completely degrade the noxious fuel oxygenate, methyl tert-butyl ether (MTBE). Strain PM1 can also degrade other gasoline components that are often co-contaminants with MTBE in groundwater, including benzene, toluene, and xylenes (BTX). The whole genome of strain PM1 was sequenced in order to elucidate the novel pathways that this bacterium uses for contaminant biodegradation as well as for its methylotrophic lifestyle. The sequence has revealed some surprising results including the presence of a large ~600 kb plasmid, presence of metal resistance/homeostasis genes and differences in predicted pathways for methylotrophy. Preliminary analysis reveals that the 4.0 Mb circular chromosome contains 3822 predicted genes and has a G+C sequence composition of 69.3% while the plasmid contains 632 predicted genes and has a G+C of 66.2%. This strain contains a single 16S rRNA gene. There are >10 repeat elements in the genome, repeated from 2 to >10 times, some as large as 30-40 kb in size. Furthermore, the plasmid has been characterized in terms of relative representation of functional groups of genes, IS elements, tRNAs and partitioning/origin of replication.

We have performed comparative genomic analysis with both phylogenetic and functional near neighbors of PM1. An overview of strain PM1 and its genome content, together with whole genome analysis and comparative analyses will be presented. The PM1 genomic sequence is available at http://genome.ornl.gov/microbial/rgel/. Strain PM1 serves as a model bacterium for petroleum contaminant degradation in methylotrophic MTBE-degrading bacteria in the environment.

39. Sequencing *Mycosphaerella* and *Cercospora* Species Will Revolutionize the Control of Major Global Threats on Wheat, Banana and Maize

Gert H. J. Kema* (gert.kema@wur.nl),[1] Larry D. Dunkle,[1] Alice C. L. Churchill,[2] Jean Carlier,[3] Andy James,[4] Manoel T. Souza, Jr.,[5] Pedro Crous,[6] Nicolas Roux,[7] Theo A. J. van der Lee,[8] Cees Waalwijk,[8] Erika Lindquist,[9] Jim Bristow,[9] and **Stephen B. Goodwin**[8]*

[1]Plant Research International B.V., Wageningen, The Netherlands; [2]Dept. of Plant Pathology, Cornell University, Ithaca, NY; [3]UMR BGPI, CIRAD, Montpellier, France; [4]CICY, Merida, Mexico; [5]EMBRAPA Genetic Resources and Biotechnology, Brasília, Brazil; [6]Fungal Biodiversity Centre, Utrecht, The Netherlands; [7]INIBAP, Montpellier, France; [8]USDA-ARS/Purdue University, West Lafayette, IN; 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 cereals, citrus, banana, and eucalypts, soft fruits such as strawberry, and horticultural crops including many *Brassica* species. A few species of *Mycosphaerella* cause disease in humans and other vertebrates. The major *Mycosphaerella* plant pathogens include *M. graminicola* of wheat and *M. fijiensis* of banana, which require global annual fungicide inputs of $400 million and $2.5 billion, respectively. Both fungi are very important to the world economy. For example, *M. graminicola* causes more than $275 million in losses annually to U.S. wheat growers, and more than 70% of the fungicides sprayed on wheat in Europe are to combat this pathogen. The International Mycosphaerella Genomics Consortium was started with the goal to sequence the genomes of *M. graminicola* and *M. fijiensis*. The former was chosen due to its genetic tractability, availability of extensive genomic resources, large worldwide research community and phylogenetic distinctiveness from other fungi being sequenced. The latter was chosen because of its worldwide impact on banana production. A joint project between the USDA-ARS/Purdue University and Plant Research International B.V. was initiated to sequence both genomes, along with 40,000 ESTs from both *M. fijiensis* and the related maize pathogen *Cercospora zeae-maydis*. The work was conducted through the Community Sequencing Program sponsored by the U.S. DOE Joint Genome Institute. The initial goals of

the projects are to: assemble 8× genomic shotgun sequences of *M. graminicola* strain IPO323 and *M. fijiensis* strain CIRAD 86; perform automated annotations of these genomic sequences and directed annotations using the ~80,000 ESTs from both *Mycosphaerella* species (the *M. graminicola* set will be made available by Syngenta); and make these sequences available publicly through a series of linked web sites for comparative analyses. Currently, the 8.9× *M. graminicola* sequence (518,271 traces) is available at NCBI. The 289,742 reads were organized into 2962 contigs spanning 37.65 Mb. These contigs were assembled into 187 scaffolds covering 39.05 Mb, giving a revised genome size of 41.2 Mb, slightly larger than estimated previously. Using a combination of gene prediction methods integrated into the JGI annotation pipeline we predicted a total of 11,395 gene models, with 80% of models supported by homology to other proteins. In addition, a draft mitochondrial assembly yielded a 43,962-base scaffold that appears to cover the complete genome. A community-wide annotation effort culminating in an annotation jamboree is anticipated for later during 2006 and will be open to all interested participants.

40. CSP/USDA/NSF–Sponsored Sequencing of *Phytophthora capsici* and Benchmarking of de Novo Genome Sequencing with 454 Technology

Stephen Kingsmore* (sfk@ncgr.org)

National Center for Genome Resources, Santa Fe, NM

41. Determination of the Baseline Proteome of the Halophilic Archaeon *Haloferax volcanii*

P. Aaron Kirkland,[1] Stanley Stevens, Jr.,[2] Jennifer A. C. Busby,[3] and **Julie A. Maupin-Furlow***[1] (jmaupin@ufl.edu)

[1]Dept. of Microbiology and Cell Science, University of Florida, Gainesville, FL; [2]Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, FL; and [3]Translational Research Institute, Dept. of Proteomics, Scripps Florida, Jupiter, FL

The genome sequence of the hemophilic archaeon *Haloferax volcanii* is nearly complete. To facilitate our understanding of this genome, we are determining the baseline proteome of *Hf. volcanii* and comparing this with the proteome deduced from the genome sequence. To this end, we have characterized the proteome of cells as they transition from log to stationary phase. In addition to a parent strain DS70 which is cured of plasmid pHV2, an isogenic proteasome mutant (GG102 *panA*) was analyzed based on the increased number of phosphoproteins detected in the proteome of this mutant strain compared to its parent, as determined by Pro-Q Diamond stained 2D-gels. Proteins of these two strains were

digested with trypsin, and peptides were separated by reverse phase and multidimensional protein identification technology (MudPIT), which incorporates back to back strong cation exchange and reverse phase chromatography steps. Peptide masses were determined by electrospray ionization quadrupole time-of-flight mass spectrometry (ESI-QTOF MS) with the Analyst QS data system (QSTAR) as well as LTQ linear ion trap tandem mass spectrometry (LTQ MS/MS). SEQUEST and MASCOT were used to correlate the mass spectra of peptides with amino acid sequences deduced from the *Hf. volcanii* genomic database. Using these methods, we identified 1,678 proteins out of 4,083 predicted proteins with conservative filtering constraints. Peptides detected included those of proteins ranging in calculated molecular mass from 5 to 243 kDa and were cytosolic, membrane-associated and integral membrane proteins. Many of the proteins (over 250) were predicted to be involved in energy production and metabolism; however, a large number of hypothetical and conserved hypothetical proteins were also detected.

42. Nitrifier Genomics – New Insights into a Simple Catabolic Lifestyle

Martin G. Klotz* (jgiusermeeting@mgklotz.com),[1] Daniel J. Arp,[2] David J. Bergmann,[3] Paul Berube,[4] Peter J. Bottomley,[2] Patrick S. G. Chain,[5,6] Brad Elmore,[7] Amal F. El-Sheikh,[1] Loren J. Hauser,[8] William J. Hickey,[9] Norman G. Hommes,[2] Alan B. Hooper,[7] Miriam L. Land,[6] Frank W. Larimer,[8] Stephanie A. Malfatti,[5,6] Jeanette M. Norton,[10] Amisha T. Poret-Peterson,[1] Luis A. Sayavedra-Soto,[2] Shawn R. Starkenburg,[2] Lisa Y. Stein,[11] Lisa M. Vergez,[5,6] Bess B. Ward,[12] and Xueming Wei[2]

[1]University of Louisville, Louisville, KY; [2]Oregon State University, Corvallis, OR; [3]Black Hills State University, Spearfish, SD; [4]University of Washington, Seattle, WA; [5]Lawrence Livermore National Laboratory, Livermore, CA; [6]DOE Joint Genome Institute, Walnut Creek, CA; [7]University of Minnesota, St. Paul, MN; [8]Oak Ridge National Laboratory, Oak Ridge, TN; [9]University of Wisconsin, Madison, WI; [10]Utah State University, Logan, UT; [11]University of California, Riverside, CA; and [12]Princeton University, Princeton, NJ

Nitrification, denitrification and nitrogen fixation, are the major processes moving inorganic nitrogen through the biogeochemical N cycle, which is carried out predominantly by bacteria. Nitrification is a two-step dissimilative process and involves the transformation of reduced inorganic nitrogen (i.e., ammonia) to more oxidized forms of nitrogen (i.e., nitrate). Ammonia is first oxidized to nitrite by ammonia-oxidizing bacteria (AOB), then nitrite is oxidized to nitrate by nitrite-oxidizing bacteria (NOB). The genome sequence of the AOB *Nitrosomonas europaea* 19718 was obtained by the DOE Joint Genome Institute as one of its first complete genomes and published in 2003. The complete first genome sequence of an NOB, *N. winogradskyi*, was just recently obtained as one

of five additional genomes of nitrifiers that are characteristic for their respective habitats. Of these, three genomes have been completely sequenced and closed, and their analysis is in progress: (A) *Nitrobacter winogradskyi* Nb-255, a facultative chemotrophic NOB that usually resides in consortia with respective AOB in soil, wastewater and freshwater/sediment environments. Highlights: an aerobic-type CO dehydrogenase, a Type I RuBisCO, 2 nitrite oxidoreductases and a glyoxylate shunt; lacks inventory needed for a complete glycolysis pathway (EMP or ED); (B) *Nitrosospira multiformis* ATCC 25196, the typical strain of AOB isolated from agricultural, forest, and grassland soils under both acid and neutral pH conditions in various regions of the world. Highlights: multiple copies of the gene clusters involved in ammonia and hydroxylamine oxidation, single copies of gene clusters encoding rRNAs, a Type I RuBisCO, urea hydrolase, urea amidolyase and a [Ni-Fe] hydrogenase; inventory to express complete central pathways (EMP, TCA, PPP); (C) *Nitrosococcus oceani* ATCC 19707, a worldwide-distributed marine AOB that is a representative of the gammaproteobacterial purple sulfur bacteria. Highlights: inventory to express complete central pathways (EMP, TCA, PPP); lacks a glyoxylate shunt; 2 functional *rrn* operons; one gene cluster each that encodes the capacity to oxidize ammonia and hydroxylamine; two copies each of the genes necessary to assemble functional complexes I and IV as well as ATP synthase (one H^+-dependent F_0F_1-type, one Na^+-dependent V-type). Two genome projects are still in the closing phase: (D) *Nitrosomonas eutropha* C-91, a betaproteobacterial AOB isolated from sewage and *Nitrobacter hamburgensis* X14, an alphaproteobacterial NOB that was isolated from corroded limestone.

The ongoing analysis of these nitrifier genomes have initiated further experimentation and comparative analyses, most of which have led to significant revisions of prior paradigms, which include that: (i) nitrifiers do have a complete tricarboxylic acid cycle; (ii) nitrifiers are capable of glycogen and sucrose synthesis; (iii) nitrifiers have polyphosphate integrated in their energy and detox metabolism; (iv) some but not all nitrifying bacteria have specific catabolic capacity such as hydrogenase, urea hydrolase and amidolyase and CO-dehydrogenase; (v) the extent of nitrification-specific genome inventory has grown beyond AMO and HAO activity-related inventory; (vi) inference into the molecular evolutionary history of nitrification-specific genome inventory points toward its origins in denitrification and sulfur oxidation capacity; and (vii) nitrifier genomes have large regions of repeat elements and are living documents of rampant invasions by bacteriophages.

43. The Compositae Genome Project and Development of a Genechip for Massively Parallel Genotyping and Gene Expression Analysis in Lettuce

Alexander Kozik* (akozik@atgc.org),[1] David G. Caldwell,[1] María José Truco,[1] Dean O. Lavelle,[1] Leah K. McHale,[1] Jason M. Argyris,[1] Marta Matvienko,[1] Oswaldo E. Ochoa,[1] Fallon Chen,[1] Smitha Mathrakott,[1] Kent J. Bradford,[1] Loren Rieseberg,[3] Steve J. Knapp,[4] Rick Kesseli,[2] David W. Still,[5] Jeffrey L. Boore,[6] Allen Van Deynze,[1] and Richard W. Michelmore[1]

[1]Genome Center, University of California, Davis, CA; [2]Dept. of Botany, University of Massachusetts, Boston, MA; [3]Dept. of Biology, Indiana University, Bloomington, IN; [4]CAGT, University of Georgia, Athens, GA; [5]Dept. of Horticulture/Plant and Soil Science, California State Polytechnic University, Pomona, CA; and [6]DOE Joint Genome Institute, Walnut Creek, CA

The initial phase of the Compositae Genome Project (CGP; http://compgenomics.ucdavis.edu) generated over 68,000 ESTs for lettuce (*Lactuca sativa* and *L. serriola*) and 67,000 ESTs for sunflower (*Helianthus annuus, H. paradoxus, H. argophyllus*) resulting in ~19,000 and ~18,000 unigenes respectively. Tools for visualization of the data were developed (http://cgpdb.ucdavis.edu) and the lettuce sequence information was subsequently used for mapping candidate genes using our core mapping population (*L. sativa* cv. Salinas x wild *L. serriola*, RIL F7:8). Candidate genes coincident with loci determining important agricultural traits, particularly disease resistance, are being investigated further. The second phase of the CGP is generating an expanded set of ESTs from five different Lactuca species (*L. sativa, L. serriola, L. saligna, L. virosa* and *L. perennis*) as well as twelve other Compositae species from five different genera. This expanded set of over 500,000 ESTs will be used for comparative genomics studies of rapidly evolving genes involved in domestication and weediness. In parallel, we have been evaluating the phenotypic diversity within and between the targeted species. Using the EST information, we are developing a 6.6 million feature Affymetrix Genechip for robust, massively-parallel genotyping based on Single Feature Polymorphisms (SFPs) as well as gene expression analysis in lettuce. Approximately 200 probes per unigene will be tiled across polymorphic coding and non-coding regions in 2 bp increments for at least ~30,000 unigenes. The design and the analysis tools we are developing should provide greater than a 95% chance of scoring each SFP correctly. In addition, probes will be designed to conserved regions of each gene and therefore may be useful for analysis of heterologous species in the Compositae.

