2012

U.S. Department of Energy

# Joint Genome Institute

Progress Report

● The DOE JGI's helical sculpture on the front cover was created by Jeff Brees of Markleeville, California.

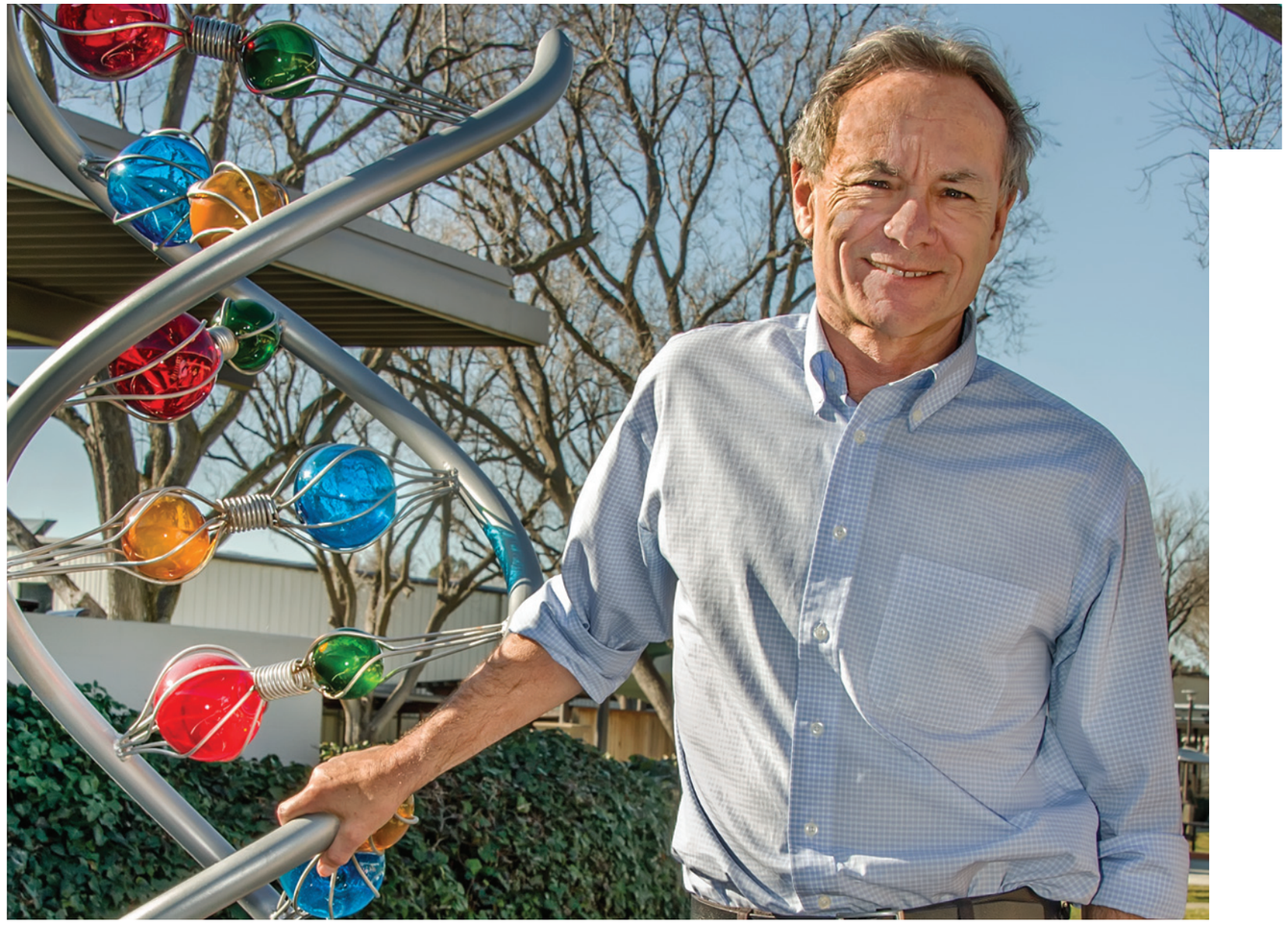
"It is interesting to contemplate a tangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about, and with worms crawling through the damp earth, and to reflect that these elaborately constructed forms, so different from each other, and dependent upon each other in so complex a manner, have all been produced by laws acting around us."

Charles Darwin

# Table of Contents

# DOE JGI Mission

The mission of the U.S. Department of Energy Joint Genome Institute (DOE JGI) is to serve the diverse scientific community as a user facility, enabling the application of large-scale genomics and analysis of plants, microbes, and communities of microbes to address the DOE mission goals in bioenergy and the environment.



# Director's Perspective

Rapid technological advances in genomics have transformed modern biology. From its inception, the Department of Energy Joint Genome Institute (DOE JGI) has been at the forefront of large-scale sequence-based science.

While sustaining its position as a leader in genome sequencing, the DOE JGI also has expanded its capabilities to achieve a deeper understanding of biological functions encoded by DNA. With the increasingly diverse data types it generates and associated computational strategies for their analysis, the DOE JGI is focused on enabling its community of users to make important new scientific discoveries to address major energy and environmental challenges.

In 2012, we completed 2,635 projects, a threefold increase over 2011. We also nearly doubled our sequencing output over the previous year, generating more than 56 trillion nucleotides of genome-sequence data from microbes and microbial communities, fungi, algae, and plants — the equivalent of nearly 19,000 human genomes. In our efforts to fill in gaps in the Tree of Life, we have added 650 genomes to the public databases just this past year. These achievements have solidified the DOE JGI as a world leader in genomics for energy and environmental research applications.

Many of our most significant accomplishments last year crossed the boundaries of taxonomy, exploring the vast complexity at the interfaces found in ecological niches. For example, through a large, interdisciplinary collaboration, we showed how the microbes associated with plant roots actually contribute to plant growth, productivity, carbon sequestration, and the ability to remediate contaminated soil and water. This study was featured on the cover of the journal *Nature*, highlighting the importance and impact of these findings.

A sampling of the many important studies completed over the past year:

- A comparative analysis of dozens of fungal species demonstrated that the end of the Carboniferous period 300-360 million years ago coincided with the evolution of white rot fungi, which break down the plant polymer lignin. The findings suggest that, instead of continuing to accumulate as peat that was transformed into coal, the bulk of plant biomass, after the appearance of white rot fungi, decayed and was released into the atmosphere as carbon dioxide (published in *Science*).

- We described the genome of foxtail millet (*Setaria italica*), a model plant for studies of the bioenergy crop switchgrass. The foxtail millet genome provided important insights into the genomic basis of traits such as cell-wall formation and for identifying the genes implicated in adaptations to particular environmental conditions (published in *Nature Biotechnology*).

- To speed up the biomass pretreatment process, researchers have been on a quest to find enzymes that are stable and active at higher temperatures. A comparison of the first end-to-end finished genomes of filamentous fungi (those with thread-like tendrils) identified enzymes that thrive in temperatures above 45°C. This discovery promises improvements for the initial step in converting fibrous materials from plants into a wide range of chemicals and biomass-based fuels (published in *Nature Biotechnology*).

- In a metagenomic study of samples collected from the Yellowstone National Park Obsidian Hot Spring and initiated through the DOE JGI 2005 Community Sequencing Program, the Great Lakes Bioenergy Research Center has generated its first issued patent for a heat-tolerant enzyme for improved biofuel production.

- In a study of the Deepwater Horizon oil spill, researchers from DOE JGI and Berkeley Lab found that marine microbial communities helped to disperse the oil plume that spread from the wellhead at a depth of 5,000 feet to the water's surface. Data analysis revealed an abundance of genes involved in the degradation of alkanes, as well as genes involved in degradation of aromatic compounds (published in *The ISME Journal*).

Additional highlights can be found on page 14 and a complete list of our publications can be found in Appendix F.

An important accomplishment in 2012 was the publication of our 10-year strategic plan, *Forging the Future — A Ten-Year Strategic Vision* (http://bit.ly/JGI-Vision). The DOE JGI mission is evolving to meet the needs of investigators and to assist them in making a real impact on the major energy and environmental scientific challenges of the future. To achieve this, researchers need access to a genomic user facility with a variety of capabilities. Massive-scale sequencing will remain one essential capability, but others are needed to solve these complex problems and to move the DOE JGI into a next-generation genome-science user facility:

- High-throughput experimental platforms to understand gene function

- Multidimensional genome annotation and data integration

- High-performance computing capabilities for challenges in large-scale sequence annotation

- Large-scale rapid DNA synthesis and genomic manipulations

- Organization of mission-oriented user communities

The DOE JGI strives to integrate these expanded activities in innovative and effective ways. This is critical if the biological sciences are to realize the full benefits and promise of genome sequencing. We have already begun to implement some of the major directions outlined in *The Ten-Year Strategic Vision*. These themes correspond well to the outcomes of a May 2012 workshop convened by DOE's Office of Biological and Environmental Research (http://bit.ly/0512GSProgram) exploring future directions for the DOE JGI.

For example, we are embarking on strategies to use the latest sequencing technologies to understand gene function. These efforts include analysis of nucleotide modifications (to understand new facets of gene expression regulation in plants and microbes) and high-throughput induction of changes in bacteria (to better understand how genes of previously unknown function contribute to fitness in a wide variety of growth conditions). These studies, and others like them, will enable the next generation of users to deploy new functional information to solve important problems in energy and the environment.

The escalating scale of the DOE JGI's current and future projects clearly requires a major investment in computational capabilities. Over the past two years, we have forged a strong alliance in high-performance computing with the National Energy Research Scientific Computing Center (NERSC). We have developed ongoing collaborations with DOE's new Systems Biology Knowledgebase (KBase) to accelerate our understanding of microbes, microbial communities, and plants. This community-driven open-source software framework and application system complements the DOE JGI's own portfolio of comparative analytical tools, such as the Integrated Microbial Genomes (IMG) suite, Phytozome (for plant comparative analysis), and MycoCosm (for fungal analysis). The traffic for these and the DOE JGI portal, the repository of all of our genome project data, is steadily rising.

Additionally, we have launched an effort to scale our in-house DNA synthesis capabilities to enable users to carry out sequence-to-function studies of unprecedented impact. Among the goals is to identify cellulose-degrading enzymes from microbes by sequence comparisons of complex environmental data sets, synthesize them, and determine their activities in high-throughput functional assays. Through this process, we are working with the DOE Bioenergy Research Centers to improve our ability to control the breakdown of plants, to reengineer microbes and fungi to produce enzymes that aid that process, and to alter plant cell walls of non-food biomass crops for renewable biofuels.

Since its founding 15 years ago, the DOE JGI remains committed to the notion of team science as conceived by Nobel Prize winner Ernest Orlando Lawrence, who founded Berkeley Lab more than 80 years ago. Our track record of productivity and marshaling the expertise and resources of interdisciplinary teams continues. In the near term, we will keep growing our network of complementary scientific interactions within the DOE national laboratory system and with other centers of excellence.

We recognize that the landscape is changing rapidly, both with the introduction of new technologies and the imperative to generate useful science from the sequencing data we produce.

I urge you to join us in tracking our progress toward fulfilling this ambitious agenda.

**Edward M. Rubin, MD, PhD**
*Director*
*DOE Joint Genome Institute*

# DOE Mission Areas

### Bioenergy

The United States is one of the world's largest consumers of petroleum and most of this energy is used for transportation and industry. This drives the DOE's focus on developing clean, sustainable alternative fuel sources. The search is on for fuels derived from cellulosic biomass — these fuels will offer energy on par with gasoline while fitting into our existing infrastructure. Sequencing projects at the DOE JGI that contribute to meeting this goal focus on one of three categories: developing plants that can be used as feedstocks for biofuel production, characterizing enzymes from fungi and microbes to break down the lignin and cellulose in plant walls, and identifying organisms that can ferment sugars into biofuels.

### Carbon Cycle

The global carbon cycle regulates the levels of atmospheric carbon dioxide and the Earth's climate. The carbon cycle is heavily dependent on the microbes that process and fix atmospheric carbon, promoting plant growth and degrading organic material. As microbes constitute the largest component of the Earth's biodiversity, understanding how they metabolize carbon, and how environmental changes affect these processes, is crucial. The DOE JGI is sequencing large numbers of microbes and microbial communities that contribute to carbon cycling. With this information, researchers can develop better predictive models that could provide more effective contributions toward reducing the effects of increasing carbon dioxide emissions on the global climate.

### Biogeochemistry

The carbon cycle is not the only process that regulates the natural environment, and the field of biogeochemistry explores the full spectrum of biological, physical, geological, and chemical processes and reactions involved. Microbes and microbial communities that can degrade or otherwise transform environmental contaminants such as toxic chemicals or heavy metals are another area of focus for the DOE JGI.