44. Improving Methods and Strategies for Sequencing Complex Genomes

Gerard R. Lazo* (lazo@pw.usda.gov), Yong Q. Gu, Frank M. You, Curt C. Crossman, Devin Coleman-Derr, and Olin D. Anderson

USDA-ARS Western Regional Research Center, Albany, CA

Sequencing technology and the tools by which to process information have improved tremendously over recent years. It is the combination of these technologies which help in adapting protocols by which to sequence complex genomes. Methods for sequencing range from random shotgun sequencing approaches using now traditional capillary electrophoresis methods to full genome sequencing using picolitre pyrosequencing approaches. Regardless of the methods used, highly repetitive regions of the genome can make it difficult to close gaps in attempts to build contiguous sequence data. A now common method used to prepare sequencing of the genome is to map out a sequencing scaffold from large-insert genome libraries using fingerprint profiling techniques, sometimes referred to as high information content fingerprinting (HICF), and assembly of the scaffolds by software (Luo *et al*., 2003). The assembled scaffolds present a directed guide for the sequencing efforts of the genome. A similar methodology was used in the design of a sequencing strategy for the individual clones of the genome library, applying the fingerprint strategy to smaller subclones and using labels to restriction endonuclease fragments cut with enzymes of four base paired recognition sites. The fingerprint profile data was extracted with the software GenoProfiler (You *et al*., 2004), and assembled with the software FPC (Soderlund *et al*., 2000). The method helped to identify and reduce the number of sequences needed to complete sequencing of a BAC clone, normally sequenced using random shotgun sequencing approaches until complete coverage was apparent. In instances where genomes contain a high content of repetitive sequences, this approach helped to leverage efficiency in the sequencing of clones where routine random shotgun sequencing yielded difficult to close gaps. We have tested this method on BAC clones derived from the wheat genome, well known for its complex genome structure.

45. A Functional Genomics Approach to Early Ascidian Development

Patrick Lemaire* (lemaire@ibdm.univ-mrs.fr)

CNRS

46. A Metagenomic Approach to Uncover the Genetic Diversity of Bacterial Degradation of the Plant Hormone Indole 3-Acetic Acid

Johan Leveau* (j.leveau@nioo.knaw.nl) and Saskia Gerards

Netherlands Institute of Ecology (NIOO-KNAW), Heteren, The Netherlands

Microbial manipulation of a host's hormone system represents an intriguing type of interorganismal interaction. *Pseudomonas putida* 1290 has previously[1] been proposed as a model organism for the study of bacterial degradation of the plant hormone indole 3-acetic acid (IAA). In the past, bacteria that can destroy IAA have received little attention, despite their ubiquitous association with plants, their likely impact on plant functioning and health, and their potential as biocontrol agents for the suppression of plant diseases associated with excess IAA production by microbial pathogens. This lack of interest is perhaps best illustrated by the fact that no bacterial genes for the degradation of IAA have yet been identified. Here, we report on the cloning and sequencing of IAA^- degradative genes from *P. putida* 1290, and on the function-based recovery of similar genes from metagenomic DNA libraries. Transposon mutagenesis of *P. putida* 1290 revealed the involvement of a *CbrAB* two-component regulatory system in IAA catabolism and confirmed the role of catechol as an intermediate in the IAA degradation pathway. Complementation of *P. putida* KT2440, which is not able to degrade IAA, with any one of a group of 4 overlapping fosmid clones from a genomic DNA library of *P. putida* 1290 identified a 16-kb DNA fragment that conferred the ability to grow on IAA as a sole source of carbon. Sequence analysis of this DNA fragment exposed several candidate genes coding for the conversion of IAA to catechol, some of which were highly similar to genes with no function assigned yet on the genomes of fully sequenced bacteria including *Burkholderia xenovorans* LB400 and *Novosphingobium aromaticivorans* DSM12444. We recognized that our complementation strategy which exploits IAA^- *P. putida* KT2440 as a host strain could easily be adapted as a functional (meta)genomics approach to recover IAA catabolic genes from (meta)genomic DNA. To this end, we screened several large-insert libraries with DNA from various sources (e.g. from newly isolated bacterial IAA degraders belonging to the species *Pseudomonas*, *Burkholderia*, or *Arthrobacter*, from grassland bulk soil, or from the *Festuca rubra* rhizosphere) for clones that conferred an IAA^+ phenotype to *P. putida* KT2440. Already, we have identified >40 such clones, and analysis of their DNA inserts will greatly advance our understanding of the types and diversity of genes underlying bacterial IAA metabolism and will provide the currently non-existing but much needed framework for a genetic approach to the phenomenon of bacterial IAA degradation and its role in plant-microbe interactions.

1. Leveau and Lindow (2005). Applied and Environmental Microbiology 71: 2365-2371.

47. Preliminary Insights from Comparison of Five Newly Sequenced Hyperthermophile Crenarchaea

Todd M. Lowe* (lowe@soe.ucsc.edu)
University of California, Santa Cruz, CA

48. Toward Resolution of Green Plant Phylogeny

Dina F. Mandoli* (mandoli@u.washington.edu),[1] Jeffrey L. Boore,[2] Karin Everett,[1] Michael J. Donoghue,[6] Kenneth G. Karol,[1] Jennifer Kueh[1,2] Brent Mishler,[3] Charlie O'Kelley,[7] Richard Olmstead,[1] Alan R. Smith,[3] Karen Renzaglia,[4] and Paul Wolf[5]

[1]Dept of Biology, University of Washington, Seattle, WA; [2]DOE Joint Genome Institute Production Center, Walnut Creek, CA; [3]University Herbarium, University of California, Berkeley, CA; [4]Dept of Plant Biology, Southern Illinois University, Carbondale, IL; [5]Dept of Biology, Utah State University, Logan, UT; [6]Dept. of Ecology and Evolutionary Biology, Yale University, New Haven, CT; and [7]Bigelow Laboratories for Ocean Sciences, West Boothbay Harbor, ME

We are funded to resolve the primary pattern of evolutionary diversification among green plants, and to establish a model for doing so that will be applicable to other groups of organisms with long evolutionary history. To achieve this goal we are 1) completing a matrix of whole genome sequences for chloroplast and mitochondria; 2) producing a comprehensive set of comparable morphological and ultrastructural data for these same taxa; and 3) incorporating inferences from across the phylogenetic hierarchy in green plants using methods designed to permit scaling across studies. We shall indicate how this work will link to other research being conducted on green plants at various scales, especially the concatenation of our data sets with theirs. We will share the methods we are using, primarily FACS and RCA, to isolate and clone genomes, show the genomes we have assembled and relay the implications of these genomes to the phylogeny of the Chlorophytes. We will list genomes we are working on and introduce new taxa we have found. We will summarize our morphological analyses and discuss advances incorporating inferences from across the phylogenetic hierarchy. Funded by NSF's Tree of Life Program, 2002-2006, DEB #0228655 (lead institution), #228432, #0228679, #0228729, #0228660, #0228576.

49. High-Throughput Protein Production via Cell-Free Methods

Meng L. Markillie* (meng.markillie@pnl.gov) and Eric Ackerman
Pacific Northwest National Laboratory, Richland, WA

Developments in disparate fields now enable high-throughput protein expression via custom and commercially-available robots for cell-free, protein expression using wheat-germ or *E. coli* extracts. We have a robot capable of producing 384 proteins per ~20 hour run in quantities of ~50 to 500 ug/protein using PCR templates or plasmids. The major expense per run is incurred by oligonucleotides and PCR reagents —— not the wheat-germ cell free extracts producing proteins directly from transcribed PCR templates, or if desired, from more expensive plasmids. Disparate sets of proteins can be targeted for expression, characterization, and optimization. The overall goal of this project is to develop and deploy scaleable capabilities for high-throughput production of active proteins utilizing cell-free methods. Protein targets will be chosen based on several criteria, but a dominant theme will be to support and extend our discovery that enzymes immobilized in properly chosen and configured nanoporous materials display remarkable properties of enhanced activity and stability of use for DOE missions in remediation, energy production, and security.

50. Sugarbeet Breeding, Genetics, and Genomics Progress

Mitch McGrath* (mitchmcg@msu.edu)
USDA-ARS and Crop and Soil Sciences, Michigan State University, East Lansing, MI

Sugarbeet is an industrial crop whose main product is sucrose, which constitutes 50 - 70% of the biomass produced during its cool temperate growing season. This simple, easily digestible, carbon source is primarily used as a sweetener, and increasingly to provide a concentrated carbon source during fermentation (e.g. ethanol fuel production). New uses for beets are being sought globally, and all parts of the plant already have an economic outlet. Expanding beet breeding to energy uses was a target during the first energy crisis, and remains a viable option today. The key will be to improve biomass production. Preliminary analyses suggest genetic variability for biomass production is available, controlled by relatively few QTL. Identifying these genes is a goal, however the tools available to dissect traits in beets to the molecular level are somewhat rudimentary. Current genomic efforts are geared to build into a complete genome sequence from which genetic markers, gene identification, and genome organization as a representative of the important and widely adaptable Chenopod plant family can be developed.

51. Genome of *Syntrophus aciditrophicus*: Lifestyle of the Thermodynamically Challenged

Michael J. McInerney* (mcinerney@ou.edu),[1] Robert Gunsalus,[2] John Campbell,[3] Lars Rohlin,[2] Housna Moutakki,[1] UnMi Kim,[2] Rebecca Krupp,[2] Luis Rio-Hernandez,[1] Jessica Sieber,[1] and Christopher G. Struchtemyer[1]

[1]Dept. of Botany and Microbiology, University of Oklahoma, Norman, OK; [2]Dept. of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, CA; and [3]Integrated Genomics, Chicago, IL

The completed genome sequence of the fatty- and aromatic-degrading, syntrophic bacterium, *Syntrophus aciditrophicus,* provides the first glimpse of the composition and architecture of the electron transfer and energy transducing systems needed to exist on marginal energy economies by a syntrophic lifestyle. Syntrophy is a thermodynamically based interaction, where the degradation of the primary substrate is thermodynamically unfavorable unless a second species maintains H_2 and/or formate concentrations produced by the first species at very low levels. Syntrophic metabolism proceeds close to thermodynamic equilibrium and requires reverse electron transfer with large negative changes in electrical potential (ΔE) to form H_2 and formate. The genome of *S. aciditrophicus* contains 3,179,300 base pairs in a single circular chromosome. Approximately 3,100 genes were identified and annotated functions were assigned to 1,618 genes (51%). Two potential regions of horizontal transfer were identified on the basis of deviation from the 51% average GC composition. One region corresponds to a Mu-like prophage; the other anomalous region includes a potential conjugative element. The genes were organized into pathways based on the functional assignments and all pathways considered essential for viability of a typical gram-negative organism appear to be present in *S. aciditrophicus*, with a few notable exceptions. The genome does not appear to encode any recognizable carbohydrate transporters for mono-, di- or even oligo saccharides. Identifiable transporters for amino acids and small peptides are also extremely limited. The lack of transport of carbohydrates, amino acids and peptides is reflected in the paucity of degradation pathways associated with life on these substrates. Interestingly, no genes homologous to the multifunctional isomerase-hydrolase-dehydrogenase (*fadB*) subunit, nor to the thiolase (*fadA*) subunit of the common bacterial fatty acid degradation pathway, are recognizable in the genome. This is particularly odd since fatty acids and aromatic substrates are growth substrates for the organism. The genome also lacks homologs for key reactions in known pathways for 2-oxoglutarate synthesis and it may be made by a novel pathway involving glutaconyl-CoA and 2-hydroxyglutaryl-CoA. Proton and sodium gradients can be formed by membrane-bound pyrophosphatases, glutaconyl-CoA decarboxlase, and proton- and sodium-dependent ATP synthetases. The ion gradients can be coupled to electron transfer by a variety of integral membrane components

such as menaquinone, Na^+-translocating NADH:quinone oxidoreductase, and a putative Na^+-translocating channel. The genome has five gene clusters with homologs to heterodisulfide reductase complexes found in Archaea; four of which also contain homologs to archaeal aldehyde oxidoreductases. Two of these clusters may be involved in aromatic ring reduction based on similarity to the newly detected, aromatic ring reductase genes in *Geobacter metallireducens*. Surprisingly, for a fermentative anaerobe, the genome of *S. aciditrophicus* lacks gene homologous to those for acetate kinase. Nine genes predicted to encode for ADP-forming acetyl-CoA synthetases, which synthesize ATP from acetyl-CoA, ADP and P_i are present. The genome sequence of *S. aciditrophicus* suggests a unique strategy for growth with fatty and aromatic acids by a syntrophic lifestyle.

52. The Genomes of Five PAH Degrading Mycobacteria

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

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

Polycyclic aromatic hydrocarbons (PAHs) are organic compounds consisting of two or more fused benzene rings. Because of their toxic, mutagenic, and carcinogenic properties, 16 PAHs are listed on the U.S. Environmental Protection Agency's priority-pollutant list. We have identified five fast growing mycobacterium isolates that degrade high molecular weight PAHs to water and carbon dioxide. Such isolates have potential in remediation of PAH-polluted sites. Mycobacterium isolates JLS, KMS and MCS were from a PAH-contaminated land-farm soil at the EPA superfund site at Libby Montana, where active remediation was being demonstrated. *M. vanbaalenii* isolate PYR-1 and *M. flavescens* were from Dr. Carl Cerniglia, being isolated from a PAH-contaminated watershed at Redfish Bay near Port Aransas, and from the Grand Calumet River sediment in northwest Indiana, respectively. Pulse field gel electrophoresis identified linear plasmids in *M.* KMS, *M.* MCS and *M. flavescens.* Draft level sequencing completed in Fall 2005, of the five PAH-degrading mycobacteria by the Joint Genomes Institute revealed genomes sizes ranging from 5.9 to 6.4 Mb and in 16 to 45 contigs. These draft level sequences were approximately 68% GC and contained between 5559 and 6012 potential open reading frames. The size of the PAH-degrading genomes is larger than the genomes for pathogenic mycobacterium isolates (about 4.4 Mb for *M. tuberculosis* H37 RV, *M. tuberculosis* CDC1551, *M. bovis* AF2122/97, *M. microti,* and 3.7 Mb *M. leprae*) but similar to *M. avium* (6.6 Mb). *M. smegmatis,* a non-pathogen, sequenced because it has been used as a surrogate for the pathogenic mycobacteria, has a faster growth time like the PAH-degraders and a similar genome size (6.9 Mb). We are initiating categorizing the sets of genes that contribute to survival of the PAH-degrading mycobacterium

in soils/water versus the genes responsible for the pathogenic niche of pathogenic mycobacteria. Additionally, we will investigate the genes involved in the PAH-degradative pathways for comparison to pathways in other microbial genera.

53. Genomic Diversity of *Oenococcus oeni* Strains Isolated from Different Winemaking Regions

David A. Mills* (damills@ucdavis.edu) and Angela Marcobal
Dept. of Viticulture and Enology, University of California, Davis, CA

Oenococcus oeni is a unique lactic acid bacteria isolated solely from fruit-mash and winery environments. Unlike many other lactic acid bacteria, *O. oeni* is tolerant to both high levels of alcohol and low pH conditions making it ideal for performing the malolactic conversion, a key secondary fermentation in the production of wine. In collaboration with DOE Joint Genome Institute, the genome sequence of *O. oeni* PSU-1 has been determined. The complete PSU-1 genome is 1,780,517 nucleotides with a GC content of 38%. 1701 ORFs could be predicted from the sequence of which 75% were functionally classified. Consistent with its classification as an obligately heterofermentative lactic acid bacteria PSU-1 genome encodes all the enzymes for the phosphoketolase pathway. Genes and pathways involved to production of flavor compounds in wine, such as the malolactic conversion, citrate utilization, glycoside activity and catabolism of amino acids were readily identified. Array-based comparisons of ten *O. oeni* isolates obtained from diverse winemaking regions indicate a high level of genome conservation with all strains possessing at least 96% of the genes present in PSU-1. Strains appeared to cluster into three groups, with no relationship between geographic locale and cluster membership. Interestingly, PSU-1, an isolate obtained from a spontaneous malolactic fermentation in Pennsylvania was shown to be nearly identical to MCW, a well-known *O. oeni* isolate obtained from a spontaneous malolactic fermentation at Matanzas Creek Winery in California. The lack of diversity apparent in the *O. oeni* species suggests a wide distribution of strains coincident with winemaking practice. The completion of the *O. oeni* genome marks a significant new phase for research on the malolactic fermentation in wine whereby the physiology, diversity and performance of *O. oeni* starter cultures can be more rigorously examined.

54. The Green Tree of Life

Brent D. Mishler* (bmishler@calmail.berkeley.edu)
University of California, Berkeley, CA
http://ucjeps.berkeley.edu/TreeofLife/

55. The Maltose Transporter Genes of *Thermotoga maritima* are Unusually Close Homologs of those from Gamma Proteobacteria

Dhaval M. Nanavati, Pascal Lapierre, J. Peter Gogarten, and **Kenneth M. Noll*** (kenneth.noll@uconn.edu)

Dept. of Molecular and Cell Biology, University of Connecticut, Storrs, CT

The corresponding proteins encoded by the two *mal* ABC transporter operons of the hyperthermophilic bacterium *Thermotoga maritima* show very high sequence similarities with one another. To determine the origin of these two operons, we carried out phylogenetic analyses of the sequences of the homologous MalE, MalF, and MalG proteins. MrBayes, parsimony and distance matrix methods showed each of these genes are most closely related to their paralog in *T. maritima* suggesting that the entire operon duplicated or that an ortholog was inherited *via* horizontal transfer. These phylogenetic analyses grouped MalF and MalG sequences into nearly identical phylogenetic relationships as those found with MalE suggesting that these genes evolved together. A phylogenetic analysis of all MalE homologs showed that the *T. maritima* homologs are most closely related to one another and represent only one of two major lineages of MalE proteins. An apparent ancient duplication of *malE* occurred and some members of the *Thermococcales* now harbor a representative of each lineage. The *malF* and *malG* lineages show no evidence of an analogous duplication event. Our analyses show that the *T. maritima* Mal protein sequences group with their respective homologs from enteric bacteria. In contrast to the MalEFG homologs, an apparent *T. maritima* MalK homolog (the ATP-binding protein) is more closely related to archaeal MalK homologs from the *Thermococcales*.

Putative structural features observed in the *T. maritima* MalF and MalG protein sequences support the phylogenetic placements for these proteins. The sequence of the *T. maritima* MalF1 homolog suggests it contains eight transmembrane helices and a large periplasmic loop similar to one found uniquely in the MalF proteins of *E. coli* and its close relatives. Hydrophilic domains of both *T. maritima* MalG proteins are predicted to have coiled coil secondary structures similar to those domains in chromosomal maintenance proteins (COG1196). This domain is unique to the *T. maritima* and *Thermotoga neapolitana* MalG homologs indicating that the *mal* genes were acquired only once by their ancestor.

We also have identified both *mal* operons in *Thermotoga sp*. strain RQ2 indicating that these *mal* operons appeared in the *Thermotogales* at least by the time the ancestor of the *Thermotoga* branch emerged. Our data suggest that the *T. maritima malEFG* genes are only distantly related to their homologs in archaeal hyperthermophiles and that they have been shared among a small number of bacterial lineages including the enteric bacteria.

56. The *Calyptogena magnifica* Symbiont Draft Genome: An Obligate, Maternally Transmitted Endosymbiont with Extensive Metabolic Capabilities

I. L. G. Newton* (irene.newton@gmail.com),[1] F. J. Stewart,[1] T. Woyke,[2] P. M. Richardson,[2] K. W. Barry,[2] J. C. Detter,[3] D. C. Bruce,[3] S. Sullivan,[4] J. A. Eisen,[4] and C. M. Cavanaugh[1]

[1]Organismic and Evolutionary Biology Department, Harvard University, Cambridge, MA; [2]DOE Joint Genome Institute, Production Genomics Facility, Walnut Creek, CA; [3]DOE Joint Genome Institute, Los Alamos National Laboratory, Los Alamos, NM; and [4]The Institute for Genomic Research, Rockville, MD

Chemoautotrophic endosymbionts form the metabolic foundation for communities at deep-sea hydrothermal vents and cold seeps. These γ-proteobacteria use reduced sulfur compounds as an energy source for carbon fixation, providing their hosts (e.g., the vestimentiferan *Riftia pachyptila*) the majority of their carbon requirements. At hydrothermal vents, the giant clam, *Calyptogena magnifica,* harbors chemoautotrophic bacteria in its gills. These bacterial symbionts are transmitted via the eggs each generation, and the mutualism between clam and bacterium is obligate —— neither has been found alone in the natural environment and the symbionts have yet to be cultured. Given that this symbiosis is maternally transmitted and is relatively old (between 50 and 100 MY), we hypothesized that the *C. magnifica* symbiont would have a reduced genome with skewed base composition, as has been found in insect endosymbiont genomes as well as most obligate intracellular species. The genome of the *C. magnifica* symbiont is being sequenced from a mixed gDNA sample (host and symbionts) by the DOE Joint Genome Institute and currently constitutes 3-5 contigs at 14x coverage. The draft of the symbiont genome is AT rich (35% GC) and estimated to be relatively small (~1.2Mb). Although reduced in size, the *C. magnifica* genome contains not only the chemoautotrophic pathways thought to be central to the symbiosis (e.g., Calvin cycle enzymes and sulfur oxidation pathways) but also many of the metabolic pathways of free-living chemoautotrophs (e.g., biosynthetic pathways for essential amino acids). This is the first description of a chemoautotrophic symbiont genome and our analyses reveal the extent of host-symbiont metabolic dependency and the evolutionary pressures on the symbiont genome. The study of extant bacterial-eukaryote symbioses will allow us to better understand how metabolic symbioses form and evolve and provide insights into the evolution of eukaryotic organelles such as mitochondria and chloroplasts.

57. Gene Expression and Genomic Analysis of the *in situ* *Geobacteraceae* during Uranium Bioremediation

Regina O'Neil* (rtarallo@microbio.umass.edu), Dawn Holmes, Lorrie Adams, Juliana Larrahando, Joy Ward, Helen Vrionis, Kelvin Gregory, Kelly Nevin, Lucie N'Guessan, and Derek Lovley

University of Massachusetts, Amherst, MA

Geobacteraceae have been shown to be the dominant microorganisms during bioremediation of a U(VI) contaminated aquifer in Rifle, CO when Fe(III) and U(VI) reduction are stimulated by the addition of acetate. Molecular studies have shown that the diversity of 16S rRNA gene sequences detected in the groundwater and sediments during the U(VI) bioremediation process is extremely low. For example, *Geobacteraceae* with 16S rRNA gene sequences closely related to the subsurface isolate *Geobacter bemidjiensis* accounted for 98% of the microbial community, and these sequences were 97-100% similar to each other. Four methods are currently being used to obtain genomic information from the dominant *Geobacteraceae* at this site: 1) shotgun sequencing of environmental genomic DNA, 2) sequencing of the genome from a single cell amplified by multiple displacement amplification, 3) gene specific sequencing using targeted degenerate *Geobacteraceae* primers, and 4) sequencing of isolates which have an exact 16S rDNA match to the predominant *Geobacteraceae*. Significant progress has been made with these methods. Shotgun sequencing of environmental genomic samples obtained during the peak of U(VI) bioremediation, when *Geobacteraceae* accounted for 99% of the microbial population, is currently underway. In addition, methods to isolate and amplify the genome of a single *Geobacteraceae* cell from the environment have also been developed, resulting in the ability to amplify over 80% of the genome from this single environmental cell. Information obtained from the targeted gene amplification approach has also been valuable, and has indicated that a number of genes that are conserved among the seven sequenced *Geobacteraceae* genomes are also found in the environment. Techniques have also been developed to isolate environmentally relevant *Geobacter* species that have 16S rRNA gene sequences that are identical to the dominant *Geobacteraceae* during uranium bioremediation.

The genomic information obtained from these approaches is currently being used to analyze the metabolic state of *Geobacteraceae* during uranium bioremediation. Quantitative RT-PCR primers targeting specific genes from the dominant *Geobacteraceae* were designed from genomic sequences obtained from the environment and used to monitor gene expression during the process of uranium bioremediation. The results to date suggest that there is a correlation between increased reduction of Fe(III) and U(VI) in the subsurface and the expression of specific *Geobacteraceae* genes.

58. The rRNA-Based Map of Biological Diversity

Norman R. Pace* (nrpace@colorado.edu),[1] J. Kirk Harris,[1] Jeffrey J. Walker,[1] John R. Spear,[2] Ruth E. Ley,[3] Phillip Hugenholtz,[4] and Falk Warnecke[4]

[1]University of Colorado, Boulder, CO; [2]Colorado School of Mines, Golden, CO; [3]Washington University School of Medicine, St. Louis, MO; and [4]DOE Joint Genome Institute, Walnut Creek, CA

Our understanding of the extent of microbial diversity has expanded dramatically over the past decade. In large part this expansion has resulted from analysis of ribosomal RNA sequences from uncultured environmental microbes. The number of known main phyla, phylogenetic divisions, of bacteria has increased from 12 in 1987 to currently ~100. Only about one-quarter of these kingdom-level clades have cultured representation and it is common that the members of these uncultured clades are dominant in their respective environments. Thus, the breadth of diversity represented by environmental sequences far outreaches that of biological experience through cultivation. The number of environmental sequences in the databases continues to expand exponentially and now exceeds the numbers from cultivars. Results with Bacteria, Eucarya and Archaea have refined substantially our view of the large-scale Tree of Life. A challenge at this time is to learn the phenotypic natures of uncultured microorganisms.

Although molecular phylogeny with rRNA gene sequences provides some perspective on the evolutionary course of the genetic machinery, rRNA does not necessarily speak to phenotype as manifest in the environment. Much overtly expressed phenotype depends on genes that come and go through lateral gene transfer or propagation through gene families. Thus, for correlations of phenotype (ecotype) with phylotype, as seen by rRNA sequence, it will be necessary to cull phenotypic information from studies such as metagenomic sequencing efforts. Current efforts of the laboratory with JGI involve sequencing deeply into the community rRNA genes of the Guerrero Negro hypersaline microbial mat and shotgun sequencing community DNA of this complex, laminated ecosystem.