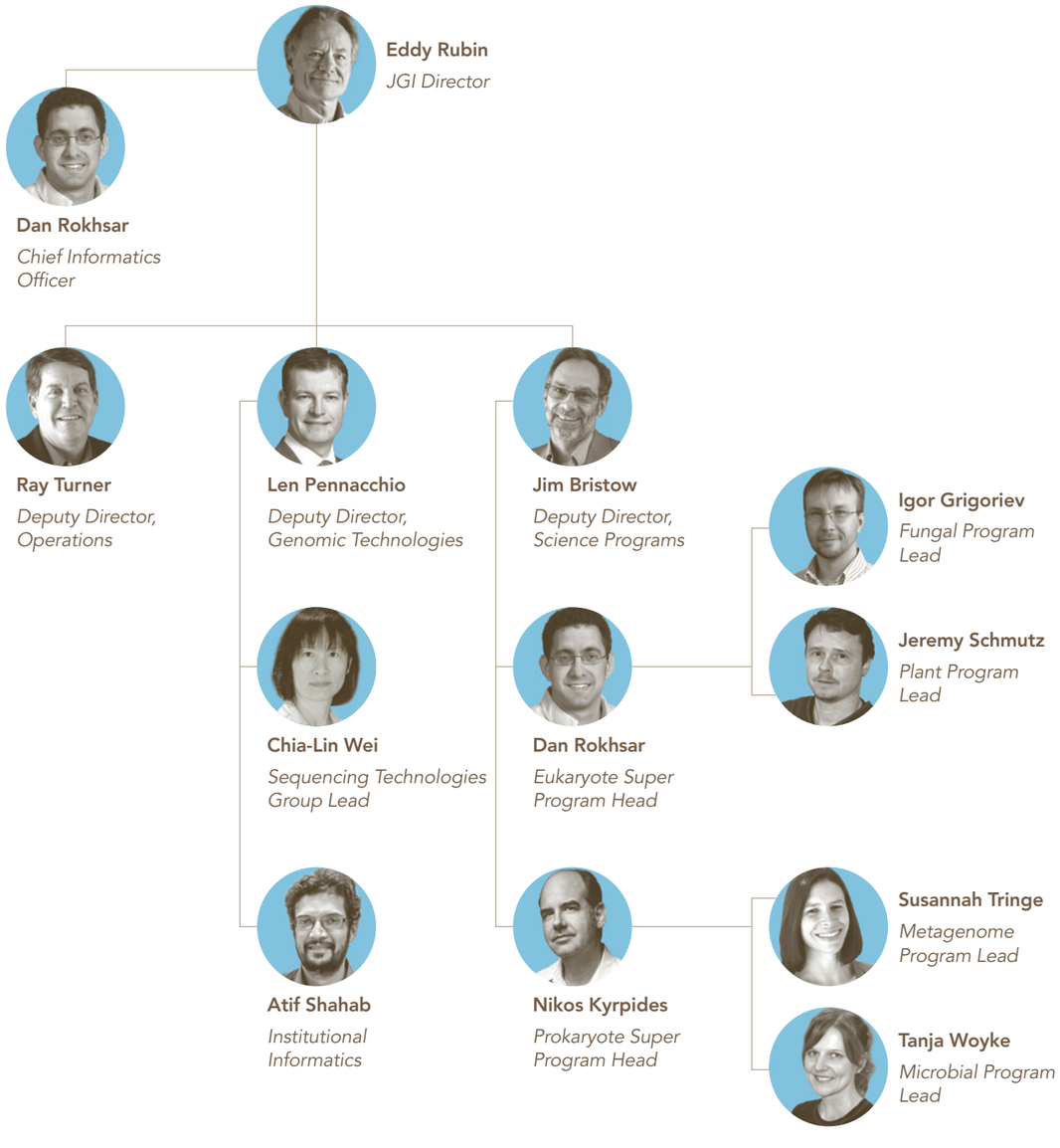
The DOE Joint Genome Institute:

# Impact

# 2012

# Organizational Structure

### Strategic Management




# The DOE Joint Genome Institute Partnership

With oversight provided by the DOE Office of Biological and Environmental Research (BER), the DOE JGI's Walnut Creek, California, headquarters draws its workforce from Lawrence Berkeley National Laboratory (Berkeley Lab) and Lawrence Livermore National Laboratory (LLNL). Additional partners and their respective roles:

Berkeley Lab's National Energy Research Scientific Computing Center (NERSC)
Computational and storage systems and support; high-performance computing applications

Los Alamos National Laboratory (LANL)
Microbial genome improvement, metagenome assembly R&D

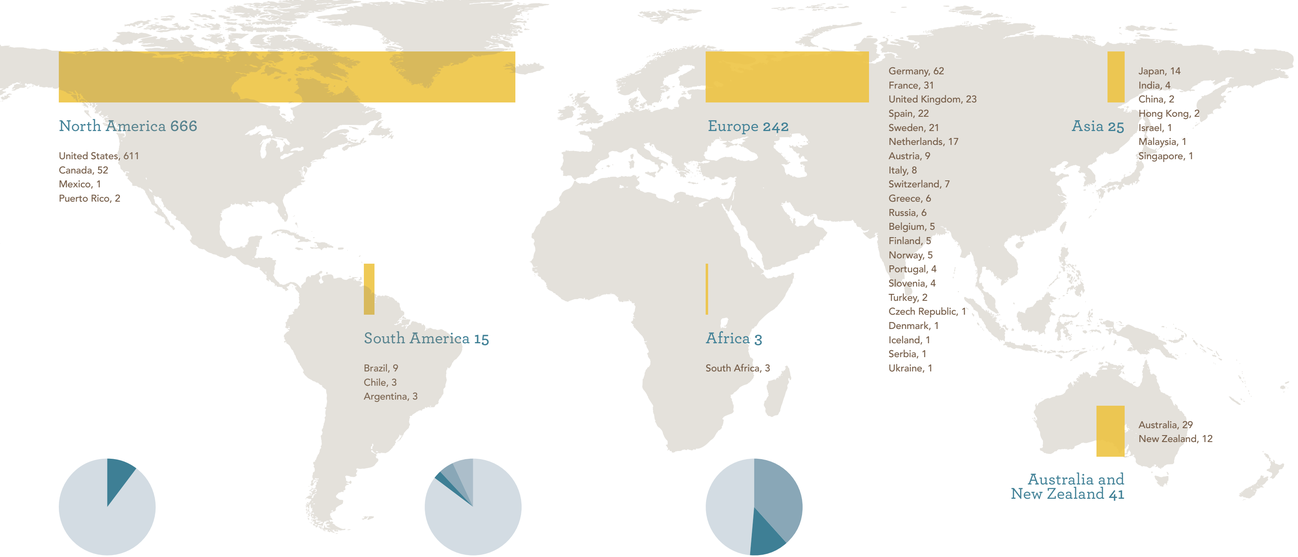
HudsonAlpha Institute for Biotechnology
Plant genome assembly, eukaryotic genome improvement

Oak Ridge National Laboratory (ORNL)
Plant genome biology

Energy Sciences Network (ESnet)
High-performance computing applications

## Fiscal Year 2012 Total Primary Users* = 992

*Principal investigators (PIs), co-PIs, collaborators, and community-based annotators that actively contribute to approved DOE JGI user projects during the fiscal year.



### North America 666

United States, 611
Canada, 52
Mexico, 1
Puerto Rico, 2

### South America 15

Brazil, 9
Chile, 3
Argentina, 3

### Europe 242

Germany, 62
France, 31
United Kingdom, 23
Spain, 22
Sweden, 21
Netherlands, 17
Austria, 9
Italy, 8
Switzerland, 7
Greece, 6
Russia, 6
Belgium, 5
Finland, 5
Norway, 5
Portugal, 4
Slovenia, 4
Turkey, 2
Czech Republic, 1
Denmark, 1
Iceland, 1
Serbia, 1
Ukraine, 1

### Africa 3

South Africa, 3

### Asia 25

Japan, 14
India, 4
China, 2
Hong Kong, 2
Israel, 1
Malaysia, 1
Singapore, 1

### Australia and New Zealand 41

Australia, 29
New Zealand, 12



### by Proposal Type

| | |
|---|---|
| Non-Bioenergy Research Centers (BRCs) | 890 |
| DOE BRCs | 102 |



### by Institution Type

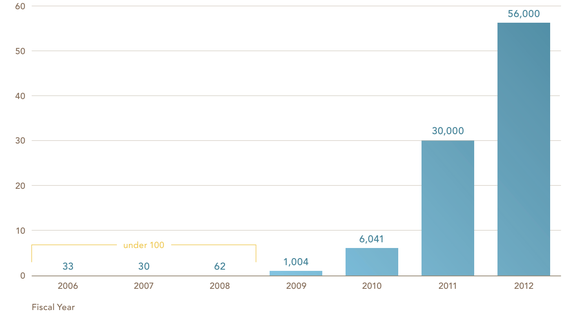
| | |
|---|---|
| Academic | 846 |
| DOE Laboratory | 68 |
| Government | 50 |
| Company | 28 |



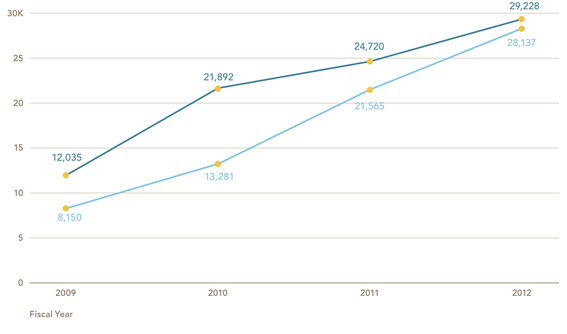
### by Location of Users

| | |
|---|---|
| USA (non-California) | 481 |
| Foreign | 381 |
| USA (California) | 130 |

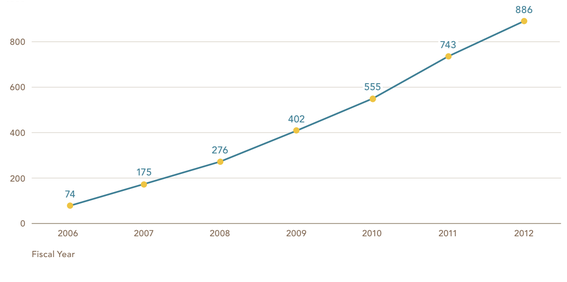
## Sequence Productivity (in billions of bases or GB)




## Citations

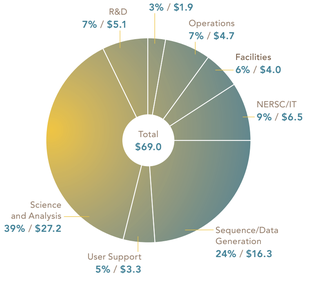



## Cumulative Number of Scientific Publications

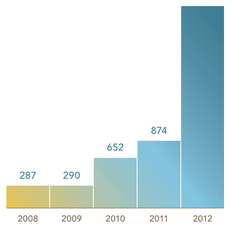



## FY12 Operating Expenses

(in millions)




## Projects Completed



# Science: Year in Review

The DOE JGI's sequencing efforts fall under the Eukaryote Super Program, which includes the Plant and Fungal Genomics Programs; and the Prokaryote Super Program, which includes the Microbial Genomics and Metagenomics Programs. In 2012, several projects made news for their contributions to energy and environment research.

## Energy

One of the biggest hurdles to achieving the federal Renewable Fuel Standard's goal of producing 36 billion gallons of biofuels annually by 2022 lies in optimizing the multistep process involved in breaking down plant biomass and then converting it into fermentable sugars that can be refined into fuel for our transportation needs. To help overcome this challenge, bioenergy researchers focus on identifying enzymes such as cellulases and ligninases from fungi and microbes.

An international team of scientists including DOE JGI researchers compared the finished fungal genomes of *Thielavia terrestris* and *Myceliophthora thermophila*, and the report appeared October 2, 2011, in *Nature Biotechnology*. These fungi thrive in high-temperature environments above 45°C and their enzymes are active at temperatures ranging from 40°C to 75°C, which would therefore be useful for accelerating (thus improving) the biofuel production process. /

(David Gilbert, DOE JGI)

"No matter how intently one studies the hundred little dramas of the woods and meadows, one can never learn all the salient facts about any one of them."

**Aldo Leopold**

*A Sand County Almanac: With Other Essays on Conservation from Round River*

**SEM image of the spider mite *T. urticae.*** *(Stephane Rombauts and Wannes Dermauw, Ghent University)*

**Inset: Web-spinning two-spotted spider mite.** *(M. Grbic)*

Tiny two-spotted spider mites extract the nutrients they need from leaves of more than 1,000 plant species, including bioenergy feedstocks and food staples. Chemically controlling their predations costs farmers $1 billion annually.

In the November 24, 2011, edition of *Nature*, an international team of researchers from more than 30 institutions reported on how the spider mite genome is shedding light on questions such as the pest's ability to rapidly develop resistance to pesticides, how this small creature can serve more broadly as a model for pest-plant interactions, and how it will likely respond in a changing environment. This research community is now employing the publicly available genomic data from the spider mite to advance the development of novel pest-control strategies as alternatives to chemical pesticides and to help to reduce environmental pollution. /

Cellulose and lignin are the most abundant biopolymers on the planet. The fungus *Phanaerochaete chrysosporium* (sequenced by the DOE JGI in 2004) and its close relative *Ceriporiopisis subvermispora* are found all over the world and interest bioenergy researchers because they possess enzymes that can break down plant biomass. An international team of researchers did a comparative genomic analysis of the two white rot fungi, reporting their results online the week of March 19, 2012, in the *Proceedings of the National Academy of Sciences*. The study revealed substantial differences among the sets of genes involved in lignocellulose degradation, providing further insight into the mechanics of how white rots do their dirty work.

"Very few fungi have the capability to degrade lignin," said study senior author and DOE JGI collaborator Dan Cullen of the U.S. Department of Agriculture Forest Service, Forest Products Laboratory (FPL). "Even fewer fungi have the ability to selectively remove lignin at such an efficient rate. *C. subvermispora* is one exception in its ability to do just that." /

Switchgrass is a perennial grass that the DOE is investigating as a prospective biofuels feedstock. Assembling its genome, however, is challenging because switchgrass has multiple copies of its chromosomes. Working in tandem with several institutions including the BioEnergy Science Center (BESC) and the DOE Joint BioEnergy Institute (JBEI) — two of the three DOE Bioenergy Research Centers — the DOE JGI has sequenced plant genomes of related candidate bioenergy crops such as sorghum and the model grass *Brachypodium*. Though both plants have been used as references for switchgrass, they last shared a common ancestor with the grass several million years ago. The genome of a much closer switchgrass relative — foxtail millet (*Setaria italica*) — was described in the May 13, 2012, edition of *Nature Biotechnology*.

"We're not thinking of *Setaria* as a biofuel crop per se but as a very informative model since its genome is so structurally close to switchgrass," said Jeff Bennetzen, a BESC researcher, the study's co-first author, and a professor at the University of Georgia. /

The coal used to generate a significant amount of electricity used by the United States is sourced from plant matter that was fossilized around 360 to 300 million years ago. DOE JGI researchers were part of the team that presented evidence in the June 29, 2012, edition of the journal *Science* suggesting that the most productive coal-forming period came to a close because of an ancestor of white rot fungi.

"We're hoping this will get into the biology and geology textbooks," said Clark University biologist David Hibbett, senior author of the comprehensive study comparing the complete genomes of dozens of species of fungi, most of which were sequenced at the DOE JGI. "When you read about coal formation it's usually explained in terms of physical processes, and that the rate of coal deposition just crashed at the end of the Permo-Carboniferous. Why was that? There are various explanations."

The explanation Hibbett and his colleagues are now following involves the evolution of fungi capable of breaking down the polymer lignin. Until this ancestral fungus' arrival, plant biomass was converted into coal during the Carboniferous period. In the fungus' presence, dead plant matter was broken down completely.

"Once you have white rot you can break down lignin, the major precursor of coal," Hibbett said. "So the evolution of white rot is a very important event in the evolution of carbon cycle." /

## Environment

The frozen soils of the Arctic trap an estimated 1,672 billion metric tons of carbon from the Earth's atmosphere and as the world warms, scientists are concerned that its release could become a major source of greenhouse

gases. As such, microbes residing in the Arctic are key players in the thawing permafrost's potential impact on the global carbon cycle.

Researchers from the DOE JGI, Berkeley Lab's Earth Sciences Division, and the U.S. Geological Survey collaborated to learn more about the permafrost microbial community. Among the findings, published online November 6, 2011, in the journal *Nature*, is the draft genome of a novel microbe that produces methane, a far more potent greenhouse gas than carbon dioxide. This microbe, not yet named, lives in the permafrost, and was assembled out of the metagenomic fragments isolated from the frigid soil.

"By applying metagenomics to study microbial community composition and function, we can help to answer questions about how the currently uncultivated and unstudied microbial species residing in permafrost cycle organic carbon and release greenhouse gases during thaw," said DOE JGI scientist and study corresponding author Janet Jansson. "This will provide valuable information that could lead to improved carbon cycle models and eventual mitigation strategies." /

After the Deepwater Horizon oil spill, several studies confirmed that various microbes played a role in the dispersal of the oil in the Gulf of Mexico. To learn more about the composition of the microbial community that responded to the spill, researchers led by DOE JGI scientist Janet Jansson collaborated with scientists at two of the Laboratory's national user facilities, the DOE JGI and the Advanced Light Source (ALS). The findings appeared in two articles and track a series of microbial species dominating the community in the waters at various times and thought to be involved in the removal of different fractions of the oil.

In the report published June 21, 2012, in *The ISME Journal*, the team described using a combination of genomics techniques to study the way the microbes responded to the influx of oil. They focused on the community's expressed functional information or metatranscriptome. In addition, they isolated single cells to identify the predominant microbial members in the deep-ocean oil plume. Using the latter technique, they were able to assemble a draft genome of what they say is the first deep-sea, oil-eating bacterium from a single cell. /

The button mushroom *Agaricus bisporus* is a familiar ingredient in many dishes but the forest fungus also plays an important role in the carbon cycle. Unlike many fungi that break down wood, said Fungal Genomics Program head Igor Grigoriev, "*Agaricus* fits neither brown rot nor white rot classifications." As he and his colleagues noted in the study published October 8, 2012, in the *Proceedings of the National Academy of Sciences*, the button mushroom has adapted to growing in a leaf-litter, humus-rich environment.

Humus contributes chemicals that drive the decomposition process, as well as adding organic matter to deficient soils and contributing to overall plant health. Analysis of two *Agaricus* strains revealed the presence of several families of well-known sugar-degrading enzymes similar to the repertoire found in wood-decaying fungi. However, the enzymes in *Agaricus* such as heme-thiolate peroxidases and etherases predominate in the presence of humus-rich soil habitats, suggesting a higher ability to metabolize complex mixtures of derivatives of lignin and other polymers.

"Our understanding of the carbon cycling role of *Agaricus* in ecosystems is a prerequisite to modeling and optimizing carbon management for sustainable forests," said study senior author Francis Martin, Head of the ARBRE Lab of Excellence at INRA, Nancy, France. /

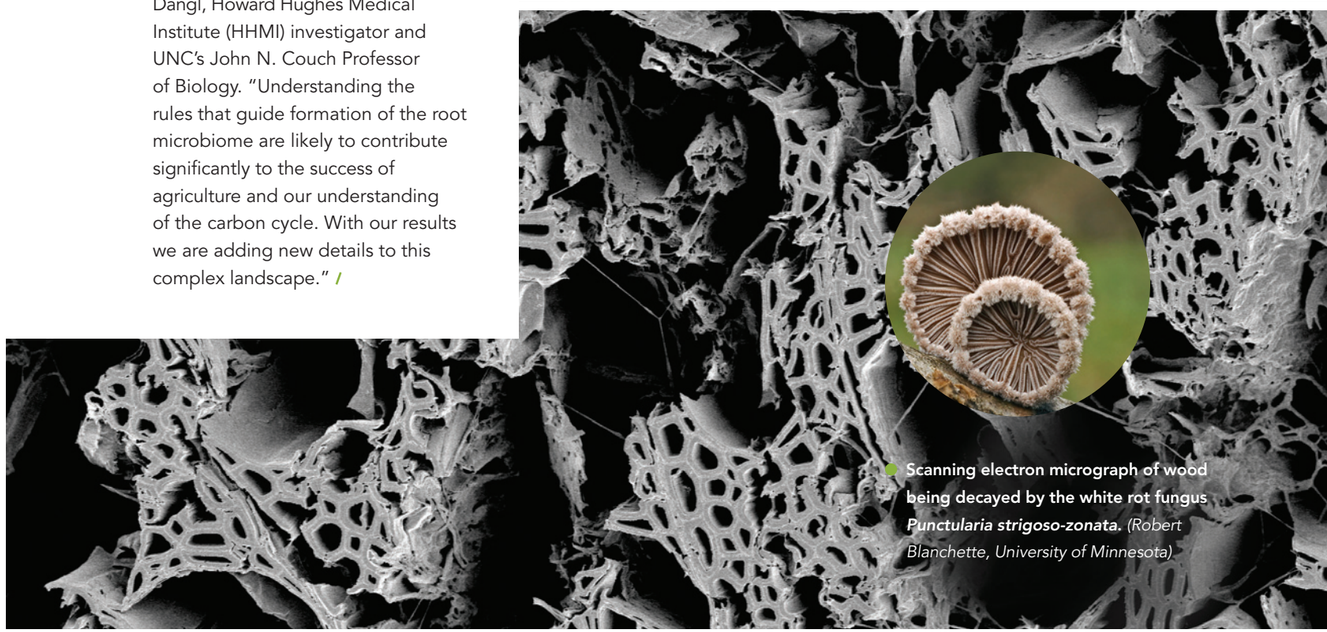
***Setaria viridis* (green foxtail) is one of many plants sequenced by the DOE JGI.** *(Kropsoq, Wikimedia Commons)*



**A surface slick in the Gulf of Mexico, taken ~1.5 km from the Deepwater Horizon wellhead.** *(Olivia Mason, LBNL).*

The microbial community inhabiting the rhizosphere and endosphere niches immediately surrounding and inside a plant's root facilitates the shuttling of nutrients and information into and out of the roots within the soil matrix. Interactions between plant roots and microbes contribute to plant health and growth through processes such as defending against pests and managing carbon and other elements for the enrichment of the soil. These underground microbial activities were highlighted in an August 2, 2012, publication in *Nature* from a team led by researchers from the DOE JGI and the University of North Carolina (UNC), Chapel Hill. The team was able to pinpoint the identity of many of the microbial players so their metabolic roles could be inferred, as well as where and when they were pressed into service.

"Science has long been fascinated with the spectrum of relationships between plant and microbe that span from pathogenic to mutually beneficial," said senior author Jeff Dangl, Howard Hughes Medical Institute (HHMI) investigator and UNC's John N. Couch Professor of Biology. "Understanding the rules that guide formation of the root microbiome are likely to contribute significantly to the success of agriculture and our understanding of the carbon cycle. With our results we are adding new details to this complex landscape." /



**Scanning electron micrograph of wood being decayed by the white rot fungus *Punctularia strigoso-zonata*.** *(Robert Blanchette, University of Minnesota)*

# Technologies and Informatics

The primary output of the Genomic Technologies Department continues to be at the forefront of advances in high-throughput DNA sequencing and analysis, enhanced by robust sample management and library construction.

Over the past two years, the DOE JGI has consolidated primary sequencing operations onto the Illumina HiSeq 2000s, complemented by the introduction of MiSeq instruments for quick project turnaround. The long reads provided by Pacific Biosciences (PacBio) RS single-molecule DNA sequencers ensure the DOE JGI has the full suite of sequencing technologies to meet user needs. For a detailed description of the Illumina HiSeq and PacBio RS systems, see Appendix B.

Additional efforts are devoted to exploring new opportunities for technological access to DOE JGI user services, including the use of large-scale single-cell genomics to study hard-to-culture environmental microbes. The fact that only a small fraction of microbes are currently readily culturable in the laboratory presents a substantial obstacle for exploring the biology of the vast majority of microbes and their potential relevance to energy and environmental applications. Culture-independent approaches such as metagenomics and, more recently, metatranscriptomics — the expressed subset of genes within a microbial community at a given point in time or in response to a given treatment or perturbation — provide a path to understanding the uncultured microbial biosphere and to addressing many questions of DOE relevance. Emerging single-cell technologies such as micro-and nanofluidic approaches provide a powerful complementary strategy to access the genetic makeup of individual uncultured community members, eliminating key challenges of metagenomic approaches such as the proper assembly and binning of complex data sets. This will also enable cell preparations for complementary single-cell proteomics and metabolomics studies of the same specimens by users, enabling systems-level studies at single-cell resolution.

Post-sequencing capabilities including gene expression (RNA) profiles and metatranscriptomics are now being offered to users. Additionally, to accelerate the linking of sequence to function, the DOE JGI is exploring the development of rapid and inexpensive approaches to designing and creating DNA fragments encompassing genes and larger segments of DNA. The DOE JGI's primary role in synthetic DNA projects will be to support users in the computational design of desired target constructs, in the creation of these large and complex DNA molecules, and in their introduction into suitable host cells. To support users in their ability to generate synthetic systems required to address energy and environmental challenges, the DOE JGI will also develop experimental strategies in which functional readouts can be closely linked to synthetic sequence.

## Informatics

Informatics supports three main areas of the DOE JGI's activities: sequencing project management, sequencing, and scientific programs. Specific science program informatics systems provide support for sequencing data processing, analysis, integration, and publication. Informatics systems seek to maintain high throughput (data sets processed weekly, monthly, and quarterly) and quality of service (reliability, robustness, and performance). The genome sequence data processing and integration activities of the DOE JGI's Science Programs have a "production" nature, in which tools developed based on computational biology methods are applied on data sets within the context of program-specific informatics systems. The Institute's comparative-analysis systems have matured over the past years and are recognized as important resources for conducting genome and metagenome studies, empowering scientists around the world to conduct studies that otherwise would be very expensive or out of reach.

Building upon the DOE JGI's seminal contributions to the Human Genome Project, DOE JGI researchers made contributions to the genomic techniques and data-management requirements of two global-scale projects related to this work. A team led by Genomic Technologies Deputy Director Len Pennacchio was part of the international consortium of researchers working on the Encyclopedia of DNA Elements (ENCODE) project, a follow-up to the Human Genome Project in that it set out to catalog the functional sequences in those 3 billion bases. Amid the flurry of publications released by the ENCODE consortium in September 2012, one of the techniques that proved useful to understanding that a huge proportion of the human genome is involved in gene regulation came from ChIP-Seq experiments that use next-generation sequencing technology to identify regulatory regions of DNA that turn gene expression on or off.

A second project involved developing and maintaining a comparative analysis system, built from the DOE JGI's well-established Integrated Microbial Genomes (IMG) suite of tools, called the Integrated Microbial Genomes and Metagenomes for the Human Microbiome Project (IMG/M HMP). To better understand the roles and functions of the microbial communities in the human body, the normal microbial makeup of healthy humans has been mapped for the first time in a major effort funded by the National Institutes of Health. As part of the larger IMG/M data system overseen by Prokaryote Super Program head Nikos Kyrpides, the IMG/M HMP "data mart" allows scientists worldwide to study big data sourced from human microbiome samples within the context of reference genomes of individual microbes that help scientists identify the microbes in a sample.

In June, several publications in *Nature* and *Public Library of Science* journals detailed five years of research from the HMP Consortium, which involves some 200 members from nearly 80 research institutions. With Victor Markowitz, head of Berkeley Lab's Biological Data Management and Technology Center that oversees the development and maintenance of the IMG/M system, Kyrpides is a co-principal investigator of HMP's Data Analysis and Coordination Center. The work also involves supercomputers at NERSC, and networking resources from ESnet.




> "The machine does not isolate man from the great problems of nature but plunges him more deeply into them."
>
> Antoine de Saint-Exupéry

# Education and Outreach

(David Gilbert, DOE JGI)

Within the DOE JGI's Genomics and Bioinformatics Education Program overseen by Cheryl Kerfeld, the Undergraduate Research in Microbial Genome Annotation — or Interpret a Genome Program — provides college and university students access to recently sequenced microbial genomes, such as those of organisms from little-known branches of the Tree of Life selected as part of the DOE JGI's Genomic Encyclopedia of Bacteria and Archaea (GEBA) project.

The students compare and annotate the genomes in the context of their own coursework, gaining hands-on knowledge of genomics and bioinformatics.