59. High-Throughput Assembly

Jasmyn Pangilinan* (jlpangilinan@lbl.gov)

DOE Joint Genome Institute, Walnut Creek, CA

60. Genomic Analysis of the Aromatic Catabolic Pathways of *Cupriavidus necator* (ex *Ralstonia eutropha)* Provides Clues for Adaptation to Aromatic Compound Catabolism in Bacteria

D. **Perez-Pantoja**,[1] R. De la Iglesia,[1] J. E. Gonzalez-Pastor,[2] V. de Lorenzo,[2] D. H. Pieper,[3] and **B**. **Gonzalez***[1] (bgonzalez@bio.puc.cl)

[1]Depto. Genetica Molecular y Microbiologia y EMBA Millennium Nucleus. FCB. P. Universidad Catolica de Chile, Chile; [2]Centro de Astrobiologia (CSIC-INTA), Spain; and [3]Division of Microbiology, GBF, Braunschweig, Germany

The complete genome sequence of *C. necator* JMP134 (ex *R. eutropha),* the best known (halo) aromatic compounds degrading beta proteobacterial strain, is available. We performed the metabolic reconstruction of aromatic compounds degradation, correlating the catabolic abilities as determined *in silico* with a complete study of the range of compounds that support growth of this bacterium. More than 50 out of 160 tested aromatic compounds were used as a sole carbon source. Catabolic genes encoding the degradation pathways for aromatics were identified by BLAST using published gene sequences as *in silico* probes. We observed a strong correlation between the catabolic abilities found *in silico* with those determined *in vivo*. At least ten central ring-cleavage pathways for aromatics, described so far in the class Proteobacteria, were encoded in this genome. Functionality of most of the predicted catabolic pathways was verified with a 50-mer oligonucleotide DNA microarray including 350 catabolic genes of *C. necator*. On the other hand, strict hierarchy of use in aromatics mixtures was observed for most chromosomally encoded functions. Gene redundancy expressed as repeated complete operons, gene encoding peripheral or central reactions, and gene regulators, is a key feature in catabolic versatility. This work confirmed the potential of complete genome sequencing to predict the metabolic abilities of microorganisms and the utility of microarrays to confirm metabolic predictions.

Work supported by a FONDECYT grant (1030493), the MIFAB Millennium Institute, the EMBA Millennium Nucleus, the DOE-JGI initiative and the ICA4-CT-2002-10011 (ACCESS) Contract of the European Union.

61. Soybean Rust Genome Project

Martha L. Posada-Buitrago* (mlposada-buitrago@lbl.gov),[1] Jeffrey L. Boore,[1] and Reid D. Frederick[2]

[1]DOE Joint Genome Institute, Walnut Creek, CA, and [2]USDA-ARS Foreign Disease-Weed Science Research Unit, Fort Detrick, MD

Soybean rust, the most devastating foliar disease responsible for significant losses of soybean crop in Africa, Asia, Australia and South

America, is caused by the fungal pathogen *Phakopsora pachyrhizi*. The pathogen was found for the first time in the continental U.S.A. in November 2004 and has spread to more than 130 counties in the Southeastern States, becoming a major threat to U.S. soybean production.

Here, we present an update on the *Phakopsora pachyrhizi* Genome Sequencing Project funded by the U.S. Department of Agriculture, Agricultural Research Service; the U.S. Department of Defense; and the U.S. Department of Energy.

62. Jazz – The JGI Genome Assembler

Nik Putnam,[1] Jarrod Chapman,[1] Dan Rokhsar,[1] **Isaac Ho*** (iyho@lbl.gov),[2] Serge Dusheyko,[2] and Craig Furman[3]

[1]DOE Joint Genome Institute, Dept. of Physics, University of California, Berkeley, CA; [2]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA; and [3]DOE Joint Genome Institute, Lawrence Livermore National Laboratory, Livermore, CA

Jazz, originally developed by graduate students Nik Putnam and Jarrod Chapman under the direction of Dr. Dan Rokshar, is a whole-genome shotgun assembler designed to handle paired-end sequencing data sets both small and large. It has been used in several high-profile assemblies, notably *Ciona intenstinalis*, *Fugu rubripes*, and *Xenopus tropicalis.*

Jazz relies on the overlap-layout-consensus paradigm made prominent by the Celera assembler; several algorithmic innovations distinguish Jazz from similar products, including evaluation of sister pairing early in the layout process, a hybrid depth/percent-id alignment scoring system, and a classical spring model for resolving complex scaffold topologies. Future directions of research include explicit handling of polymorphism, adaptation to cutting-edge prototype sequencing technologies, and the seamless integration of various forms of linking data such as restriction maps.

63. Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium beijerinckii*

Flavia A. Ramirez* (framirez@uiuc.edu)

University of Illinois, Urbana-Champaign, IL

64. Massive Horizontal Gene Transfer in the *Amborella* Mitochondrial Genome

Aaron O. Richardson* (arichard@bio.indiana.edu)
Indiana University, Bloomington, IN

65. Genotypic Analysis of European and U.S. Populations of *Phytophthora ramorum* by Microsatellite Markers Derived from Genome Sequence

Noah Rosenzweig* (noahr@nature.berkeley.edu) and Matteo Garbelotto
Dept. of ESPM-ES, University of California, Berkeley, CA

Sudden oak death (SOD) caused by a newly described *Phytophthora* species (*Phytophthora ramorum*), has recently become a major disease of concern in western U.S. coastal forest ecosystems and European (EU) nurseries and private estates. Genetic analysis was used to elucidate the genetic structure of the pathogen in its known worldwide range, with the aim of understanding its population biology and its migration patterns. Microsatellite analysis of twelve polymorphic simple sequence repeats (SSR) identified from the genome sequence of *P. ramorum*, revealed three distinct clades among 151 isolates. Populations recovered from U.S. forest and EU nursery isolates clustered into two well-supported separate clades, while isolates from a U.S. nurseries belonged to a third novel clade. *Phytophthora ramorum* genotypic diversity from populations recovered from nurseries and forests was significantly higher 91% compared to 18% respectively. Multilocus SSR analysis determined populations of *P. ramorum* collected from U.S. forests may reproduce clonally and are likely descendants of a single introduced individual. The genetic structure of populations recovered from EU nurseries displayed higher complexity, including multiple, closely related genotypes. All three clades were identified in some U.S. nurseries, including genotypes that perfectly matched the U.S. wild and the EU nursery SSR genotypic profiles, emphasizing the role of commercial plant trade in the introduction and migration of this pathogen. The combined SSR or microsatellite, sequencing and morphological analyses suggest the three clades represent distinct evolutionary lineages, all exotic to the U.S. and probably Europe and of unknown origin. Analysis using two hyper-variable tetra-repeat microsatellite loci of over 200 additional isolates, allowed for genotyping of asexually generated individuals. Results from a statewide monitoring study indicated different genotypes arise and dominate locally. In one instance, evidence of selection against old genotypes by newer ones was obtained. Genotyping allowed for the elucidation of possible migration patterns of the pathogen. Our study documents the local and regional adaptation of an introduced pathogen through the creation of novel genotypes via the accumulation of favorable mutations and/or somatic recombination.

66. Comparison of Four Microbial Genomes: *Desulfovibrio vulgaris*, *Syntrophobacter fumaroxidans*, *Methanosarcina barkeri*, and *Methanospirillum hungatei*

Johannes C. M. Scholten* (johannes.scholten@pnl.gov), Fred J. Brockman, and Weiwen Zhang

Biological Sciences Department, Pacific Northwest National Laboratory, Richland, WA

Our project team at Pacific Northwest National Laboratory (PNNL) is developing a method to predict biochemical behavior in a simple, four-microorganism community.

A set of four microorganisms form together a "simple" quad-culture community that represent a versatile sulfate-reducer that is also a syntroph (*Desulfovibrio vulgaris*), a syntrophic specialist that prefers not to be a sulfate-reducer (*Syntrophobacter fumaroxidans*), a versatile methanogen that can live acetoclastically or hydrogenotrophically (*Methanosarcina barkeri*), and a specialist hydrogenotrophic methanogen (*Methanospirillum hungatei*). The syntrophic and methanogenic guilds interact intimately with each other, their physiologies are very well known, the genomes are sequenced, the two foundational dual-cultures are well understood, and we believe this quad-culture is the most constrained microbial system available to develop an approach for identifying, understanding, and modeling dynamic and coupled inter-species molecular networks within a "simple" microbial community. Thus, the quad-culture model system is a relatively highly constrained microbial community and represents an optimal system to learn how to identify inter-species metabolic networks and interactions. Through this study, we will characterize phenotypic traits important, for example, to the sulfate-reducing and/or syntrophic lifestyles of *D. vulgaris* and *S. fumaroxidans*. The system we are using allows us to study different microbial interactions in a four-microorganism or less complex community.

The organisms *D. vulgaris* and *S. fumaroxidans* are closely related to each other as are *M. barkeri* and *M. hungatei* given that they belong to the *delta-Proteobacteria* and *Archaea*, respectively. The goal of our research is to describe carbon and energy flow in the microbial community using transcriptomic, proteomic and macroscopic process data. In collaboration with The Nimblegen Systems, we have developed customer oligonucleotide microarrays for each of these four species. In order to minimize the possible cross-hybridization during microarray analysis, extra efforts have been made to compare the genome sequences of all four species, and only the probes unique to each genome were picked and used in the microarrays. In addition, we also used the Integrated Microbial Genomes website (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi) to compare the genomes of *D. vulgaris*, *S. fumaroxidans*, *M. barkeri* and *M. hungatei*. In our paper we present some preliminary results we obtained by comparing two, three of four genomes at the same time.

Furthermore we list the genes that might be of concern for cross-hybridization.

67. Next Generation DNA Sequencing: Clonal Single Molecule Array Technology and its Application to Genetic Analysis

Gary P. Schroth* (gschroth@solexa.com)
Solexa, Hayward, CA

68. High-Throughput Eukaryotic Whole-Genome Shotgun Assembly

Harris Shapiro* (hjshapiro@lbl.gov), Jasmyn Pangilinan, and Hank Tu
DOE Joint Genome Institute, Walnut Creek, CA

Historically, the amount of time needed to run, analyze, and release whole-genome-shotgun (WGS) assemblies of eukaryotic genomes has been on the order of months or (usually) years. This has allowed for detailed human inspection and correction of assembly results, as well as extensive customization of WGS assemblers, sometimes on a per-release basis.

As the scale of WGS sequencing has increased, the amount of time available for such pre- and post-assembly analysis has been drastically reduced. The DOE Joint Genome Institute presently sequences 20 – 30 eukaryotic genomes per year, ranging in size from 15 MB to more than 1 GB. Keeping up requires running, assessing, and releasing an assembly every 1-2 weeks. The situation is further 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.

For most organisms, the main bottleneck is the person-hours needed to make sense of the results. Even at the current level of sequencing throughput, it is no longer feasible to address this by simply assigning additional people. We describe the automated pre- and post-assembly quality control (QC) checks that we have implemented, to try to identify potential assembly problems as soon as possible, and to speed their resolution. We also discuss some of the initial results of these processes, e.g. that most eukaryotic sequencing projects have some amount of potentially symbiotic prokaryotic contamination associated with them.

69. Utilization and Characterization of cDNA and Protein from the Deep-Sea Thermal Vent Worm *A. pompejana* for Structural Biology

David S. Shin* (davess@scripps.edu),[1] Michael DiDonato,[2] David P. Barondeau,[3] Greg Hura,[3] Chiharu Hitomi,[3] J. Andrew Berglund,[3] S. Craig Cary,[3] and John A. Tainer[3]

[1]Skaggs Institute for Chemical Biology and Scripps Research Institute, La Jolla, CA; [2]Protein Sciences Genomics Institute, Novartis Research Foundation, San Diego, CA; [3]Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA; [4]Institute of Molecular Biology, University of Oregon, Eugene, OR; and [5]College of Marine Studies, University of Delaware, Lewes, DE

Structural biology is used to visualize protein conformations and complex interfaces. Three-dimensional protein structures not only aid in our understanding of protein function but also guide mutational analyses and drug design. Many key human protein sample preparations have proven difficult for atomic level structural analyses, particularly those for X-ray crystallography and small-angle X-ray scattering (SAXS) in solution, due to their inherent instability. Therefore, more stable proteins derived from thermophilic unicellular organisms are often characterized in lieu of their less stable counterparts. For example, we have used the archaeaons *Pyrococcus furiosus* and *Archaeoglobus fulgidus* for structure determination of fundamental DNA repair proteins, whose defects produce genomic instability phenotypes. Although we aim to extend our analyses of additional DNA repair proteins, as well as downstream signaling and cell-cycle control partners, many are missing from unicellular thermophiles. *Alvinella pompejana*, a worm that inhabits deep-sea hydrothermal vents, may provide a missing resource of thermostable eukaryotic proteins.

A. pompejana is a polychaetous annelid that inhabits the deep-sea thermal vents that extend along 9°N to 32°S within the axial summit caldera of the East Pacific Rise. Adult worms are found within self-constructed tubes in the hottest sections of black smoker chimneys. The mixing of iron- and sulfide-rich vent fluids with ambient seawater generates solid FeS (black smoke) within a local area that is nearly devoid of detectable oxygen. The water temperature in the worms' environment reaches up to 84°C and chemical reactions at the vents drive the pH of the water to ~5.5. The water also contains a toxic soup of heavy atoms (Ag, Cd, Co, Cu, Fe, Mg, Mn, Ni, Pb, Sr, V, Zn) at levels up to 1000x to that of ambient waters.

A. pompejana samples were collected at a depth of ~2.5 km (1.55 miles) using the deep sea submersible, DSV Alvin. An *in situ* stabilization chamber that bathes the worms in a nucleic acid stabilization solution was used to preserve mRNA. cDNA sets for sequencing and cloning recombinant genes were then generated.