As their annotation platform, students use the Integrated Microbial Genomes Annotation Collaboration Toolkit (IMG-ACT), a wiki/Web portal fusion that lets them work with existing genome data sets and record their discoveries. The platform is the result of a collaboration between the Education Program and faculty members from several universities around the country. Through events such as workshops held at the DOE JGI, traveling workshops supported by an NSF award (RCN-UBE-DBI 0954829), and the annual ASM-JGI Bioinformatics Institute in Washington, D.C., the Genomics and Bioinformatics Education Program helps keep undergraduate and graduate faculty current on the latest in genomics-based research and assists in incorporating genomics and bioinformatics into their life-science curricula. The program has become so successful that an entire issue of *Biochemistry and Molecular Biology Education* (January 2013) highlights the diverse ways faculty use the IMG-ACT system across the life sciences curriculum.

Since the program launch in 2008, over 5,600 students under more than 250 instructors at 100+ institutions have used IMG-ACT.

The DOE JGI has also been proactive in putting its computational tools in the hands of the genomics community. In the past year alone, more than half a million unique visitors accessed the various systems developed to allow the genomics community to use these genomic data sets and tools. Several times a year, researchers from around the world gather for five-day Microbial Genomics and Metagenomics workshops built around the IMG system. These workshops include two days of intensive seminars and three days of hands-on tutorials and are a community resource for comparative analysis and annotation of all publicly available genomes from three domains of life in a uniquely integrated context. The DOE JGI also holds tutorials focused on its other data repositories, including the Phytozome plant genomics portal and the MycoCosm fungal genomics portal.

Aside from allowing students and researchers to explore DNA sequences, the DOE JGI recently turned its efforts toward introducing the general public to the concepts of DNA synthesis and synthetic biology. In collaboration with the Joint BioEnergy Institute, the DOE JGI is developing Bioscriber, a software module that shows how information from nature in the form of enzymes identified through genome sequencing can be harnessed to break down and convert plant biomass into advanced, non-ethanol biofuels. The tutorial level of Bioscriber debuted at the Open House hosted by Berkeley Lab on October 13, 2012, and the next phase of the ongoing project will introduce additional features.

# Safety

The DOE JGI Safety Team — a network of groups working to maintain a strong, robust safety culture — recognizes that world-class scientific achievement requires a safe working environment.

In recent years, the DOE JGI has seen a significant reduction in recordable injuries, the result of continuing efforts to develop a stronger employee safety culture, increase management feedback and involvement in safety, and improve existing safety programs. In the past year, there have been no recordable injuries.

The DOE JGI has shared its success in changing safety culture and creating a proactive program by contributing several best practices and other resources, which can be found at: http://www.jgi.doe.gov/whoweare/ergonomics/.

Several employee-led safety groups actively promote and support safety at the DOE JGI. One of the most active is the Safety Culture Committee, composed of employees who promote safety culture and provide safety-related feedback to management and the Employee Safety Committee. The group generates safety posters, conducts periodic safety-related surveys, sponsors an employee safety recognition program, sponsors safety theme months, and conducts annual safety fairs.

Several projects related to safety and ergonomics have been entered in the annual Applied Ergonomics Ergo Cup® competition. In 2012, the DOE JGI entry to the internationally recognized competition was titled, "In the Beginning, There IS Ergo." Submitted in the Ergonomic Program Improvement Initiatives category, the DOE JGI team showcased the Lean Six Sigma-driven approach to strategic planning that has led to ergonomics being incorporated earlier upstream into new-employee orientation and new-technology implementation procedures. Competing against large companies including GE Healthcare, Honda, Boeing, and Toyota, the DOE JGI team brought home the first runner-up plaque.

Another important employee-led safety group is the Emergency Response Team (ERT). In the event of a major disaster, this volunteer group of employees provides emergency response and first-aid services until professional responders can arrive. The ERT also runs annual emergency evacuation drills, using Community Emergency Response Training (CERT), which is sponsored by the Federal Emergency Management Agency (FEMA), as the training model for its members. The CERT-based training program developed at the DOE JGI has been adopted by Berkeley Lab, with several ERT members from the DOE JGI serving as instructors.

"If you are going to do an experiment, you need to think two or three moves ahead: 'This is what I think is going to happen...This is what might happen instead...This is how I'd respond if it does.'"

**Berkeley Lab Director Paul Alivisatos** *on building a safety culture based on the same logic that drives the scientific method*

# Appendices



## Appendix A:

## Glossary

**Annotation:** The process of identifying the locations of genes in a genome and determining what those genes do to improve accuracy of genetic information collected.

**Archaea:** One of the three domains of life (Eukarya and Bacteria being the others) that include primitive microorganisms that can tolerate extreme environmental conditions (temperature, acid, etc.).

**Assembly:** Compilation of an organism's overlapping DNA sequences that have been clustered together based on their degree of sequence identity or similarity.

**Barcoding:** The practice of appending known unique synthetic DNA sequences to sequencing libraries to allow pooling of libraries for next-generation sequencing, after which sequence data can be assigned to particular libraries or samples based on the barcode sequence.

**Base:** A unit of DNA. There are four bases: adenine (A), guanine (G), thymine (T), and cytosine (C). The sequence of bases is the genetic code.

**Base pair:** Two DNA bases complementary to one another (A and T or G and C) that join the complementary strands of DNA to form the characteristic double helix of DNA.

**Bioremediation:** The use of microorganisms to break down contaminants and other unwanted substances in waste and other substances.

**Biogeochemistry:** A study of the biosphere's interactions with the Earth's chemical environment.

**Bioinformatics:** The use of computers to collect, store, and analyze biological information.

**Bioprospecting:** Searching nature for genes and proteins that can be applied to help scientists solve energy and environment challenges.

**Bioscriber:** A scientist who uses synthetic biology to coax microbes into producing helpful compounds such as biofuels or antibiotics.

**Bridge amplification:** A proprietary technique used by Illumina sequencing platforms to generate single-stranded clusters of template DNA.

**Carbon cycle:** The biogeochemical process by which carbon is exchanged between the planet's atmosphere, land, and oceans.

**CAZymes:** Carbohydrate-Active enzymes that can break down plant polysaccharides such as cellulose into small sugars.

**Contig:** A group of cloned (copied) pieces of DNA representing overlapping regions of a particular chromosome.

**Coverage:** The number of times a region of the genome has been sequenced during whole-genome shotgun sequencing.

**Curation:** Analysis of genome annotations to improve and maintain data presentation.

**Draft genome:** The term for an incomplete genome sequence. It can be applied to a wide range of sequences, from those that have the minimum amount of information needed for submission to a public database, to assembled genomes that have undergone manual and automatic review but still have sequence errors that need to be corrected.

**Enzyme:** A protein used to induce or speed up a chemical reaction.

**Eukaryotes:** The domain of life containing organisms that consist of one or more cells, each with a nucleus and other well-developed intracellular compartments. Eukaryotes include all animals, plants, and fungi.

**Finished genome:** In accordance with the 1996 Bermuda standard, a gapless sequence with a nucleotide error rate of one or less in 10,000 bases.

**Flow cell:** Resembles a microscopic slide only with eight channels on which DNA samples are loaded for analysis on the Illumina sequencing platforms.

**Fluorescence-activated cell sorting:** A specialized type of flow cytometry used to study and purify cells. A heterogeneous mixture of cells passes through laser beams and is sorted into two or more containers, one cell at a time, based upon the specific light-scattering and fluorescent characteristics of each cell.

**Fosmid:** A vector suitable for cloning genomic inserts approximately 40 kilobases in size.

**GenBank:** Open-access, publicly available collection of annotated sequences submitted by individual laboratories and large-scale sequencing centers that is overseen by the National Center for Biotechnology Information.

**Halophile:** A microbe that thrives in environments with high salinity.

**Informatics:** The study of the science of information.

**Library:** An unordered collection of clones containing DNA fragments from a particular organism or environment that together represent all the DNA present in the organism or environment.

**Mapping:** Charting the location of genes on chromosomes.

**Metagenomics (also environmental genomics or community genomics):** The study of genomes recovered from environmental samples through direct methods that do not require isolating or growing the organisms present. This field of research allows the genomic study of organisms that are not easily cultured in a laboratory.

**Metatranscriptomics:** The study of the region of the complete genetic code that is transcribed into RNA molecules and provides information on gene expression and gene function.

**Microbe:** Another name for a microorganism.

**Microbiome:** A defined environment within which a community of microbes exists and interacts.

**Molecular Cloning:** The use of specialized DNA technology to produce multiple exact copies of a single gene or other segment of DNA to obtain enough material for further study.

**Multiple displacement amplification (MDA):** Method of amplifying tiny amounts of DNA in a cell so that it can be used for sequencing through single-cell genomics.

**Paired-end reads:** DNA library preparation technique that lets researchers look at both the forward and reverse template strands of a large DNA fragment and that provides positional information.

**Polymerase chain reaction (PCR):** A method of DNA amplification.

**Phylogeny:** The evolutionary history of a molecule such as a gene or protein, or a species.

**Prokaryotes:** Unlike eukaryotes, these organisms, (e.g., bacteria) are characterized by the absence of a nuclear membrane and by DNA that is not organized into chromosomes.

**Promoter:** A region of DNA that sends signals to a cell to tell it where a gene begins and when the gene is read. An inducible promoter only signals the cell under certain conditions while a constitutive promoter is always signaling the cell.

**Psychrophile:** A cold-loving microbe that optimally grows in environments with temperatures of 15°C (60°F) or less.

**Read length:** The number of nucleotides tabulated by the DNA analyzer per DNA reaction well.




**Rhizosphere:** Microecosystem defined by a thin layer of soil where plant roots interact with microorganisms in the soil.

**Sequence:** Order of nucleotides (base sequence) in a nucleic acid molecule. In the case of DNA sequence, it is the precise ordering of the bases (A, T, G, C) from which the DNA is composed.

**Sequencing by synthesis:** Proprietary sequencing technique used by Illumina systems in which four fluorescently labeled nucleotides determine the sequence of a DNA fragment, one base at a time.

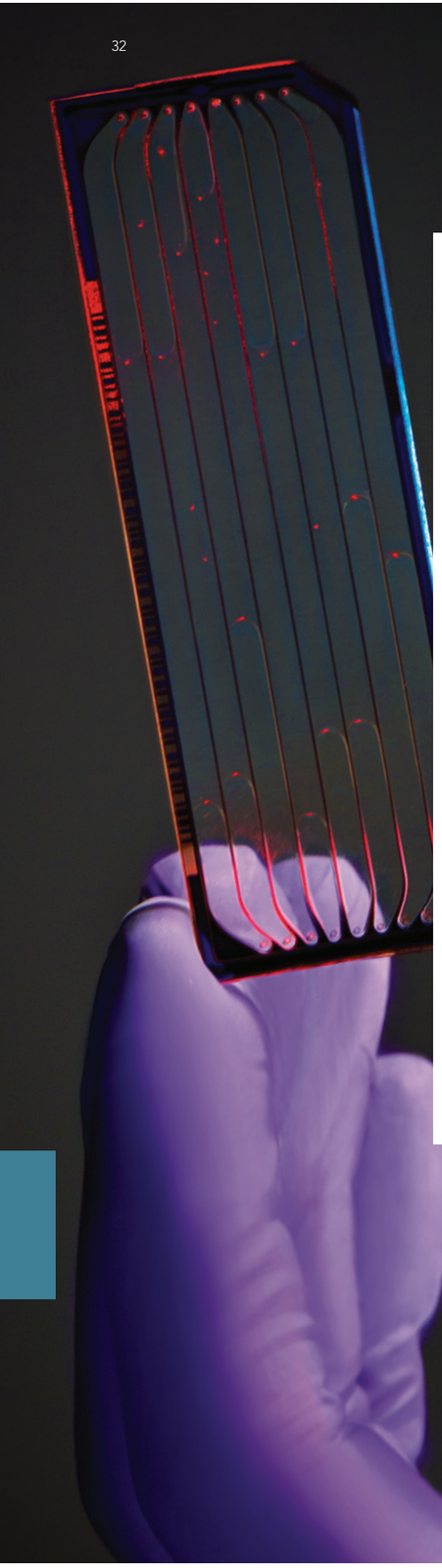
**Single-cell genomics:** Method for sequencing a genome using DNA derived from a single cell that is used to study uncultured or nonculturable organisms.

**Single-molecule real-time (SMRT) sequencing:** Single-molecule DNA sequencing performed in zero-mode waveguide (ZMW) chambers on a chip.

**Subcloning:** The process of transferring a cloned DNA fragment from one vector to another.

**Synthetic biology:** A field of research concerned with purposeful editing of biological systems. For the DOE JGI's objectives, this process refers to assembling DNA sequence fragments with the goal of synthesizing sequences to experimentally validate their functions and applications.

**Transcriptome:** A collection of all the RNA transcripts in a given cell that serves as a snapshot of global gene expression.



# Appendix B:
# Sequencing Technologies

Compared with sequence data generated in 2007 by the Sanger platform, adoption of Illumina technologies and subsequent platform improvements increased sequence data generation over a thousandfold by 2012. The DOE JGI has consolidated primary sequencing operations onto the Illumina HiSeq 2000 platform, incorporated MiSeq instruments into the sequencing process, and has installed a second PacBio RS single-molecule DNA sequencer.



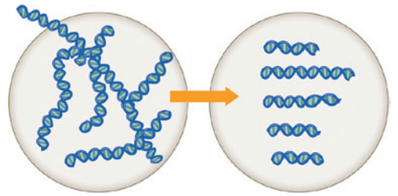
## Illumina HiSeq Sequencing Technology

The Illumina approach relies on attaching fragmented template genomic DNA prepared in a sample library to a planar, optically transparent surface on a flow cell and amplifying it to form colony clusters of the template. These clusters of templates are sequenced using a four-color DNA sequencing-by-synthesis technology that employs reversible dye-terminators with removable fluorescence and blocking groups. This highly parallel approach can generate 1 billion paired sequencing reads and 325 billion bases (gigabases) per 2x150 flow-cell run. The DOE JGI currently operates eight of these HiSeq instruments.

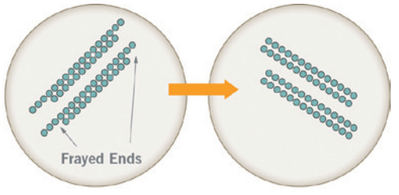
The four types of fluorescently labeled nucleotides are pumped into the flow cell at each cycle and complementary nucleotides are incorporated into the nucleic acid chain being synthesized in each cluster of templates. The incorporated nucleotides contain a blocking group attached to the fluorescent label that prevents any additional nucleotides from being incorporated. Using laser excitation, the incorporated nucleotide's fluorescent label is imaged and the nucleotide's identity is determined. The blocking group and fluorescent label are then cleaved off of the nucleotide and the sequence cycle is repeated to determine the next nucleotide incorporated. Each of the sequencing cycle images is compiled and processed to produce base sequences for each DNA template cluster. Applications are *de novo* sequencing, where there is no reference available, and resequencing, where short sequence reads are aligned against a reference. The genetic differences on the sequences are called using a specially developed data pipeline.

In late 2011, the DOE JGI implemented Illumina's latest instrument, the MiSeq, which offers a smaller-scale, cost-effective alternative rapid-sequencing platform capable of performing 2x150 runs in 24 hours as compared with the 16.5 days it takes on the HiSeq. While the MiSeq only generates a fraction of the data of a HiSeq — about 3 Gb on a MiSeq flow cell versus 325 Gb on a HiSeq flow cell — the rapid run times make it beneficial for applications where minimal sequencing reads are needed or when sequencing data are needed quickly. The MiSeq platform is also capable of longer read lengths and currently supports runs up to 2x250. The DOE JGI currently operates six of these MiSeq instruments.

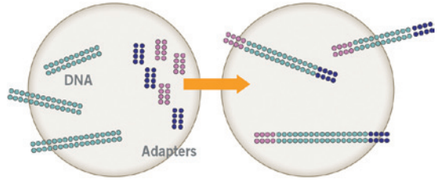
1. **Sonication**
Genomic DNA is fragmented into 100-500 base-pair fragments by sonication to create a library.



2. **Fragment End Repair**
Sonication creates frayed DNA ends that must be blunted or repaired.

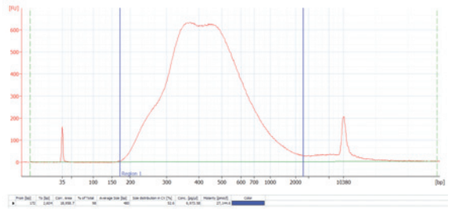



3. **A-Tailing and Adapter Ligation**
Adapters are ligated to each end of the A-tailed DNA fragment.

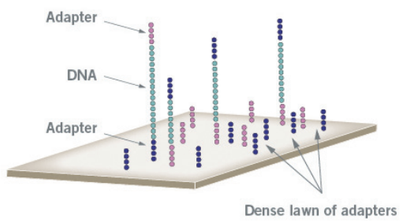


Appendix B continued

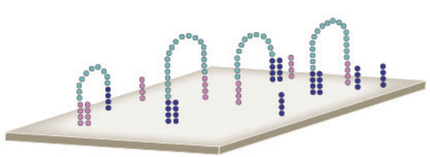
**4. QC Check**
The electropherogram shows the size and concentration of the final library. This library size also confirms the ligation of adapters.



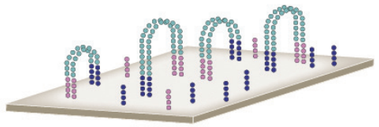
**5. cBOT Cluster Generation System**
Sodium hydroxide creates single-stranded DNA that is then randomly bound to the top and bottom of each channel in the flow cell.



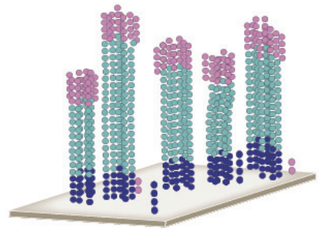

**6. Bridge Formation**
Free DNA end binds to complementary primer to form a bridge.



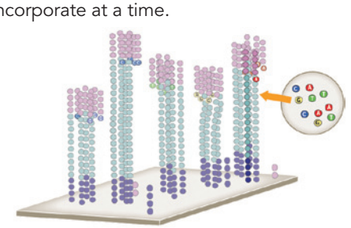
**7. Bridge Amplification**
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification. Fragments become double-stranded DNA bridges. Thirty-five cycles of amplification create clusters of identical DNA fragments.



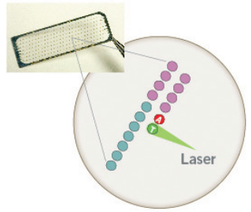
**8. Finished Flow Cell**
By completion of amplification, several million dense clusters of single-stranded DNA have been generated in each channel of the flow cell, with a sequencing primer attached.



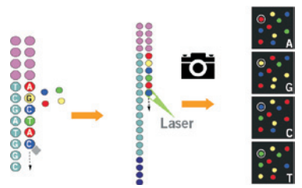
**9. DNA Sequencing**
To initiate the first sequencing cycle and determine the first base, all four labeled reversible terminators and DNA polymerase enzyme are first added. Only one base can incorporate at a time.



**10. Base Calling**
Lasers excite the fluorescent tags and the images are captured via CCD camera. The identity of the first base in each cluster is recorded, then the fluorescent tag is removed.




**11. Sequencing by Synthesis**
In the first cycle, the first base is incorporated. Its identity is determined by the signal given off and then recorded. In subsequent cycles, the process of adding sequencing reagents, removing unincorporated bases, and capturing the signal of the next base to identify is repeated.




**12. Dual Flow Cells**
Once the top surface of the flow-cell channel has been scanned, the imaging step is repeated on the bottom surface.

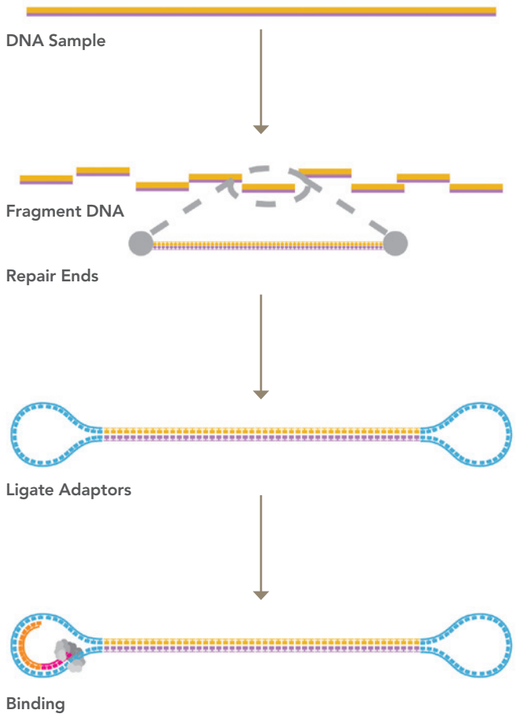
## PacBio Sequencing Technology

The Pacific Biosciences single-molecule real-time (SMRT™) DNA sequencer utilizes nano-scale zero-mode waveguide (ZMW) chambers on a chip in which to sequence DNA. The single DNA polymerase enzyme is anchored to the bottom of the ZMW chamber along with a single molecule of prepped DNA template. The ZMW chamber creates an illuminated observation space that is small enough to observe and detect only a single nucleotide at a time. As sequencing begins the nucleotides, each type labeled with a fluorescent dye is added to an array of ZMWs. When a complementary nucleotide is incorporated into the synthesizing strand at the bottom of the ZMW, that single nucleotide florescence is detected and the incorporated nucleotide is identified. Once a nucleotide is incorporated, the fluorescent dye is cleaved off and diffuses out of the ZMW, where its florescence is no longer observable. The next complementary base is then incorporated and detected and the process repeats throughout the length of the run.

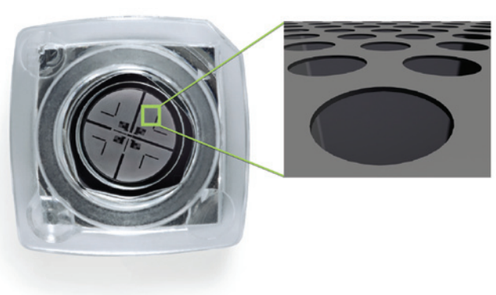
Sequencing is performed on a SMRT cell that contains an array of 150,000 ZMWs. A typical two-hour sequencing run of the SMRT cell generates on average 35,000 sequencing reads and each read has an average read length of 3,000 to 5,000 base-pairs, with the longest reads over 20,000 base-pairs. This approach currently yields up to 100 million bases per SMRT cell and the instrument has an option to load multiple SMRT cells in single run. The DOE JGI currently operates two of these PacBio sequencer instruments.

The SMRT sequencer relies on PacBio's RS DNA Template Preparation Kit to convert sample DNA into the proprietary SMRTbell™ library format for single-molecule real-time sequencing. The SMRTbell DNA template preparation method creates a unique, structurally linear, and topologically circular DNA morphology.

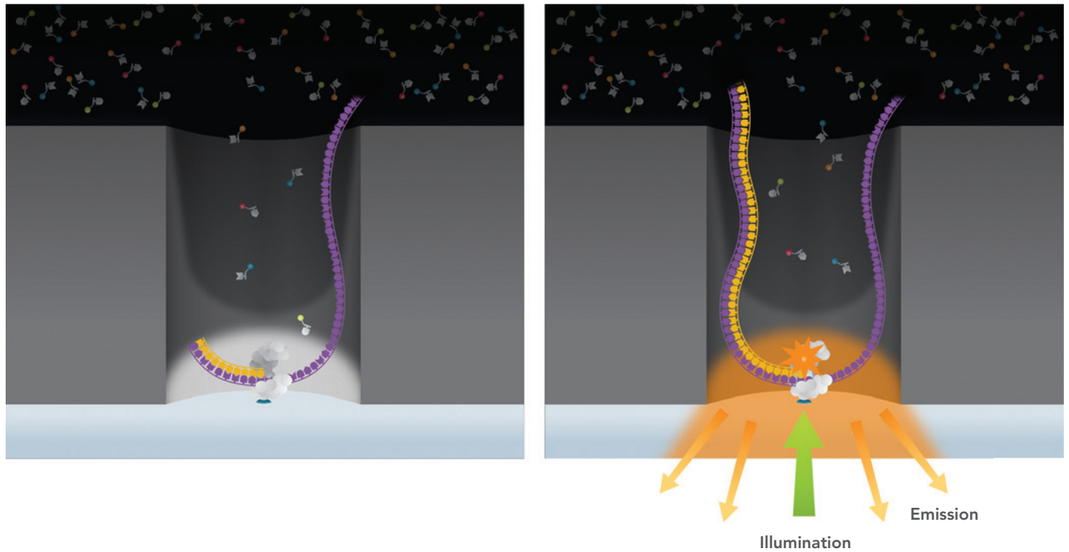
1. DNA sample prep is done away from the instrument and requires as little as 1 microgram (µg) of starting material for a typical prep. The starting DNA is sheared into double-stranded linear structures with sizes ranging from 200 base pairs to 10,000 base pairs, and then attached to the SMRTbell adapters, which produce a topologically closed circle, enabling consensus sequencing of the same template. The front of the machine contains two drawers for sample loading — one for DNA and reagents, the other for up to 16 SMRT cells.



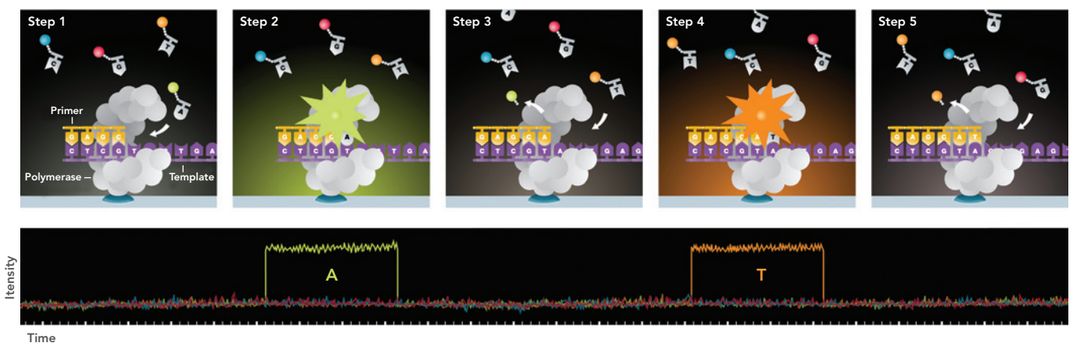

2. **SMRT Cell**
   Each SMRT cell is patterned with 150,000 zero-mode waveguides (ZMWs) measuring 100 nm across, and each ZMW contains a single DNA polymerase. The ZMW is the window through which DNA sequencing can be monitored in real time. The PacBio RS system continuously monitors ZMWs in sets of 75,000 at a time. Each SMRT cell can be run in minutes.



3. **SMRT Sequencing**
   When an active polymerase is immobilized at the bottom of each ZMW, nucleotides diffuse into the chamber. Each of the four nucleotides is tagged with fluorescent markers on the terminal phosphate, not the base. Since only the bottom 30 nm of the ZMW is illuminated, only those nucleotides near the bottom fluoresce. When the correct nucleotide is detected by the polymerase, it is incorporated into the growing DNA strand in a process that takes milliseconds.




4. **Base Calling**
   Four light-sensitive cameras collect the pulses emitted by fluorescent tags, allowing the observation of biological processes. Algorithms then translate the information captured by the optics system and convert the light pulses into either an A, C, G, or T base call. A consensus sequence can then be assembled by aligning the different fragments from each ZMW based on common sequences.