Our preliminary validation of *A. pompejana's* utility for structural biology was initiated using copper zinc superoxide dismutase (SOD), whose human homolog is involved in familial amyotrophic lateral sclerosis or Lou Gehrig's disease, a very common neurological disorder that affects roughly 1 in 200,000 people. *A. pompejana's* SOD protein resulted in X-ray diffraction to 0.99 Å resolution using protein crystallography and also was stable enough to generate a *de novo* low resolution structure using SAXS. These results set the stage for *A. pompejana's* use in structural biology and formed the basis of our FY2005 CSP project proposal to sequence *A. pompejana* cDNA.

70. Preliminary Analysis and Comparison of the *Pseudomonas fluorescens* Pf0-1 Genome with *P. fluorescens* Strains SBW25 and Pf5

Mark W. Silby* (mark.silby@tufts.edu),[1] Ana M. Cerdeño-Tárraga,[2] and Stuart B. Levy[1]

[1]Dept. of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, MA, and [2]Pathogen Sequencing Unit, Wellcome Trust Sanger Institute, Hinxton-Cambridge, UK

Pseudomonas fluorescens are a diverse group of bacteria commonly found in a wide range of terrestrial and aquatic habitats. Like other member of the genus *Pseudomonas*, they are metabolically versatile and highly adaptable. A number of *P. fluorescens* strains are being investigated for use in biocontrol or bioremediation. The DNA sequence of strain Pf0-1 was recently completed by the DOE Joint Genome Institute, and strain SBW25 has been sequenced at the Sanger Institute. We are in the process of annotation and preliminary analysis of these two genome sequences. The *P. fluorescens* genomes will provide us clues on how this species successfully colonizes a range of environments. Detailed comparison with the published *Pseudomonas* genomes should improve understanding of the differences between pathogenic and non-pathogenic members of the genus. It appears that *Pseudomonas* sp. share conserved 'core' genes, but significant differences exist in other genomic regions even between strains of the same species. For example, non-ribosomal peptide synthetases responsible for siderophore production have only limited similarity, possibly indicating functionally important domains. Our initial comparison aims to highlight some of these differences between Pf0-1, SBW25, and the published genome sequence of *P. fluorescens* Pf5. The Pf0-1 genome sequence and preliminary gene predictions can be found at http://genome.jgi-psf.org/finished_microbes/psefl/psefl.home.html, and the SBW25 sequence with preliminary predictions can be found at http://www.sanger.ac.uk/Projects/P_fluorescens.

71. Unlocking the Genomic and Ecophysiological Diversity of *Verrucomicrobia*

Hauke Smidt* (hauke.smidt@wur.nl),[1] Erwin Zoetendal,[1] Muriel Derrien,[1] Caroline M. Plugge,[1] Peter H. Janssen,[2] and Willem M. de Vos[1]
[1]Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands, and [2]Dept. of Microbiology and Immunology, University of Melbourne, Victoria, Australia

Isolates of the bacterial phylum *Verrucomicrobia* have in common that they are specialized to the utilization of sugars. Their abundance and activity in a large variety of aquatic and terrestrial ecosystems, as well as in the intestine of vertebrates, indicates that this until recently unknown group of microorganisms plays a major role in the global carbon cycle. This calls for whole-genome sequencing of representatives of the six different monophyletic subdivisions. (*Chthoniobacter flavus*, *Opitutus terrae*, *Akkermansia muciniphila*, and soil isolate "Ellin514"), plus a deep evolutionary relative from the candidate phylum *Lentisphaerae* (*Victivallis vadensis*).

The *Verrucomicrobia* have until recently been overlooked by conventional cultivation-based approaches. The group is an excellent example of bacteria that are hardly detected by culturing but frequently found in 16S rRNA gene libraries from a wide variety of ecosystems, including aquatic and terrestrial environments as well as the mammalian gastrointestinal tract. On average, 5% of all surveyed sequences can be assigned to this phylum. In addition to being abundant, members of the phylum *Verrucomicrobia* also appear to be active members of the soil microbial community. They produce 1 to 9.8% of the bacterial 16S rRNA in soils. The recently obtained human intestinal isolates, *Akkermansia muciniphila* and *Victivallis vadensis*, have been identified as normal inhabitants of the human intestinal ecosystem. Using genus-specific 16S rRNA-targeted FISH analysis, *Akkermansia muciniphila* was found to have a mean abundance of 1% of the total fecal bacterial community. This interaction could be related to the ability of *A. muciniphila* to degrade mucin. However, only little or nothing is known about the roles of these bacteria, in terms of their metabolic properties and, in case of intestinal isolates, their interaction with the host environment.

During the past five years, the application of improved approaches for cultivation has yielded the isolation of several members of this phylum, now for the first time allowing for a thorough intra-phylum comparison. However, although this phylum is of similar phylogenetic depth and environmental relevance as the *Proteobacteria*, only 2 genome projects are currently ongoing. Therefore, we propose whole genome sequencing of representatives of 4 of the 6 monophyletic subdivisions, and a deep evolutionary relative from the candidate phylum *Lentisphaerae*.

The availability of a comprehensive set of genomes, which for the first time will allow for sound comparative genomic approaches to organisms

from this phylum, can be expected to be of high relevance in the fields of taxonomy and evolution. Whole-genome sequencing of the proposed organisms will provide a wealth of information on their genetic potential and hence will allow for the further understanding and comparison of their metabolic capacities and interaction with their environments, including the mammalian host environment and GI tract microbiota, through comparative and functional genomics.

72. Toward a Comprehensive Overview of the *Bacillus cereus* Group Genomes

Alexei Sorokin* (sorokine@jouy.unra.fr),[1] Sandrine Auger,[1] Nathalie Galleron,[1] S. Dusko Ehrlich,[1] Eugene Goltsman,[2] Paul Richardson,[2] and Alla Lapidus[2]

[1]Génétique Microbienne, CRJ INRA, Jouy-en-Josas, France, and [2]DOE Joint Genome Institute, Walnut Creek, CA

The *Bacillus cereus* group consists of gram-positive, spore-forming bacteria with an impact on human activities due to their pathogenic properties. The most famous pathogens of animals and insects, *B. anthracis* and *B. thuringiensis*, carry their toxins in large plasmids. However, the bacteria of this group are also involved in mild food poisoning, causing rather noxious, although not fatal, vomiting or diarrhea. The emetic group strains appear to represent a very narrow phylogenetic group, and the corresponding toxin is encoded by a plasmid (Ehling-Schulz et al, 2005). Contrary to that, the characterized diarrheic toxins (Nhe, Hbl and CytK) are always present in the chromosome.

Three different strains were selected for complete sequencing, in addition to those reported earlier, to provide the comprehensive genomic characterization of the group. The work is being done as a collaboration of INRA (France), DOE Joint Genome Institute (Walnut Creek, USA) and CNS *Génoscope* (Evry, France). One of the strains, NVH398-91, is now completely sequenced. It was isolated from a severe food poisoning case in France (Lund et al, 2000). Surprisingly, this strain contains a gene for only one of known diarrheic toxins — CytK, which was shown to be synthesized in elevated amounts and to be particularly effective (Brillard and Lereclus, 2004; Fagerlund et al, 2004). The strain is rather different from other representatives of the *B. cereus* group and has a compact chromosome of 4.1 Mb, compared to 5.2-5.4 Mb genomes of other sequenced strains. This strain can also be regarded as the first bacterium of the group in which a highly effective toxin is chromosome-encoded. Divergence of the strain from others and its phenotypic particularities inspires to consider it as a representative of a new species, for which we propose a name *Bacillus cytotoxis*. However, only one such strain was described so far.

The second strain being sequenced, *Bwe* KBAB4, isolated from forest soil near Versailles (France), is a typical representative of recently recognized species — *B. weihenstephanensis* (Lechner et al, 1998; Sorokin et al, 2006). It contains a total of 5.6 Mb of genetic material, which includes genes to encode Nhe and Hbl, but lacks genes for CytK and HlyII. Although closely related strains are frequently reported to be food contaminants, they are in almost all cases not poisonous. However, the presence of Nhe and Hbl encoding genes suggests that such strains can also become toxic.

Finally, the third strain, *B. cereus* F0837/76, can be regarded as one of typical strains that can cause a severe diarrheic food poisoning. It synthesizes at least two characteristic toxins, Nhe and Hbl, in elevated amounts and was isolated from a contaminated surgical wound in a hospital (Turnbull et al, 1979; Beecher and MacMillan, 1990; Lund and Granum, 1997). Genomically the strain appeared to be very closely related to *B. anthracis* and this is probably the closest to it *B. cereus* strain known. It does not however contain the relevant pathogenic plasmids.

73. *Methanosarcina barkeri* Genome: Comparative Analysis with *M. acetivorans* and *M. mazei* Reveals Extensive Rearrangement within Methanosarcinal Genomes

Kevin R. Sowers* (sowers@umbi.umd.edu)

University of Maryland Biotechnology Institute, Rockville, MD

74. *Hydra* as a Model Organism for Studying Evolution of Metazoan Genomes

Rob Steele* (resteele@uci.edu)

Dept. of Biological Chemistry, University of California, Irvine, CA

A complete understanding of the evolution of metazoan genomes will require genome sequences from taxa at all levels of the metazoan phylogenetic tree. Of particular interest are organisms from such early-diverging phyla as Cnidaria, since they will be informative with regard to the nature of the earliest metazoan genomes. The cnidarian *Hydra* has been used as a model organism for experimental studies, particularly studies of developmental processes, for nearly 250 years. Genomic resources are now being developed for *Hydra*. These include a draft genome sequence that is nearing completion at the J. Craig Venter Institute and a dataset of approximately 150,000 ESTs. Analyses of these data have revealed a number of interesting evolutionary phenomena. As has been deduced from studies of EST datasets of limited size from two other cnidarians (*Acropora* and *Nematostella*), the *Hydra* data support the conclusion there has been substantial gene loss from the model bilaterians

Drosophila and *C. elegans*. *Hydra* carries out trans-spliced leader addition to a subset of its mRNAs, an interesting biological process and a useful one for identifying the 5′ ends of mRNAs. Surprisingly, trans-spliced leader addition is absent from *Acropora* and *Nematostella*, suggesting that it was acquired relatively late during the cnidarian radiation. We have identified a *Hydra* gene which is a homologue of the *flp* genes in the parabasalid protist *Trichomonas vaginalis*. This gene has not been found in any other genome sequence or EST dataset, including the genome sequence of *Nematostella*, a member of the same phylum as *Hydra*. These results are consistent with lateral transfer of the *flp* gene from a parabasalid protist into *Hydra*. Using sequences from the *Hydra* genome project, we have assembled the region containing the *flp* gene. Immediately upstream of the *flp* gene is a gene encoding the 140 kDa subunit of replication factor C (RFC140). The distance between the end of the RFC140 gene and the beginning of the *flp* gene is about 400 nucleotides. This short distance and the fact that the *Hydra flp* mRNA undergoes trans-spliced leader addition suggests that the *flp* gene forms an operon with the RFC140 gene. Trans-splicing is predicted to facilitate functional incorporation of laterally transferred genes into the *Hydra* genome since it would allow expression of such genes without the need for them to have their own promoters. Recently, the Bosch lab at the University of Kiel has developed a method for generating stably transgenic *Hydra*. This will allow functional studies of the many interesting genes being identified from the *Hydra* EST and genome projects.

75. Comparative Genomic Analysis of Freshwater, Groundwater, and Marine *Caulobacter* Strains

Craig Stephens* (cstephens@scu.edu)

Santa Clara University, Santa Clara, CA

76. Functional Genomics and Thermal Biology of the Porcelain Crab, *Petrolisthes cinctipes*

Jonathon Stillman* (stillmaj@sfsu.edu)

San Francisco State University, San Francisco, CA

77. *Roseobacter denitrificans*: A Ubiquitous Marine Phototroph with a Peculiar Carbon Fixation Pathway

Wesley D. Swingley* (wswingley@msn.com),[1] Sumedha Gholba,[2] Steven D. Mastrian,[3] Heather J. Matthies,[4] Jicheng Hao,[3] Hector Ramos,[2] Chaitanya R. Acharya,[2] Amber L. Conrad,[3] Heather L. Taylor,[3] Liza C. Dejesa,[3] Maulik K. Shah,[3] Maeve E. O'Huallachain,[3] Michael T. Lince,[4] Robert E. Blankenship,[4] J. Thomas Beatty,[5] and Jeffrey W. Touchman[3]

[1]School of Life Sciences, Arizona State University, Tempe, AZ; [2]Dept. of Computational Biosciences, Arizona State University, Tempe, AZ; [3]Translational Genomics Research Institute, Phoenix, AZ; [4]Dept. of Chemistry and Biochemistry, Arizona State University, Tempe, AZ; and [5]University of British Columbia, Vancouver, BC

Purple aerobic phototrophic bacteria (APBs) are the only bacteria that perform photosynthesis in the presence of oxygen, but do not produce oxygen as a result. These highly adaptive APB compose greater than 10% of the microbial community in some euphotic upper ocean waters and are potentially major contributors to the fixation of the greenhouse gases CO and CO_2. We describe genomic sequence features of the APB *Roseobacter denitrificans* that reveal clues to its physiology. The genome does not contain any genes that code for known carbon-fixation pathways and most notably missing is a gene for the Calvin cycle enzyme RubisCO. However, we describe the presence and potential importance of a mixotrophic CO_2-fixation pathway analogous to the C_4 cycle found in many plants. We suggest that this pathway functions to fix much of the CO_2 for the formation of cellular components, but does not allow for autotrophic growth. While some genes that code for the redox-dependent regulation of photosynthetic machinery are present, some light sensors and transcriptional regulatory motifs found in other anoxygenic photosynthetic bacteria are absent.