# Appendix C:
# FY 2013 CSP Projects

The 2013 Community Sequencing Program (CSP) call invited researchers to submit proposals that advance capabilities in fields such as plant-microbe interactions, microbes involved in carbon capture and greenhouse gas emission, and metagenomics — the characterization of complex collections of microbes from particular environmental niches.

| Proposer | Affiliation | Project Description |
|---|---|---|
| Bartley, Laura | University of Oklahoma | Switchgrass resequencing to enhance biofuel production |
| Bartley, Laura | University of Oklahoma | *Panicum virgatum* RNA sequencing to identify gene expression changes related to biofuel traits |
| Blanchard, Jeffrey | University of Massachusetts | Microbial community dynamics in a long-term soil warming chronosequence |
| Boeke, Jef | Johns Hopkins University | Synthesis of the largest yeast chromosome, chromosome IV, and the synthetic yeast genome Sc2.0 |
| Brune, Andreas | Max-Planck Institute, Marburg, Germany | Metagenomics and metatranscriptomics of the gut microbiota of higher termites |
| Constant, Philippe | INRS, Canada | Metagenomic and metatranscriptomic analysis of soil biogeochemical processes sustained by interspecific transfer of molecular hydrogen |
| Cottrell, Matthew | University of Delaware | Actively growing bacteria in coastal waters of the west Antarctic Peninsula identified by metatranscriptomic analysis |
| Denef, Vincent | University of Michigan | The bacterial component of the microbial loop in the Laurentian Great Lakes and their role in the carbon cycle. |
| Deutschbauer, Adam | LBNL | Functional Encyclopedia of Bacteria and Archaea (FEBA) |



| Proposer | Affiliation | Project Description |
|---|---|---|
| Francis, Christopher | Stanford University | Metagenomic and biogeochemical characterization of marine ammonia-oxidizing archaeal communities in a coastal upwelling system |
| Glass, N. Louise | UC Berkeley | The Fungal Nutritional ENCODE project |
| Greenshields, Dave | Novozymes | Genome sequencing of phosphate-solubilizing *Penicillium* species to understand fungal contributions to the phosphorus cycle |
| Hallam, Steven | University of British Columbia, Canada | Opening a single-cell genomic window on microbial ecotype selection in expanding marine oxygen minimum zones |
| Jansson, Janet | LBNL | Next Generation Ecosystem Experiment (NGEE) in the Arctic |
| Juenger, Tom | University of Texas at Austin | Resequencing diverse collections and mapping resources for *Panicum hallii* |
| Kema, G. H. J. | Wageningen University and Research Centre, Netherlands | Deciphering the interactome of Dothideomycete-bioenergy crops |
| Klenk, Hans-Peter | DSMZ, Germany | Genomic Encyclopedia of Archaeal and Bacterial Type Strains, Phase II: From individual species to whole genera |
| Martin, Francis | INRA, France | Mycorrhizal Genomics Initiative: Exploring the symbiotic transcriptomes |
| Nichols, Nancy | USDA-ARS | Genome sequence of the inhibitor-tolerant Ascomycete *Coniochaeta ligniaria* NRRL30616 |
| Nusslein, Klaus | University of Massachusetts | Profiling metatranscriptomic consequences of Amazon deforestation at different spatial scales |
| Picard, Kathryn | Duke University | Functional genomics of the saprotrophic-symbiotrophic fungus *Rhizidium phycophilum* and its algal partner *Bracteacoccus* sp.: Increasing algal biomass through symbiosis |
| Record, Eric | Aix-Marseille University, France | Survey of lignocellulolytic capabilities over the order *Polyporales* (Fungi, Basidiomycetes) |
| Sczyrba, Alex | Bielefeld University, Germany | Metagenome, metatranscriptome, and single-cell genome sequencing to uncover the microbiology and functional potential of biogas-producing microbial communities from production-scale biogas plants |
| Shapiro, Lucy | Stanford University | Defining the essential symbiosis genome of *Sinorhizobium meliloti* |
| Siegel, Justin | University of Washington | Enhancement of biofuel production and carbon fixation pathways through metagenomic enzyme design |
| Stacey, Gary | University of Missouri | DOE JGI Flagship Plant Gene Atlas Pilot |
| Stepanauskas, Ramunas | Bigelow Laboratory for Ocean Sciences | Enigmatic life underneath us: Genomic analysis of deep subsurface microorganisms |
| Teeling, Hanno | Max-Planck Institute, Bremen, Germany | COGITO (Coastal Microbe Genomic & Taxonomic Observatory) |

# Appendix D:

# Advisory and Review Committee Members

## The Scientific Advisory Committee (SAC)

The Scientific Advisory Committee is a board convened by the DOE JGI Director to provide a scientific and technical overview of the DOE JGI. Responsibilities of this board include providing technical input on large-scale production sequencing, new sequencing technologies, and related informatics; overview of the scientific programs at the DOE JGI; and overview of the Community Sequencing Program (CSP). A crucial job of the committee is to take the input from the CSP Proposal Study Panel on prioritization of CSP projects and, with the DOE Office of Biological and Environmental Research (BER) concurrence, set the final sequence allocation for this program.

**Members**

Bruce Birren (chair), *Broad Institute*

Jeff Dangl, *University of North Carolina*

David Dooling, *Monsanto*

Joe Ecker, *Salk Institute*

Glenn Kubiak, *Lawrence Berkeley National Laboratory*

Nancy Moran, *Yale University*

Julian Parkhill, *The Sanger Institute*

James Tiedje, *Michigan State University*

Alexandra Z. Worden, *Monterey Bay Aquarium Research Institute*



## CSP Proposal Study Panel (PSP)

**Members**

Nina Agabian, *University of California at San Francisco*

Chris Amemiya, *Benaroya Research Institute at Virginia Mason*

Gary L. Andersen, *Lawrence Berkeley National Laboratory*

Jo Ann Banks, *Purdue University*

John Battista, *Louisiana State University*

Fred Brockman, *Pacific Northwest National Laboratory*

Zac Cande, *University of California at Berkeley*

Patrick Chain, *Lawrence Livermore National Laboratory*

Jonathan C. Cohen, *UT Southwestern Medical Center*

Nigel Dunn-Coleman, *Genencor International*

Joe Ecker, The Salk Institute for Biological Studies

Katrina Edwards, *Woods Hole Oceanographic Institution*

Kelly Frazer, *Perlegen Sciences, Inc.*

Richard Harland, *University of California at Berkeley*

Derek Lovley, *University of Massachusetts*

David Mills, *University of California at Davis*

Alison Murray, *Desert Research Institute*

Arend Sidow, *Stanford University*

Nipam Patel, *University of California at Berkeley*

Karin Remington, *National Institute of General Medical Sciences*

John Taylor, *University of California at Berkeley*

Naomi Ward, *The Institute for Genomic Research*

Bart Weimer, *Utah State University*

**DOE JGI Ex-Officio Members**

James Bristow (PSP Chairman), *DOE Joint Genome Institute*

Daniel Rokhsar, *DOE Joint Genome Institute*

Eddy Rubin, *Director, DOE Joint Genome Institute*

## The Informatics Advisory Committee (IAC)

**Members**

Adam Arkin, *Division Director, Physical Biosciences Division, Lawrence Berkeley National Laboratory*

David Dooling, *Assistant Director, Genome Center, Washington University, St Louis*

Saul Kravitz, *Principal Systems Engineer, Center for Connected Government, MITRE*

Stan Letovsky, *Vice President and Chief Informatics Officer at SynapDx*

Jill Mesirov, *Associate Director and Chief Informatics Officer, Broad Institute (IAC Chair)*

Granger Sutton, *Senior Director of Informatics, J. Craig Venter Institute*

Kathy Yelick, *Associate Lab Director, Computing Sciences, Lawrence Berkeley National Laboratory*

## Plant Program User Advisory Committee

**Members**

Joe Ecker, *The Salk Institute for Biological Studies*

Jeff Dangl, *University of North Carolina*

Stephen Moose, *University of Illinois*

Sabeeha Merchant, *University of California at Los Angeles*

Gary Stacey, *University of Missouri*

Thomas Mitchell-Olds, *Duke University*

Eva Huala, *Carnegie Institute/TAIR*

Appendix D continued

## Fungal Program User Advisory Committee

**Members**

Scott Baker, *Pacific Northwestern National Laboratory*

Randy Berka, *Novozymes*

Ronal de Vries, *CBS (Netherlands)*

Audrey Gasch, *Great Lakes Bioenergy Research Center*

N. Louise Glass, *University of California at Berkeley*

Stephen Goodwin, *Purdue University*

David Hibbett, *Clark University*

Francis Martin, *INRA (France)*

Joseph Spatafora, *Oregon State University*

Adrian Tsang, *Concordia University (Canada)*

## Prokaryotic Super Program Advisory Committee Meeting

**Members**

Cameron Currie, *University of Wisconsin*

Ed DeLong, *MIT*

Jed Fuhrman, *University of Southern California*

George Garrity, *Michigan State University*

Steve Hallam, *University of British Columbia (Canada)*

Phil Hugenholtz, *University of Queensland (Australia)*

Bob Landick, *Great Lakes Bioenergy Research Center*

Folker Meyer, *Argonne National Laboratory*

Nancy Moran, *Yale University*

Mary Ann Moran, *University of Georgia*

Karen Nelson, *J. Craig Venter Institute*

Rich Roberts, *New England BioLabs*

Doug Rusch, *J. Craig Venter Institute*

Ramunas Stepanauskas, *Bigelow Laboratory for Ocean Sciences*

Niels van der Lelie, *RTI*




# Appendix E:

# 2012 Genomics of Energy & Environment Meeting

## Keynote Speakers

Science writer Carl Zimmer delivered the opening keynote, sharing his thoughts on covering the genomics beat for publications such as the *New York Times* and *Discover* magazine. Starting with an anecdote on the first time he realized the power of genomics, he discussed the challenges he and his colleagues face in describing why each genome sequence now routinely reported is significant.

Biochemist Steve Benner from the Foundation for Applied Molecular Evolution closed the meeting with a keynote acknowledging the differences between the scientist's perspective and the public's perspective of genomic studies. He urged the audience to take the "planetary biology" approach in considering the importance of genomics and how multiple genomes enable researchers to study both human and global histories and respond accordingly.

Appendix E continued

**Other Featured Speakers** (in order of appearance):

Jody Banks, *Purdue University*

Fred G. Gmitter, Jr., *Citrus Research and Education Center, University of Florida*

David Hibbett, *Clark University*

Robert Goldberg, *University of California at Los Angeles*

Loren Rieseberg, *University of British Columbia, (Canada)*

**Scientific Session: Forging the Vision for the DOE JGI**
Adam Arkin, *Lawrence Berkeley National Laboratory*
Dan Drell, *DOE Office of Biological and Environmental Research*
Eddy Rubin, *DOE Joint Genome Institute*
Richard Sayre, *New Mexico Consortium and Los Alamos National Laboratory*
Christina Smolke, *Stanford University*

Justin P. Gallivan, *Center for Fundamental and Applied Molecular Evolution, Emory University*

Eske Willerslev, *University of Copenhagen (Denmark)*

Adam Arkin, *Lawrence Berkeley National Laboratory*

Shawn Kaeppler, *University of Wisconsin-Madison*

Michael C. Schatz, *Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory*

Matthew B. Sullivan, *University of Arizona*

Ramunas Stepanauskas, *Bigelow Laboratory for Ocean Sciences*

Richard Sayre, *New Mexico Consortium and Los Alamos National Laboratory*

Stephen J. Giovannoni, *Oregon State University*

Wellington Muchero, *Oak Ridge National Laboratory*

Annika Mosier, *University of California at Berkeley*

Patrick Shih, *University of California at Berkeley*

Siobhan Brady, *University of California at Davis*

N. Louise Glass, *EBI, University of California at Berkeley*

Stan D. Wullschleger, *Environmental Sciences Division, Oak Ridge National Laboratory*

Christina Smolke, *Stanford University School of Medicine*

Len Pennacchio, *DOE Joint Genome Institute*

Learn more about the meeting talks at http://bit.ly/UM7agenda

Videos of the 2012 Meeting talks are available on the DOE JGI's SciVee channel at http://www.scivee.tv/node/47395



# Appendix F:
# 2011-12 Publications


1. Abt B et al. Complete genome sequence of the termite hindgut bacterium *Spirochaeta coccoides* type strain (SPN1T), reclassification in the genus *Sphaerochaeta* as *Sphaerochaeta coccoides* comb. nov. and emendations of the ... *Stand Genomic Sci*. 2012 May 25;6(2):194-209. Epub 2012 May 4. doi:10.4056/sigs.2796069

2. Anderson I et al. Genome sequence of the thermophilic sulfate-reducing ocean bacterium *Thermodesulfatator indicus* type strain (CIR29812T). *Stand Genomic Sci*. 2012 May 25;6(2):155-64. Epub 2012 May 4. doi:10.4056/sigs.2665915

3. Anderson I et al. Genome sequence of the homoacetogenic bacterium *Holophaga foetida* type strain (TMBS4$^T$). *Stand Genomic Sci*. 2012 May 25;6(2):174-84. Epub 2012 May 4. doi: 10.4056/sigs.2746047

4. Anderson I et al. Genomics of aerobic cellulose utilization systems in actinobacteria. *PLoS One*. 2012;7(6):e39331. Epub 2012 Jun 18.

5. Anderson I et al. Complete genome sequence of the moderately thermophilic mineral-sulfide-oxidizing firmicute *Sulfobacillus acidophilus* type strain (NALT). *Stand Genomic Sci*. 2012 July 20;6(3). doi:10.4056/sigs.2736042

6. Anderson I et al. Complete genome sequence of *Ferroglobus placidus* AEDII12DO. *Stand Genomic Sci*. 2011 Oct 15;5(1):50-60.