78. Novel Proteins from an Acidic, Hot, Metal-Rich Ecosystem

Michael P. Thelen* (mthelen@llnl.gov),[1] **Jill Banfield**,[3] Steven Singer,[1] Christopher Jeans,[1] Mona Hwang,[1] Nathan VerBerkmoes,[2] Manesh Shah,[2] Robert Hettich,[2] Clara Chan,[3] Vincent Denef,[3] and Gene Tyson[3]

[1]Biosciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA; [2]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; and [3]Dept. of Earth and Planetary Sciences, University of California, Berkeley, CA

Although a large portion of proteins deduced from microbial genomes do not match any annotated sequences, it is likely that many of these "hypothetical proteins" are important, in some cases essential, for

microbial fitness under natural conditions. The void in our knowledge regarding this dark side of genomics, representing 30 – 40% of protein sequences, is a formidable obstacle to nearly every investigation enabled by full genome sequence information. In our study of acidophilic biofilm communities, we are focused directly on the problem of hypothetical proteins within a well-defined system in one of the most extreme limits of habitability. These microbes assemble in very acidic streams of underground mine tunnels to form floating, matrix-bound, self-sustaining communities that derive electrons from iron (II) while in turn catalyzing the dissolution of pyrite (FeS_2). Such conditions result in the biological generation of acid mine drainage (AMD), ultimately causing many orders of magnitude greater metal dissolution and environmental acidification than abiotic forces alone. Since AMD is so prevalent and potentially disastrous to U.S. inland waters associated with principle energy resources such as coal and uranium, understanding the mechanisms by which biofilm communities generate AMD is a priority with DOE's Office of Biological and Environmental Research.

In addition to the detailed geochemical and microbiological description of AMD biofilms, the foundation of our current investigation rests on both genomic and proteomic datasets for each of several dominant organisms. Characterization of abundant species in one such biofilm and the associated genome analyses resulted in reconstruction of near complete or partial genomes for two different *Leptospirillum* bacteria and three archaea,[1] and further work at the JGI is now extending the sequences within this biofilm and also to ones with differing population structures. In our attempt to understand essential biofilm metabolic activities and the partitioning of functions between different organism types, proteins native to an environmental biofilm were analyzed by shotgun MS proteomics. This confirmed that many hypothetical proteins are expressed at detectable levels in the major biofilm organisms, and also indicated that a majority of abundant extracellular proteins are novel.[2] This groundwork has enabled us to examine proteins expressed in biofilms harvested from several neighboring locations, in carefully measured geochemical conditions, using a combination of chromatographic techniques, MS identification and computational analyses/predictions.[3] These methods provide us with a powerful approach to isolate many of the novel proteins along with any known proteins associated with them, and to begin assigning functions using biochemical assays. Several key enzymatic activities that are suspected to play important roles in biofilm formation and function, such as hydrolases, oxidoreductases and polysaccharide synthases, will be examined in parallel. The development of this type of systematic approach for the validation and description of hypothetical proteins, used within a defined biological system, is a test case for the full utility of proteogenomic information.

1. Tyson et al, 2004, Nature 428:37-43.
2. Ram et al, 2005, Science 308:1915-20.
3. Banfield et al, 2005, OMICS 9:301-33.

79. The Hox Cluster of a Model Chelicerate Arthropod—The Oribatid Mite *Archegozetes longisetosus*

R. H. Thomas* (rthomas@zoology.siu.edu),[1,2] J. A. Mackenzie-Dodds,[2] H. M. Fourcade,[3] and J. L. Boore[3]

[1]Dept. of Zoology, Southern Illinois University, Carbondale, IL; [2]Dept. of Zoology, Natural History Museum, London, UK; and [3]Evolutionary Genomics Program, DOE Joint Genome Institute, Walnut Creek, CA

The arthropods are the most diverse group of animals and are usually divided into four major groups: insects, crustaceans, myriapods (centipedes, millipedes and others), and chelicerates (horseshoe crabs and arachnids, which contain mites, spiders, scorpions and other eight-legged animals). Of these chelicerate groups, mites are by far the most diverse, both in numbers of species and in ecological breadth. Here we report on the sequence of the first chelicerate arthropod Hox cluster from a free-living oribatid mite, *Archegozetes longisetosus*. We have demonstrated that the mite homologs of the Drosophila genes *zen* and *ftz* both behave as Hox genes in this mite, lending support to the notion that the ancestral arthropod Hox cluster contained ten genes. The Archegozetes homolog of *abdominal-A* is not in its usual position between the mite homologs of *Ultrabithorax* and *Abdominal-B*. All attempts to locate an *abdominal-A* homolog elsewhere in the genome have failed. We conclude that *abd-A* is lost in this animal — the first verified loss of a Hox gene reported in an arthropod. There are no obvious simplifications or losses in the mite opisthosoma that suggest an immediate explanation for the loss of *abdominal-A*.

80. Mining of Switchgrass Microsatellites from EST Data for Mapping and Diversity Analysis

Christian M. Tobias* (ctobias@pw.usda.gov)

U.S. Department of Agriculture, Agricultural Research Service

81. Plant-Plant Interactions: Parasitic Plant Responses in Host Plant Interactions

Alexey Tomilov,[1] Natalya Tomilova,[1] Manuel Torres,[2] Russell Reagan,[1] Tatiana Fillapova,[1] Pradeepa Gunathilake,[1] and **John I. Yoder***[1] (jiyoder@plantsciences.ucdavis.edu)

[1]Dept. of Plant Sciences, University of California, Davis, CA, and [2]Center for Applied Genetic Technologies, University of Georgia, Athens, GA

In natural environments roots of different plants grow closely entwined and frequently contact. When roots of parasitic plants in the Orobanchaceae contact roots of neighboring plants, they develop haustoria

which attach to and invade host root tissues. When mature, haustoria connect parasite and host roots and provide the physiological conduit through which host resources are translocated into the parasite. The consequences of parasitism can be dire for the host plant and some of the world's worst agricultural pests are parasitic weeds. We are identifying genes that encode early haustorium development in parasitic Orobanchaceae in order to 1) discover the genetic basis for the origin(s) of parasitism in plants; 2) determine how parasitic plants identify and discriminate hosts, and; 3) develop strategies for engineering crop plants for genetic resistance against parasitic weeds.

Triphysaria is a parasitic plant that grows as a common springtime wildflower in coastal bluffs, serpentine slopes and grassland fields throughout the Pacific Coast of North America. *Triphysaria* is closely related to the internationally devastating weeds *Striga* and *Orobanche* and provides an experimentally tractable model for these aggressively invasive, highly restricted, agricultural pests. *Triphysaria* develop knob shaped haustoria at their root tips when seedlings are treated with host (*Arabidopsis*) roots, host root exudates, or purified root factors. Under these in vitro conditions haustorium development is robust, rapid and highly synchronous.

We used suppression subtractive hybridization (SSH) to enrich cDNA libraries for transcripts regulated in *Triphysaria* roots within a few hours after exposure to different haustorium inducing treatments. About ten thousand SSH products have been sequenced to date with an addition twenty thousand SSH products and thirty thousand full length cDNAs in the JGI sequencing pipeline. Using a combination of bioinformatics and cDNA arrays we showed that many of the genes expressed early in haustorium development are similarly regulated when exposed to cytotoxins released by allelopathic plants, even when the allelochemicals are inactive as haustoria inducers. Protein annotations suggest that many of these genes encode xenobiotic detoxification enzymes. A significant fraction of the parasite transcriptome regulated by host contact is predicted to function in electron or chemical transport. Two distinct quinone oxidoreductases regulated in *Triphysaria* roots by host contact have been expressed in *E. coli*, purified and characterized. Gene transfer technologies have been developed for *Triphysaria* roots and these are being used to determine the function of these and other candidate parasite genes. Our results suggest that host recognition involves a redox associated signal transduction system recruited in evolution from genes functioning in rhizosphere chemical detoxification.

82. Genome Sequence Comparisons of the Eukaryotic Plant Pathogens *Phytophthora sojae* and *Phytophthora ramorum*

Sucheta Tripathy* (stripath@vt.edu)
Virginia Polytechnic University and State University, Blacksburg, VA

83. Progress on the Human Gut Microbiota Project

Parag Vaishampayan* (pavaishampayan@lbl.gov),[1] Jenna Morgan,[1] Jeff Froula,[1] Yogesh Shouche,[2] Howard Ochman,[3] and M. Pilar Francino[1]
[1]Evolutionary Genomics Department, DOE Joint Genome Institute, Walnut Creek, CA; [2]National Centre for Cell Science, University of Pune, India; and [3]Departments of Biochemistry and Molecular Biophysics, and Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ

The diverse community of gut microbes plays an essential role in human health, contributing to the digestive process, promoting gut maturation, modulating the immune system, and interacting with pathogens in several complex ways. We have initiated efforts to study complex gut microbial community structure and their coding capabilities.

Human fecal samples from two healthy anonymous volunteers (mother and infant) were collected at the University of Arizona in 2004 and 2005, in collaboration with the laboratory of Howard Ochman. High molecular weight DNA for large insert library construction was prepared from these samples. From these preparations, we have produced 4 different fosmid libraries, two from adult samples (2004 and 2005) and two infant samples (2004, 1 month old and 2005, 11 months old), each containing 50,000 to 70,000 clones (40 kb insert size). We have also prepared 16s libraries from the DNA agarose plugs from which the fosmid libraries were prepared.

We have initiated screenings of the fosmid libraries with 16s rRNA primers as well as a set of 11 primer sets directed towards universally conserved bacterial proteins (Santos and Ochman 2004) for phylogenetic analysis. Currently we are optimizing protocols to avoid amplification of the host *E. coli* strain genes. The different alternative protocols include fosmid induction, effective fosmid purification, use of DNA exonucleases and use of terminator primers to abort PCR extension of *E. coli* genes.

We aim to further characterize these libraries by screening different types of genes involved in microbe-host interactions. We will perform phenotypic screenings to identify clones producing antibiotic molecules potentially involved in competition between the microbial community members. The information on phylogenetic diversity, coding capabilities related to commensalism and pathogenicity and antibiotic production will provide an ample biological background describing the GI microbiotic community.

In future we have plans to study succession and development of gut microbiota of new born infant as collaborative efforts with a research institution in India. To clarify the initial acquisition and subsequent colonization of bacteria in an infant within the few months after birth, phylogenetic analysis will be performed using 16S rDNA sequences from the DNA isolated from feces. In addition, fosmid libraries will be prepared from these DNA samples and comparative analyses will be performed between the American and Indian microbial succession data.

84. A Unique Diversity of Cytochrome C-Dependent Dehydrogenases Found in the *Pantoea citrea* Genome

Fernando Valle* (fvalle@genencor.com), Mai Du, Bob Caldwell, Harun Rashid, and Tim Dodge

Genencor International/Danisco, Palo Alto, CA

Some bacteria utilize the extracellular oxidation of substrates as a metabolic strategy to generate membrane potential. The best known microorganism with this type of metabolism is *Gluconobacter oxidans*, who is capable of oxidizing several carbohydrates using six periplasmic dehydrogenases. Two of these dehydrogenases contain a cytochrome C subunit.

We have developed a very efficient fermentation process to produce 2-keto-L-gulonate (2-KLG) from glucose using the bacteria *Pantoea citrea* as the production host. 2-KLG is an intermediate for the synthesis of Ascorbic acid.

To understand the potentials and limitations of *P. citrea* as a production host, and to directed some of our Metabolic Engineering efforts, its genome was sequenced.

Genome analysis showed that *P. citrea* possesses an array of genes to modify carbohydrates extracellularly. Among them, 19 operons containing cytochrome c-dependent extracellular dehydrogenases were found. So far, *P. citrea* is the only bacteria possessing this astonishing number of paralogs. An analysis of these 19 operons will be presented.

85. *Dinoroseobacter shibae*, A New Aerobic Phototrophic Bacterium from the *Roseobacter* Lineage with a Complex Genome Structure Producing Long-Chain Acylated Homoserine Lactones

Irene Wagner-Döbler* (iwd@gbf.de),[1] Silke Pradella,[2] Brian Tindall,[2] Rüdiger Pukall,[2] Stefan Schulz,[3] Hanno Biebl,[1] and Martin Allgaier[4]
[1]GBF, Braunschweig, Germany; [2]German Collection for Microorganisms and Cell Cultures, Braunschweig, Germany; [3]Technical University, Braunschweig, Germany; and [4]Leibniz-Institute for Aquatic Ecology, Neu-Globsow, Germany

First described 15 years ago, the genus *Roseobacter* has become the core of a group of marine bacteria which are the focus of intense research because of their taxonomic diversity and abundance in the world's oceans as well as their important contribution to the global carbon and sulfur cycles and thus to climate change. *Dinoroseobacter shibae* is a novel genus within this group that was isolated from toxic marine dinoflagellates[2] and is able to perform aerobic anoxygenic photosynthesis, a way of generating energy using bacteriochlorophyll *a* which contributes up to 10% to the primary production in the ocean and is presently poorly understood. *Dinoroseobacter shibae* carries the *pufLM* genes of the photosynthesis reaction center and has fully functional light harvesting complexes I and II including bacteriochlorophyll *a* synthesis[4]. *Dinoroseobacter shibae* has a complex genome, harbouring seven linear plasmids which are most probably conjugative and together comprise 861 kb of extrachromosomal information[3]. Finally, *Dinoroseobacter shibae* has been found to produce highly unusual quorum sensing signalling compounds, i.e. acylated homoserine lactones (AHLs) with side chain lengths of up to 18 carbon atoms, the longest AHL side chain length presently known. Three different AHL signals have been discovered by their ability to stimulate a luxR-type reporter strain, as well as by GC-MS analysis of concentrated culture supernatants resulting in structure elucidation, namely C8-HSL, C18-en-HSL and C18-dien-HSL[1]. It is conceivable that some of the genetic determinants, e.g. the luxR type receptor and the luxI type synthase genes, might be located on the many linear plasmids discovered in this organism, from where they could easily be transferred to related species.