7. Anderson IJ et al. Complete genome sequence of *Halopiger xanaduensis* type strain (SH6T). *Stand Genomic Sci*. 2012 Mar 5;6:1. doi:10.4056/sigs.2505605

8. Aylward FO et al. Metagenomic and metaproteomic insights into bacterial communities in leaf-cutter ant fungus gardens. *ISME J*. 2012 Sep;6(9):1688-701. doi: 10.1038/ismej.2012.10. Epub 2012 Mar 1.

9. Baelum J et al. Deep-sea bacteria enriched by oil and dispersant from the Deepwater Horizon spill. *Environ Microbiol*. 2012 Sep;14(9):2405-16. doi: 10.1111/j.1462-2920.2012.02780.x. Epub 2012 May 23.

10. Baker SE et al. Phylogenomic analysis of polyketide synthase-encoding genes in *Trichoderma*. Microbiology. 2012 Jan;158(Pt 1):147-54. Epub 2011 Nov 17.

11. Bayer T et al. *Symbiodinium* transcriptomes: Genome insights into the dinoflagellate symbionts of reef-building corals. *PLoS One*. 2012; 7(4): e35269. doi:10.1371/journal.pone.0035269

12. Beckham GT et al. Harnessing glycosylation to improve cellulase activity. *Curr Opin Biotechnol*. 2012 Jun;23(3): 338-45. Epub 2011 Dec 18.

13. Beckloff N et al. Bacterial genome annotation. *Methods Mol Biol*. 2012;881:471-503.

14. Bennetzen JL et al. Reference genome sequence of the model plant *Setaria*. *Nat Biotechnol*. 2012 May 13;30(6):555-61. doi: 10.1038/nbt.2196

15. Berg Miller ME et al. Phage-bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. *Environ Microbiol*. 2012 Jan;14(1):207-27. doi:10.1111/j.1462-2920.2011.02593.x. Epub 2011 Oct 17.

Appendix F continued


16. Berka RM et al. Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. *Nat Biotechnol*. 2011 Oct 2. doi: 10.1038/nbt.1976.

17. Bini E et al. Complete genome sequence of *Desulfurispirillum indicum* strain S5T. *Stand Genomic Sci*. 2011 5:3 doi:10.4056/sigs.2425302 Epub 2011 Dec 22.

18. Blanc G et al. The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biol*. 2012 May 25;13(5):R39. doi: 10.1186/gb-2012-13-5-r39

19. Boden R et al. Complete genome sequence of the aerobic marine methanotroph *Methylomonas methanica* MC09. *J Bacteriol*. 2011 Dec;193(24):7001-2.

20. Brisson VL et al. Metagenomic analysis of a stable trichloroethene-degrading microbial community. *ISME J*. 2012 Mar 1. doi: 10.1038/ismej.2012.15. [Epub ahead of print]

21. Brumm PJ et al. Complete genome sequence of *Paenibacillus* strain Y4.12MC10, a novel *Paenibacillus lautus* strain isolated from Obsidian Hot Spring in Yellowstone National Park. *Stand Genomic Sci*. 2012 July 27;6(3). doi:10.4056/sigs.2605792

22. Casler MD et al. The switchgrass genome: Tools and strategies. *Plant Gen*. 2011 Nov; 4(3): 273–282. doi:10.3835/plantgenome2011.10.0026

23. Chan CX et al. Analysis of *Porphyra* membrane transporters demonstrates gene transfer among photosynthetic eukaryotes and numerous sodium-coupled transport systems. *Plant Physiol*. 2012 Apr;158(4):2001-12.

24. Chan CX et al. *Porphyra* (Bangiophyceae) transcriptomes provide insights into red algal development and metabolism. *J Phycol*. 2012. doi: 10.1111/j.1529-8817.2012.01229.x

25. Chang YJ et al. Non-contiguous finished genome sequence and contextual data of the filamentous soil bacterium *Ktedonobacter racemifer* type strain (SOSP1-21). *Stand Genomic Sci*. 2011 Oct 15;5(1): 97-111. Epub 2011 Oct 1.

26. Chertkov O et al. Complete genome sequence of *Hirschia baltica* type strain (IFAM 1418T). *Stand Genomic Sci*. 2011 5:3 doi:10.4056/sigs.2205004 Epub 2011 Dec 23.

27. Chow V et al. Complete genome sequence of *Paenibacillus* sp. strain JDR-2. *Stand Genomic Sci*. 2012 Mar 5;6:1. doi:10.4056/sigs.2374349

28. Copeland A et al. Complete genome sequence of the aerobic, heterotroph *Marinithermus* hydrothermalis type strain (T1T) from a deep-sea hydrothermal vent chimney. *Stand Genomic Sci*. 2012 Mar 5;6:1. doi:10.4056/sigs.2435521

29. Copeland A et al. Complete genome sequence of the halophilic and highly halotolerant *Chromohalobacter salexigens* type strain (1H11T). *Stand Genomic Sci*. 2011 5:3 doi:10.4056/sigs.2285059 Epub 2011 Dec 30.

30. Copeland A et al. Complete genome sequence of the aquatic bacterium *Runella slithyformis* type strain (LSU 4T). *Stand Genomic Sci*. 2012 May 25;6(2):145-54. Epub 2012 May 4. doi:10.4056/sigs.2665915

31. Copeland A et al. Complete genome sequence of the orange-red pigmented, radioresistant *Deinococcus proteolyticus* type strain (MRP$^{T}$). *Stand Genomic Sci*. 2012 May 25;6(2):240-50. Epub 2012 Apr 27. doi:10.4056/sigs.2756060

32. Cui Y et al. Fine mapping of the Bsr1 barley stripe mosaic virus resistance gene in the model grass *Brachypodium distachyon*. *PLoS One*. 2012;7(6):e38333.

33. Davis JR et al. Genome sequence of *Amycolatopsis* sp. strain ATCC 39116, a plant biomass-degrading Actinomycete. *J Bacteriol*. 2012 May;194(9):2396-7.

34. Dichosa AE et al. Artificial polyploidy improves bacterial single cell genome recovery. *PLoS One*. 2012;7(5):e37387. Epub 2012 May 22.

35. Dilly GF et al. Exploring the limit of metazoan thermal tolerance via comparative proteomics: thermally induced changes in protein abundance by two hydrothermal vent polychaetes. *Proc Biol Sci*. 2012 Aug 22;279(1741):3347-56. Epub 2012 May 2.

36. Earl D et al. Assemblathon 1: A competitive assessment of de novo short read assembly methods. *Genome Res*. 2011 Dec;21(12):2224-41. Epub 2011 Sep 16.

37. Fernandez-Fueyo E et al. Comparative genomics of *Ceriporiopsis subvermispora* and *Phanerochaete chrysosporium* provide insight into selective ligninolysis. *Proc Natl Acad Sci U S A*. 2012 Apr 3;109(14):5458-63. Epub 2012 Mar 20.

38. Floudas D et al. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science*. 2012 Jun 29;336(6089):1715-9.

39. Fostona M et al. Chemical, ultrastructural and supramolecular analysis of tension wood in *Populus tremula x alba* as a model substrate for reduced recalcitrance. *Energy Environ*. Sci. 2011. Advance Article. doi:10.1039/C1EE02073K

40. García-Amado MA et al. Bacterial diversity in the cecum of the world's largest living rodent (*Hydrochoerus hydrochaeris*). *Microb Ecol*. 2012 May;63(4):719-25. Epub 2011 Nov 15.

41. Garcia SL et al. Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. *ISME J*. 2012 Jul 19. doi: 10.1038/ismej.2012.86. [Epub ahead of print]

42. Giovannelli D et al. Complete genome sequence of *Thermovibrio ammonificans* HB-1T, a thermophilic, chemolithoautotrophic bacterium isolated from a deep-sea hydrothermal vent. *Stand Genomic Sci*. Epub 2012 Sep 26;7(1). doi:10.4056/sigs.2856770

43. Godoy-Vitorino F et al. Differences in crop bacterial community structure between hoatzins from different geographical locations. *Res Microbiol*. 2012 Apr;163(3):211-20. Epub 2012 Jan 21.

44. Göker M et al. Complete genome sequence of the thermophilic sulfur-reducer *Desulfurobacterium thermolithotrophum* type strain (BSAT) from a deep-sea hydrothermal vent. *Stand Genomic Sci*. 2011 5:3 doi:10.4056/sigs.2465574 Epub 2011 Dec 30.

45. Göker M et al. Genome sequence of the moderately thermophilic, amino-acid-degrading and sulfur-reducing bacterium *Thermovirga lienii* type strain (Cas60314$^{T}$). *Stand Genomic Sci* 2012 May 25;6(2):230-9. Epub 2012 May 4.doi:10.4056/sigs.2726028

46. Goodstein DM et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012 Jan;40 (Database issue):D1178-86. Epub 2011 Nov 22.

47. Graham DE et al. Microbes in thawing permafrost: the unknown variable in the climate change equation. *ISME J*. 2012 Apr;6(4):709-12. doi: 10.1038/ismej.2011.163

48. Grote J et al. Draft genome sequence of strain HIMB100, a cultured representative of the SAR116 clade of marine *Alphaproteobacteria*. *Stand Genomic Sci*. 2011;5(3). doi:10.4056/sigs.1854551 Epub 2011 Dec 23.

49. Gori F et al. The metagenomic basis of anammox metabolism in Candidatus *'Brocadia fulgida'*. *Biochem Soc Trans*. 2011 Dec 1;39(6):1799-804.

50. Grbić M et al. The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature*. 2011 Nov 23;479(7374):487-92. doi: 10.1038/nature10640

51. Grigoriev IV et al. The Genome Portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Res*. 2012 Jan;40(1):D26-32. Epub 2011 Nov 22.

52. Han CS et al. Complete genome sequence of the sulfur compounds oxidizing chemolithoautotroph *Sulfuricurvum kujiense* type strain (YK-1T). *Stand Genomic Sci*. 2012 Mar 5;6:1. doi:10.4056/sigs.2456004

53. Han CS et al. Complete genome sequence of the facultatively chemolithoautotrophic and methylotrophic alphaproteobacterium *Starkeya novella* type strain (ATCC 8093T). *Stand Genomic Sci*. Epub 2012 Sep 26;7(1). doi:10.4056/sigs.3006378

54. Harhangi HR et al. Hydrazine synthase, a unique phylomarker to study the presence and biodiversity of anammox bacteria. *Appl Environ Microbiol*. 2011Dec 2. [Epub ahead of print]

55. Hollister EB et al. Mesophilic and thermophilic conditions select for unique but highly parallel microbial communities to perform carboxylate platform biomass conversion. *PLoS One*. 2012;7(6):e39689. Epub 2012 Jun 22.

56. Human Microbiome Project Consortium. A framework for human microbiome research. *Nature*. 2012 Jun 13;486(7402):215-21. doi: 10.1038/nature11209

Appendix F continued


57. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012 Jun 13;486(7402):207-14. doi: 10.1038/nature11234

58. Huntemann M et al. Complete genome sequence of the facultatively anaerobic, appendaged bacterium *Muricauda ruestringensis* type strain (B1$^T$). *Stand Genomic Sci*. 2012 May 25;6(2):185-93. Epub 2012 May 4. doi:10.4056/sigs.2786069

59. Isanapong J et al. High-quality draft genome sequence of the Opitutaceae bacterium strain TAV1, a symbiont of the wood-feeding termite *Reticulitermes flavipes*. *J Bacteriol*. 2012 May;194(10):2744-5.

60. Izquierdo JA et al. Complete genome sequence of *Clostridium clariflavum* DSM 19732. *Stand Genomic Sci*. 2012 Mar 5;6:1. doi:10.4056/sigs.2456004

61. Jiang K et al. Complete genome sequence of *Thauera aminoaromatica* strain MZ1T. *Stand Genomic Sci*. 2012 July 20;6(3). doi:10.4056/sigs.2696029

62. Joshi HJ et al. 1001 Proteomes: a functional proteomics portal for the analysis of *Arabidopsis thaliana* accessions. *Bioinformatics*. 2012 May 15;28(10):1303-6.

63. Jung S et al. Whole genome comparisons of *Fragaria, Prunus* and *Malus* reveal different modes of evolution between Rosaceous subfamilies. *BMC Genomics*. 2012 Apr 4;13(1):129.

64. Kallimanis A et al. Complete genome sequence of Mycobacterium sp. strain (Spyr1) and reclassification to *Mycobacterium gilvum* Spyr1. *Stand Genomic Sci*. 2011 Oct 15;5(1):144-53. Epub 2011 Oct 1.

65. Khudyakov JI et al. Global transcriptome response to ionic liquid by a tropical rain forest soil bacterium, *Enterobacter lignolyticus*. *Proc Natl Acad Sci* U S A. 2012 Aug 7;109(32):E2173-82.

66. Kimelman A et al. A vast collection of microbial genes that are toxic to bacteria. *Genome Res*. 2012 Apr;22(4):802-9. Epub 2012 Feb 1.

67. Kinney JN et al. Elucidating the essential role of the conserved carboxysomal protein CcmN reveals a common feature of bacterial microcompartment assembly. *J Biol Chem*. 2012 May 18;287(21):17729-36. Epub 2012 Mar 29.

68. Kiss H et al. Complete genome sequence of the filamentous gliding predatory bacterium *Herpetosiphon aurantiacus* type strain (114-95T). *Stand Genomic Sci*. 2011 5:3 doi:10.4056/sigs.2194987 Epub 2011 Dec 23.

69. Klassen JL et al. Draft genome sequence of Streptomyces sp. strain Wigar10, isolated from a surface-sterilized garlic bulb. *J Bacteriol*. 2011 Dec;193(24):6999-7000.