The genome sequence of *Dinoroseobacter shibae* will make it possible to unravel the mechanisms of aerobic anoxygenic photosynthesis and cell density dependent gene expression and thus help to understand the adaptations of this organism to its ecological niche in the ocean.

1. Wagner-Döbler et al. (2005) Discovery of complex mixtures of novel longchain quorum sensing signals in free-living and host-associated marine Alphaproteobacteria. ChemBioChem 6(12):2195-206.
2. Biebl et al. (2005) *Dinoroseobacter shibae*, gen. nov., sp. nov., a new aerobic phototrophic bacterium isolated from dinoflagellates. Int. J. System. Evol. Microbiol. **55**: 1089 - 1096.

3. Pradella et al. (2004) Genome Organization and Localization of the *pufLM* Genes of the Photosynthesis Reaction Center in Phylogenetically Diverse Marine *Alphaproteobacteria*. Appl. Environ. Microbiol. **70**:3360-3369.
4. Allgaier et al. (2003) Aerobic anoxygenic photosynthesis in *Roseobacter* clade bacteria from diverse marine habitats. Appl. Environ. Microbiol. **69**:5051-5059.

86. Genome Sequence of the Cellulolytic Gliding Bacterium *Cytophaga hutchinsonii*

G. Xie[1,2], D. C. Bruce[1,2], J. F. Challacombe[1,2], O. Chertkov[1,2], P. Gilna[1,2], C. S. Han[1,2], B. Henrissat[3], S. Lucas[2,4], M. Misra[1], G. L. Myers[1], P. Richardson[2,5], R. Tapia[1,2], N. N. Thayer[1], L. S. Thompson[1,2], T. S. Brettin[1,2], and **M. J. McBride*** (mcbride@uwm.edu)[6]

[1]Los Alamos National Laboratory, Los Alamos, NM; [2]DOE Joint Genome Institute, Walnut Creek, CA; [3]CNRS and Universites Aix-Marseille I & II, Marseille, France; [4]Lawrence Livermore National Laboratory, Livermore, CA; [5]Lawrence Berkeley National Laboratory, Berkeley, CA; [6] University of Wisconsin, Milwaukee, WI

We present the complete DNA sequence and metabolic analysis of the aerobic cellulolytic Gram-negative soil bacterium *Cytophaga hutchinsonii,* which belongs to the phylum *Bacteroidetes* (also known as the *Cytophaga-Flavobacterium-Bacteroides* group). The genome consists of a single circular 4.43 Mb chromosome containing 3805 open reading frames (ORFs), with 37% GC content, of which 1969 ORFs (52%) have been assigned a tentative function. Two of the most striking characteristics of *C. hutchinsonii* are its rapid gliding motility over surfaces and its contact dependent digestion of crystalline cellulose. The mechanism of *C. hutchinsonii* motility is not known but its genome contains homologs for each of the *gld* genes that are required for gliding of the distantly related bacteroidete *Flavobacterium johnsoniae*. *Cytophaga-Flavobacterium* gliding appears to be novel and does not involve well-studied motility organelles such as flagella or type IV pili. Many genes thought to encode proteins involved in cellulose utilization were identified. These include putative endo-1, 4-β-glucanases and β-glucosidases. Surprisingly, genes encoding obvious exoglucanases were not detected. Since such enzymes are usually needed for efficient cellulose digestion *C. hutchinsonii* either has novel exoglucanases or has an unusual method of cellulose utilization. Genes encoding typical cellulosome components were absent but the genome appears to encode 74 very-high-molecular-weight proteins that may be part of the cell-bound cellulolytic machinery. Most of these are predicted to be extracellular or cell surface proteins and many have putative polysaccharide lytic D5 domains. 184 transporters have been identified, but PTS systems are absent. Biosynthetic pathways for all common amino acids, nucleotides, vitamins and coenzymes appear to be present. Pathways needed for oxidative respiration of glucose appear intact and a modified oxidative branch of the pentose phosphate pathway is also predicted.

87. Origin and Evolution of Vertebrate Neural Crest Gene Network—Insights from Comparative Genomics Analysis of *Amphioxus* ESTs

Jr-Kai Sky Yu* (jkyu@caltech.edu),[1] Yutaka Satou,[2] Tadasu Shin-I,[3] Yuji Kohara,[3] Nori Satoh,[2] Linda Z. Holland,[4] and Marianne Bronner-Fraser[1]

[1]Division of Biology, California Institute of Technology, Pasadena, CA; [2]Dept. of Zoology, Graduate School of Science, Kyoto University, Kyoto, Japan; [3]National Institute of Genetics, Shizuoka, Japan; and [4]MBRD, Scripps Institution of Oceanography, University of California, San Diego, CA

The neural crest is a cell type unique to vertebrates that arises in the embryo from the border of the neural plate and adjacent non-neural ectoderm. Neural crest cells undergo an epithelial/mesenchymal transition to become a migratory population of cells that gives rise to numerous differentiated cell types, including the craniofacial skeleton and much of the peripheral nervous system. Evolution of the neural crest is thought to be responsible for formation of many of the complex structures of the vertebrate head that are lacking in invertebrate chordates. The cephalochordate, *amphioxus*, is an evolutionary model for understanding the origin of vertebrates since its embryonic body plan is very vertebrate-like. We have taken a comparative approach using a large scale EST analysis to identify *amphioxus* homologs of vertebrate genes involved expressed at the neural plate border in a putative gene regulatory network. Our preliminary survey shows that while the early stages of neural plate border patterning genes appear conserved between *amphioxus* and vertebrates, later deployment of some transcription factors at the neural plate border and presumptive neural crest region seems to be a vertebrate innovation. Such an understanding of the evolution of gene regulatory events gives important insights into how developmental genes may have been co-opted for new roles in patterning the neural crest as a novel vertebrate innovation.

88. Metagenomic Analysis of Microbial Communities in Uranium-Contaminated Groundwaters

Jizhong Zhou* (jzhou@rccc.ou.edu),[1,6] Terry Gentry,[1] Chris Hemme,[1,6] Liyou Wu,[1] Matthew W. Fields,[2] Chris Detter,[3] Kerrie Barry,[3] David Watson,[1] Christopher W. Schadt,[1] Paul Richardson,[3] James Bristow,[3] Terry Hazen,[4] James Tiedje,[5] and Eddy Rubin[3]

[1]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Dept. of Microbiology, Miami University, Oxford, OH; [3]DOE Joint Genome Institute, Walnut Creek, CA; [4]Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [5]Center for Microbial Ecology, Michigan State University, East Lansing, MI; and

[6]Institute for Environmental Genomics, Dept. of Botany and Microbiology, University of Oklahoma, Norman, OK

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 contaminated with very high levels of nitrate, uranium and other heavy metals and pH 3.7. Sequence analysis of this groundwater sample based on a 16S rDNA library revealed 10 operational taxonomic units (OTUs) at the 99.6% cutoff with >90% of the OTUs represented by an unidentified y-proteobacterial species related to the genus *Frateuria.* Additional OTUs were related to a J3-proteobacterial species of the genus *Azoarcus.* Three clone libraries with different DNA fragment sizes (3, 8 and 40 kb) were constructed, and 70 Mb raw sequences were obtained using a shotgun sequencing approach. The raw sequences were assembled into 2770 contigs totaling 6 Mb which were further assembled into 224 scaffolds (1.8 kb-2.4 Mb). Preliminary binning of the scaffolds suggests 4 primary phylotypes (FRC Gamma Groups I & II (*Frateuria*), FRC Beta Group I (*Burkholderia*) and FRC Beta Group II (*Herbaspirillum*), with the FRC Gamma Group I representing the dominant phylotype (1 IX genome coverage). Genes identified from the sequences were consistent with the geochemistry of the site, including multiple nitrate reductase and metal resistance genes. Despite the low *species* diversity of the samples, evidence of strain diversity within the identified species was observed. Analysis with functional gene arrays containing 23,000 probes designed based on these community sequences as well as genes important for biogeochemical cycling of C, N, and S, along with metal resistance and contaminant degradation suggested that the dominant species could be biostimulated during *in situ* uranium reduction experiments at the FRC. These results also suggest that the dominant species could play a direct or indirect role in the bioremediation of uranium.

89. "Omics" of *Dehalococcoides* spp.

Stephen H. Zinder* (shz1@cornell.edu)
Dept. of Microbiology, Cornell University, Ithaca, NY

Dehalococcoides spp. are members of the *Chloroflexi* that can reductively dechlorinate diverse chloroorganics including the solvents tetrachloroethene (PCE) and trichloroethene (TCE), chlorobenzenes, and PCBs. The genome of the PCE-dechlorinating *D. ethenogenes* strain 195 was 1.47 Mb and encoded 17 predicted reductive dehalogenases (RDHs). The genome of the chlorobenzene dechlorinating *Dehalococcoides* strain CBDB1 was 1.40 Mb and encoded 32 predicted RDHs. There is near complete synteny between the two strains in regions encoding "housekeeping" genes, whereas regions encoding RDHs show signs of genetic exchange and rearrangement. Two more *Dehalococcoides* genome sequences are being determined by JGI. Transcriptional analyses of RDH expression in *D. ethenogenes* grown on PCE or TCE by qPCR showed high expression of DET0079 and DET0318, predicted to encode TCE RDH and PCE RDH respectively. Highly expressed oxidoreductases included the Hup, Vhu, and Hym hydrogenases and an ORF predicted to encode a formate dehydrogenase (FDH). Tandem mass spectroscopy of tryptic peptides generated from membrane enriched fractions from *D. ethenogenes* showed high coverage of the RDHs and oxidoreductases described above. The high expression of the predicted FDH was surprising since *D. ethenogenes* does not utilize formate. Examination of the predicted amino acid sequence revealed a serine encoded in place of cysteine or selenocysteine at a critical residue. Also showing high peptide coverage was DET1407, and analysis of this protein suggests it is part of the S-layer cell wall in *D. ethenogenes*. These studies show that *Dehalococcoides* spp. are highly evolved to use halogenated compounds and demonstrate the ability of "omics" to shed light on organisms difficult to study in the laboratory.

Attendees

Current as of March 10, 2006

Catherine C. Adley
University of Limerick
catherine.adley@ul.ie

Andrea Aerts
DOE Joint Genome Institute
alaerts@lbl.gov

Nina Agabian
University of California, San Francisco
nina.agabian@ucsf.edu

Eric E. Allen
University of California, Berkeley
eallen@nature.berkeley.edu

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

Anne J. Anderson
Utah State University
anderson@biology.usu.edu

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

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

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

Jody Banks
Purdue University
banksj@purdue.edu

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

Stephan R. Berosik
Applied Biosystems
berosisr@appliedbiosystems.com

Alison M. Berry
University of California, Davis
amberry@ucdavis.edu

Diana L. Berry
Stony Brook University
diberry@optonline.net

Ilka Bischofs
University of California, Berkeley
ibbischofs@lbl.gov

Hans J. Bohnert
University of Illinois, Urbana-Champaign
bohnerth@life.uiuce.edu

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

Jeffrey L. Boore
DOE JGI, Lawrence Berkeley National Laboratory
JLBoore@Berkeley.edu

Chris Bowler
Ecole Normale Superieure
cbowler@biologie.ens.fr

Elbert W. Branscomb
Lawrence Livermore National Laboratory
branscomb1@llnl.gov

Lambert Brau
Murdoch University
lbrau@murdoch.edu.au

Jim Bristow
DOE Joint Genome Institute
JBristow@lbl.gov

Fred Brockman
Pacific Northwest National Laboratory
Fred.Brockman@pnl.gov

Eoin Brodie
Lawrence Berkeley National Laboratory
elbrodie@lbl.gov

Shane Brubaker
DOE Joint Genome Institute
shanebrubaker@sbcglobal.net

Donald A. Bryant
The Pennsylvania State University
dab14@psu.edu

Andrew Cameron
California Institute of Technology
acameron@caltech.edu

Attendees

Patrick Chain
Lawrence Livermore National Laboratory
chain2@llnl.gov

Cathy Chang
Diversa Corporation
cchang@diversa.com

Quen J. Cheng
Lawrence Berkeley National Lab
QJCheng@lbl.gov

Penny Chisholm
Massachusetts Institute of Technology (MIT)
chisholm@MIT.edu

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

Dylan Chivian
Lawrence Berkeley National Laboratory
DCChivian@lbl.gov

Richard M. Clark
Max-Planck-Institute for Developmental Biology
richard.clark@tuebingen.mpg.de

John K. Colbourne
Indiana University
jcolbour@cgb.indiana.edu

Matthew Coleman
Lawrence Livermore National Laboratory
coleman16@llnl.gov

Lorenza Conterno
Cornell University, N.Y. State Agriculture Experiment Station
lc98@cornell.edu

Luis M. Corrochano
Universidad de Sevilla
corrochano@us.es

Adrienne M. Cottrell
University of Georgia
acottrel@uga.edu

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

Christopher Daum
DOE Joint Genome Institute
daum1@llnl.gov

Charles N. David
Ludwig Maximilian University, Munich
david@zi.biologie.uni-muenchen.de

Seana K. Davidson
University of Washington
skdavid@u.washington.edu

Scott C. Dawson
University of California, Berkeley
sdawson@nature.berkeley.edu

Anthony De Tomaso
Stanford University
tdet@stanford.edu

Paramvir S. Dehal
DOE Joint Genome Institute
psdehal@lbl.gov

Edward F. DeLong
Massachusetts Institute of Technology
delong@mit.edu

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

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

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

Ilenys Diaz-Muniz
USDA-ARS at NCSU
idiazmu@ncsu.edu

Larry J. Dishaw
All Children's Hospital, H. Lee Moffitt Cancer Center and Research Institute
ldishaw@hsc.usf.edu

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

Nicole Dubilier
Max Planck Institute for Marine Microbiology
ndubilie@mpi-bremen.de

Jennifer L. DuBois
University of Notre Dame
jdubois@nd.edu

Serge Dusheyko
DOE Joint Genome Institute
sdusheyko@lbl.gov

Joseph Ecker
University of California, San Diego
ecker@salk.edu

Eric Edsinger-Gonzales
University of California, Berkeley
eegonzales@berkeley.edu

Michael Egholm
454 Life Sciences
megholm@454.com

Jonathan A. Eisen
The Institute for Genomic Research
jonathan_eisen@mac.com

Drew Endy
Massachusetts Institute of Technology (MIT)
endy@mit.edu

Marsha Fenner
DOE Joint Genome Institute
MWFenner@lbl.gov

Matthew W. Fields
Miami University
fieldsmw@muohio.edu

Peg Folta
Lawrence Livermore National Laboratory
pfolta@llnl.gov

Thomas Fowler
Southern Illinois University, Edwardsville
tfowler@siue.edu

Maria Pilar Francino
DOE Joint Genome Institute
mpfrancino@lbl.gov

Marvin E. Frazier
J. Craig Venter Institute
mfrazier@venterinstitute.org

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

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

Jill O. Fuss
Lawrence Berkeley National Laboratory
jfuss@lbl.gov

Eric Gaidos
University of Hawaii
gaidos@hawaii.edu

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

Paul Gilna
Los Alamos National Laboratory
pgil@lanl.gov

Tijana Glavina Del Rio
DOE Joint Genome Institute
glavinadelrio1@llnl.gov

Frank O. Gloeckner
Max Planck Institute for Marine Microbiology
fog@mpi-bremen.de

Bernardo Gonzalez
Facultad de Ciencias Biologicas. P. Universidad Catolica de Chile
bgonzalez@bio.puc.cl

Stephen B. Goodwin
USDA-ARS and Purdue University
sgoodwin@purdue.edu

Joe W. Gray
Lawrence Berkeley National Laboratory
JWGray@lbl.gov

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

Arthur Grossman
Carnegie Institution of Washington
arthurg@stanford.edu

Steven J. Hallam
Bacterial Adaptation and Response Network
shallam@mit.edu

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

Emily E. Hare
U.C. Berkeley and LBNL
ehare@berkeley.edu

Richard Harland
U.C. Berkeley
harland@berkeley.edu

Jonathan K. Harris
University of Colorado
jjharris@colorado.edu

Loren J. Hauser
Oak Ridge National Laboratory
hauserlj@ornl.gov

Dennis Hedgecock
University of Southern California
dhedge@usc.edu

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

Thomas Henick-Kling
Cornell University
th12@cornell.edu

Isaac Y. Ho
DOE Joint Genome Institute
iyho@lbl.gov

Roger Hohnsbeen
Researcher
roger@hohnsbeen.com

Attendees

Dawn E. Holmes
University of Massachusetts
dholmes@microbio.umass.edu

Amy J. Horneman
University of Maryland School of Medicine
ahornema@epi.umaryland.edu

Benjamin Horowitz
JGI PGF
horowitz2@llnl.gov

Stephen J. Horton
Union College
hortons@union.edu

Ping Hu
Lawrence Berkeley National Lab
phu@lbl.gov

Phil Hugenholtz
DOE Joint Genome Institute
PHugenholtz@lbl.gov

Lonnie O. Ingram
University of Florida-IFAS
ingram@ufl.edu

Natalia Ivanova
DOE Joint Genome Institute
NNIvanova@lbl.gov

Thomas W. Jeffries
USDA, FS, FPL
twjeffri@wisc.edu

Paul R. Jensen
Scripps Institution of Oceanography
pjensen@ucsd.edu

Jhaveri Jinal
DOE Joint Genome Institute
jajhaveri@lbl.gov

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

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

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

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

Staci Kane
Lawrence Livermore National Laboratory
kane11@llnl.gov

Samuel Kaplan
University of Texas
Health Science Center at Houston
samuel.kaplan@uth.tmc.edu

Kenneth G. Karol
University of Washington
kkarol@u.washington.edu

Gerrit H. Kema
Plant Research International B.V.
GERT.KEMA@WUR.NL

Cheryl A. Kerfeld
University of California, Los Angles
kerfeld@mbi.ucla.edu

Arshi Khan
DOE Joint Genome Institute
khan5@llnl.gov

Nicole King
University of California, Berkeley
nking@berkeley.edu

Stephen Kingsmore
NCGR
sfk@ncgr.org

Martin G. Klotz
University of Louisville
mgk@mgklotz.com

Michael L. Knotek
Knotek Scientific Consulting
m.knotek@verizon.net

Arthur Kobayashi
DOE Joint Genome Institute
kobayashi1@llnl.gov

Thomas D. Kocher
University of New Hampshire
tdk@cisunix.unh.edu

Alexander Kozik
University of California, Davis
akozik@atgc.org

Jerome J. Kukor
Rutgers University
kukor@aesop.rutgers.edu

Alan Kuo
DOE Joint Genome Institute
akuo@lbl.gov

Nikos Kyrpides
DOE Joint Genome Institute
NCKyrpides@lbl.gov

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

Gerard R. Lazo
USDA-ARS Western Regional Research Center
lazo@pw.usda.gov

Jared R. Leadbetter
California Institute of Technology
jleadbetter@caltech.edu

Patrick Lemaire
CNRS
lemaire@ibdm.univ-mrs.fr

Johan Leveau
Netherlands Institute of Ecology
(NIOO-KNAW)
j.leveau@nioo.knaw.nl

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

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

Athanasios Lykidis
DOE Joint Genome Institute
alykidis@lbl.gov

Ronald C. Mackenzie
University of Texas Health Science Center at Houston
ronald.c.mackenzie@uth.tmc.edu

Jon K. Magnuson
Pacific Northwest National Laboratory
jon.magnuson@pnl.gov

Dina F. Mandoli
University of Washington
mandoli@u.washington.edu

Reinhold C. Mann
Oak Ridge National Laboratory
mannrc@ornl.gov

Betty K. Mansfield
Oak Ridge National Laboratory
mansfieldbk@ornl.gov

Meng Markillie
Pacific Northwest National Laboratory
meng.markillie@pnl.gov

Eric J. Mathur
Diversa Corporation
eric.mathur@diversa.com

Julie A. Maupin-Furlow
University of Florida
jmaupin@ufl.edu

Mark J. McBride
University of Wisconsin, Milwaukee
mcbride@uwm.edu

Sandra McFarland
DOE Joint Genome Institute
Somcfarland@lbl.gov

J. Mitchell McGrath
USDA-ARS
mitchmcg@msu.edu

Michael J. McInerney
University of Oklahoma
mcinerney@ou.edu

David Mead
Lucigen
dmead@lucigen.com

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

Galit Meshulam-Simon
Stanford University
meshulam@stanford.edu

Axel Meyer
University of Konstanz
axel.meyer@uni-konstanz.de

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

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

Marissa Mills
Oak Ridge National Laboratory
MILLSMD@ORNL.gov

Brent D. Mishler
University of California, Berkeley
bmishler@calmail.berkeley.edu

Herve Moreau
Oceanological Observatory of Banyuls
h.moreau@obs-banyuls.fr

Jenna L. Morgan
DOE Joint Genome Institute
jlmorgan@lbl.gov

Kenneth H. Nealson
University of Southern California
knealson@usc.edu

Irene L. Newton
Harvard University
garcia@fas.harvard.edu

Shirley F. Nishino
Georgia Institute of Technology
sn81@ce.gatech.edu

Attendees

Kenneth M. Noll
University of Connecticut
kenneth.noll@uconn.edu

Howard Ochman
University of Arizona
hochman@email.arizona.edu

Graham W. O'Hara
Murdoch University
gohara@murdoch.edu.au

Regina A. O'Neil
University of Massachusetts
rtarallo@microbio.umass.edu

Bobby Otillar
DOE Joint Genome Institute
RPOtillar@lbl.gov

Norman R. Pace
University of Colorado-Boulder
nrpace@colorado.edu

Jasmyn Pangilinan
DOE Joint Genome Institute
jlpangilinan@lbl.gov

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

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

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

Jayant Patil
DOE Joint Genome Institute
jmpatil@lbl.gov

Renè A. Perrier
DOE Joint Genome Institute
RAPerrier@lbl.gov

Yvette M. Piceno
Lawrence Berkeley National Laboratory
ympiceno@lbl.gov

Nicolas Pinel
University of Washington
dc4el@u.washington.edu

Samuel Pitluck
DOE Joint Genome Institute
s_pitluck@lbl.gov

Darren Platt
DOE Joint Genome Institute
dmplatt@lbl.gov

Martha L. Posada-Buitrago
DOE Joint Genome Institute
mlposada-buitrago@lbl.gov

Silke Pradella
German Collection of Microorganisms
and Cell Cultures (DSMZ)
silke_pradella@t-online.de

Flavia A. Ramirez
University of Illinois, Urbana-Champaign
framirez@uiuc.edu

Jason Raymond
Lawrence Livermore National Laboratory
raymond20@llnl.gov

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

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

Aaron O. Richardson
Indiana University
arichard@bio.indiana.edu

Kirsti M. Ritalahti
Georgia Institute of Technology
krita@ce.gatech.edu

Dan Rokhsar
DOE Joint Genome Institute
DSRokhsar@lbl.gov

Noah Rosenzweig
University of California
noahr@nature.berkeley.edu

Eddy Rubin
DOE Joint Genome Institute
EMRubin@lbl.gov

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

Nori Satoh
Kyoto University
satoh@ascidian.zool.kyoto-u.ac.jp

Thomas M. Schmidt
Michigan State University
tschmidt@msu.edu

Johannes C. Scholten
Pacific Northwest National Laboratory
johannes.scholten@pnl.gov

Gary P. Schroth
Solexa
gschroth@solexa.com

David A. Sela
University of California
dasela@ucdavis.edu

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

Harris Shapiro
DOE Joint Genome Institute
hjshapiro@lbl.gov

Thomas J. Sharpton
University of California, Berkeley
sharpton@berkeley.edu

Zhen Shi
University of Illinois, Urbana-Champaign
zshi@uiuc.edu

David S. Shin
Scripps Research Institute
davess@scripps.edu

Mark Silby
Tufts University School of Medicine
mark.silby@tufts.edu

Hauke Smidt
Wageningen University
hauke.smidt@wur.nl

Alexei V. Sorokin
Gènètique Microbienne CRJ INRA
sorokine@jouy.inra.fr

Kevin R. Sowers
University of Maryland
Biotechnology Institute
sowers@umbi.umd.edu

Shawn R. Starkenburg
Oregon State University
starkens@onid.orst.edu

Rob Steele
University of California, Irvine
resteele@uci.edu

Craig Stephens
Santa Clara University
cstephens@scu.edu

Jonathon Stillman
San Francisco State University
stillmaj@sfsu.edu

Wesley D. Swingley
Arizona State University
wswingley@msn.com

John A. Tainer
Scripps Research Institute
jat@scripps.edu

Hanno Teeling
Max Planck Institute for Marine Microbiology
hteeling@mpi-bremen.de

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

Richard H. Thomas
Southern Illinois University
rthomas@zoology.siu.edu

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

Ravi P. Tiwari
Murdoch University
r.tiwari@murdoch.edu.au

Christian M. Tobias
USDA, ARS
ctobias@pw.usda.gov

Tamas Torok
Lawrence Berkeley National Laboratory
ttorok@lbl.gov

Sucheta Tripathy
Virginia Polytechnic University and
State University
sutripa@vbi.vt.edu

Stephan Trong
DOE Joint Genome Institute
trong1@llnl.gov

Hank Tu
DOE Joint Genome Institute
hctu@lbl.gov

Jerry Tuskan
Oak Ridge National Laboratory
tuskanga@ornl.gov

Tina L. Tyler-Gresham
Idaho State University
tyletina@isu.edu

Gene W. Tyson
University of California, Berkeley
gtyson@nature.berkeley.edu

Parag A. Vaishampayan
parag_vai@yahoo.co.in

Fernando Valle
Genencor International
fvalle@danisco.com

Claire Vieille
Michigan State University
vieille@msu.edu

Attendees

Irene Wagner-Dobler
German Biotechnology Research Centre (GBF)
iwd@gbf.de

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

Jens Walter
University of Otago
jens.walter@stonebow.otago.ac.nz

Ronald A. Walters
Intelligence Technology Innovation Center
ra.walters@pnl.gov

Boris Wawrik
Rutgers University
bwawrik@yahoo.com

Sharlene C. Weatherwax
U.S. Department of Energy
sharlene.weatherwax@ science.doe.gov

David A. Weisblat
University of California, Berkeley
weisblat@berkeley.edu

Greg Werner
DOE Joint Genome Institute
gmwerner@lbl.gov

Rachel Whitaker
University of California, Berkeley
rwhitaker@nature.berkeley.edu

Larry J. Wilhelm
Oregon State University
wilhelml@science.oregonstate.edu

John Willis
Duke University
jwillis@duke.edu

Alexandra Z. Worden
Rosenstiel School of Marine and Atmospheric Science
aworden@rsmas.miami.edu

Martin Wu
The Institute for Genomic Research
mwu@tigr.org

Dongying Wu
The Institute for Genomic Reasearch
dwu@tigr.org

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

Jr-Kai Sky Yu
California Institute of Technology
jkyu@caltech.edu

Xueling Zhao
DOE Joint Genome Institute
xzhao@lbl.gov

Jizhong Zhou
University of Oklahoma
jzhou@rccc.ou.edu

Stephen H. Zinder
Cornell University
shz1@cornell.edu

Gerben J. Zylstra
Rutgers University
zylstra@aesop.rutgers.edu

Author Index

Authors

Notes

Notes