70. Klenk HP et al. Genome sequence of the ocean sediment bacterium *Saccharomonospora marina* type strain (XMU15$^T$). *Stand Genomic Sci*. 2012 May 25;6(2):265-75. Epub 2012 May 15. doi:10.4056/sigs.2655905

71. Klenk HP et al. Genome sequence of the soil bacterium *Saccharomonospora azurea* type strain (NA-128($^T$)). *Stand Genomic Sci*. 2012 May 25;6(2):220-9. Epub 2012 May 5.

72. Knight R et al. Unlocking the potential of metagenomics through replicated experimental design. *Nat Biotechnol*. 2012 Jun 7;30(6):513-20. doi: 10.1038/nbt.2235.

73. Koren S et al. Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nat Biotechnol*. 2012 Jul 1. doi: 10.1038/nbt.2280.

74. Kostka JE et al. Genome sequences for six *Rhodanobacter strains*, isolated from soils and the terrestrial subsurface, with variable denitrification capabilities. *J Bacteriol*. 2012 Aug;194(16):4461-2. doi: 10.1128/JB.00871-12

75. Lamour K et al. Genome sequencing and mapping reveal loss of heterozygosity as a mechanism for rapid adaptation in the vegetable pathogen *Phytophthora capsici*. *Mol Plant Microbe Interact*. 2012 Oct;25(10):1350-60. doi: 10.1094/MPMI-02-12-0028-R. Epub 2012 Jun 20.

76. Lapidus A et al. Genome sequence of the filamentous, gliding *Thiothrix nivea* neotype strain (JP2T). *Stand Genomic Sci*. 2011 5:3 doi:10.4056/sigs.2344929 Epub 2011 Dec 30.

77. Leung K et al. A programmable droplet-based microfluidic device applied to multiparameter analysis of single microbes and microbial communities. *Proc Natl Acad Sci* U S A. 2012 May 15;109(20):7665-70. Epub 2012 Apr 30.

78. Liolios K et al. The Metadata Coverage Index (MCI): A standardized metric for quantifying database metadata richness. *Stand Genomic Sci*. 2012 July 20;6(3). doi:10.4056/sigs.2675953

79. Lucas-Elío P et al. Complete genome sequence of the melanogenic marine bacterium *Marinomonas mediterranea* type strain (MMB-1T). *Stand Genomic Sci*. 2012 Mar 5;6:1. doi:10.4056/sigs.2545743

80. Lucas-Elío P et al. Complete genome sequence of *Marinomonas posidonica* type strain (IVIA-Po-181T). *Stand Genomic Sci*. Epub 2012 Sep 27;7(1). doi:10.4056/sigs.2976373

81. Lundberg DS et al. Defining the core *Arabidopsis thaliana* root microbiome. *Nature*. 2012 Aug 2;488:86-90.

82. Luo C et al. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS One*. 2012;7(2):e30087. Epub 2012 Feb 10.

83. Luo C et al. Individual genome assembly from complex community short-read metagenomic datasets. *ISME J*. 2012 Apr;6(4):898-901. doi: 10.1038/ismej.2011.147. Epub 2011 Oct 27.

84. Mackelprang R et al. Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature*. 2011 Nov 6;480(7377):368-71. doi: 10.1038/nature10576.

85. Markowitz V et al. IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res*. 2012 Jan;40(1):D115-22.

86. Markowitz VM. IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Res*. 2012 Jan;40(1):D123-9. Epub 2011 Nov 15.

87. Markowitz VM et al. IMG/M-HMP: A metagenome comparative analysis system for the Human Microbiome Project. *PLoS One*. 2012;7(7):e40151. Epub 2012 Jul 5.

88. Martinez RJ et al. Complete genome sequence of *Rahnella* sp. Strain Y9602, a Gammaproteobacterium isolate from metal- and radionuclide-contaminated soil. *J Bacteriol*. 2012 Apr;194(8):2113-4.

89. Martinez RJ et al. Complete genome sequence of *Rahnella aquatilis* CIP 78.65. *J Bacteriol*. 2012 Jun;194(11):3020-1.

90. Martinez-Garcia M et al. Capturing single cell genomes of active polysaccharide degraders: an unexpected contribution of *Verrucomicrobia*. *PLoS One*. 2012;7(4):e35314. Epub 2012 Apr 20.

91. Marx CJ et al. Complete genome sequences of six strains of the genus methylobacterium. *J Bacteriol*. 2012 Sep;194(17):4746-8.

92. Mason OU et al. Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *ISME J*. 2012 Sep;6(9):1715-27. doi: 10.1038/ismej.2012.59. Epub 2012 Jun 21.

93. Mavromatis K et al. Permanent draft genome sequence of the gliding predator *Saprospira grandis* strain Sa g1 (= HR1). *Stand Genomic Sci*. May 25;6(2):210-9. Epub 2012 May 4.doi:10.4056/sigs.2816096

94. Meincke L et al. Complete genome sequence of *Polynucleobacter necessarius* subsp. asymbioticus type strain (QLW-P1DMWA-1T). *Stand Genomic Sci*. 2012 Mar 5;6:1. doi:10.4056/sigs.2395367

95. Muyzer G et al. Complete genome sequence of *Thioalkalivibrio* sp. K90mix. *Stand Genomic Sci*. 2011 5:3 doi:10.4056/sigs.2315092 Epub 2011 Dec 23

96. Neupane S et al. Complete genome sequence of the rapeseed plant-growth promoting *Serratia plymuthica* strain AS9. *Stand Genomic Sci*. 2012 Mar 5;6:1. doi:10.4056/sigs.2595762

97. Neupane S et al. Complete genome sequence of *Serratia plymuthica* strain AS12. *Stand Genomic Sci*. 2012 May 25;6(2):165-73. Epub 2012 May 1. doi:10.4056/sigs.2705996

98. Neupane S et al. Complete genome sequence of the plant-associated *Serratia plymuthica* strain AS13. *Stand Genomic Sci*. Epub 2012 Sep 26;7(1). doi:10.4056/sigs.2966299

Appendix F continued


99. Novichkov PS et al. RegPrecise web services interface: programmatic access to the transcriptional regulatory interactions in bacteria reconstructed by comparative genomics. *Nucleic Acids Res*. 2012 Jul;40(Web Server issue):W604-8. Epub 2012 Jun 14.

100. Olson A et al. Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen. *New Phytol*. 2012 Jun;194(4): 1001-13. doi: 10.1111/j.1469-8137.2012.04128.x. Epub 2012 Mar 28.

101. Padamsee M et al. The genome of the xerotolerant mold *Wallemia sebi* reveals adaptations to osmotic stress and suggests cryptic sexual reproduction. *Fungal Genet Biol*. 2012 Mar;49(3):217-26. Epub 2012 Feb 4.

102. Pagani I et al. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res*. 2012 Jan;40(1):D571-9. Epub 2011 Dec 1.

103. Pati A et al. ClaMS: A classifier for metagenomic sequences. *Stand Genomic Sci*. 2011 Nov 30;5(2): 248-53. Epub 2011 Nov 28.

104. Peng Z et al. Generation of long insert pairs using a Cre-LoxP inverse PCR approach. *PLoS One*. 2012;7(1):e29437. Epub 2012 Jan 9.

105. Persson T et al. Genome sequence of "Candidatus *Frankia datiscae*" Dg1, the uncultured microsymbiont from nitrogen-fixing root nodules of the dicot *Datisca glomerata*. *J Bacteriol*. 2011 Dec;193(24):7017-8.

106. Piškur J et al. The genome of wine yeast *Dekkera bruxellensis* provides a tool to explore its food-related properties. *Int J Food Microbiol*. 2012 Jul 2;157(2):202-9.

107. Plett JM et al. Phylogenetic, genomic organization and expression analysis of hydrophobin genes in the ectomycorrhizal basidiomycete *Laccaria bicolor*. *Fungal Genet Biol*. 2012 Mar;49(3):199-209. Epub 2012 Jan 28.

108. Prochnik S et al. The cassava genome: Current progress, future directions. *Trop Plant Biol*. 2012 Mar;5(1):88-94. Epub 2012 Jan 5.

109. Queiroux C. A comparative genomics screen identifies a *Sinorhizobium meliloti* 1021 sodM-like gene strongly expressed within host plant nodules. *BMC Microbiol*. 2012 May 15;12:74. doi: 10.1186/1471-2180-12-74.

110. Rawat SR et al. Complete genome sequence of *Terriglobus saanensis* strain SP1PR4T, an Acidobacteria from tundra soil. *Stand Genomic Sci*. Epub 2012 Sep 26;7(1). doi:10.4056/sigs.3036810

111. Rhee MS et al. Complete genome sequence of a thermotolerant sporogenic lactic acid bacterium, *Bacillus coagulans* strain 36D1. *Stand Genomic Sci*. 2011 5:3 doi:10.4056/sigs.2365342 Epub 2011 Dec 22.

112. Rineau F et al. The ectomycorrhizal fungus *Paxillus involutus* converts organic matter in plant litter using a trimmed brown-rot mechanism involving Fenton chemistry. *Environ Microbiol*. 2012 Jun;14(6):1477-1487. doi: 10.1111/j.1462-2920.2012.02736.x. Epub 2012 Mar 30.

113. Roberts EW et al. Isolation and characterization of the *Prochlorococcus* carboxysome reveals the presence of the novel shell protein CsoS1D. *J Bacteriol*. 2012 Feb;194(4):787-95. Epub 2011 Dec 9.

114. Schleheck D et al. Complete genome sequence of *Parvibaculum lavamentivorans* type strain (DS-1T). *Stand Genomic Sci*. 2011 5:3 doi:10.4056/sigs. 2215005 Epub 2011 Dec 22.

115. Scholz MB et al. Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Curr Opin Biotechnol*. 2012 Feb;23(1):9-15. Epub 2011 Dec 9.

116. Schuster A et al. A versatile toolkit for high throughput functional genomics with *Trichoderma reesei*. *Biotechnol Biofuels*. 2012 Jan 2;5(1):1. [Epub ahead of print]

117. Sessitsch A et al. Functional characteristics of an endophyte community colonizing rice roots as revealed by metagenomic analysis. *Mol Plant Microbe Interact*. 2012 Jan;25(1):28-36. Epub 2011 Oct 4.

118. Sharma MK et al. A genome-wide survey of switchgrass genome structure and organization. *PLoS One*. 2012 Apr 12; 7(4): e33892. doi:10.1371/journal.pone.0033892

119. Siddaramappa S et al. A comparative genomics perspective on the genetic content of the alkaliphilic haloarchaeon *Natrialba magadii* ATCC 43099T. *BMC Genomics*. 2012 May 4;13(1):165. doi: 10.1186/1471-2164-13-165.

120. Siddaramappa S et al. Complete genome sequence of *Dehalogenimonas lykanthroporepellens* type strain (BL-DC-9T) and comparison to "*Dehalococcoides*" strains. *Stand Genomic Sci*. 2012 May 25;6(2):251-64. Epub 2012 May 15. doi:10.4056/sigs.2806097

121. Siddaramappa S et al. Horizontal gene transfer in *Histophilus somni* and its role in the evolution of pathogenic strain 2336, as determined by comparative genomic analyses. *BMC Genomics*. 2011 Nov 23;12(1):570. [Epub ahead of print]

122. Siebers B et al. The complete genome sequence of *Thermoproteus tenax*: A physiologically versatile member of the Crenarchaeota. *PLoS One*. 2011;6(10):e24222. Epub 2011 Oct 7.

123. Slavov GT et al. Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. *New Phytol*. 2012 Nov;196(3):713-725. doi: 10.1111/j.1469-8137.2012.04258.x. Epub 2012 Aug 3.

124. Srivastava A et al. A zinc finger family transcription factor AbVf19 is required for the induction of a gene subset important for virulence in *Alternaria brassicicola*. *Mol Plant Microbe Interact*. 2012 Apr;25(4):443-52. Epub 2011 Dec 20.

125. Suen G et al. Complete genome of the cellulolytic ruminal bacterium *Ruminococcus albus 7*. *J Bacteriol*. 2011 Oct;193(19):5574-5.

126. Suzuki H et al. Comparative genomics of the white-rot fungi, *Phanerochaete carnosa* and *P. chrysosporium*, to elucidate the genetic basis of the distinct wood types they colonize. *BMC Genomics*. 2012 Sep 2;13:444. doi: 10.1186/1471-2164-13-444.

127. Svenning MM et al. Genome sequence of the Arctic methanotroph *Methylobacter tundripaludum* SV96. *J Bacteriol*. 2011 Nov;193(22):6418-9. Epub 2011 Jul 1.

128. Swaminathan K et al. A framework genetic map for *Miscanthus sinensis* from RNAseq-based markers shows recent tetraploidy. *BMC Genomics*. 2012 Apr 24;13(1):142.

129. Swingley WD et al. Coordinating environmental genomics and geochemistry reveals metabolic transitions in a hot spring ecosystem. *PLoS One*. 2012;7(6):e38108. Epub 2012 Jun 4.

130. Swithers KS et al. Genome sequence of *Kosmotoga olearia* strain TBF 19.5.1, a thermophilic bacterium with a wide growth temperature range, isolated from the Troll B oil platform in the North Sea. *J Bacteriol*. 2011 Oct;193(19):5566-7.

131. Swithers KS et al. Genome sequence of *Thermotoga* sp. strain RQ2, a hyperthermophilic bacterium isolated from a geothermally heated region of the seafloor near Ribeira Quente, the Azores. *J Bacteriol*. 2011 Oct;193(20):5869-70.

132. Tisserant E et al. The transcriptome of the arbuscular mycorrhizal fungus *Glomus intraradices* (DAOM 197198) reveals functional tradeoffs in an obligate symbiont. *New Phytol*. 2012 Feb;193(3):755-69. doi: 10.1111/j.1469-8137.2011.03948.x. Epub 2011 Nov 16.

133. Vandera E et al. Heterologous expression and characterization of two 1-hydroxy-2-naphthoic acid dioxygenases from *Arthrobacter phenanthrenivorans*. *Appl Environ Microbiol*. 2012 Feb;78(3):621-7. Epub 2011 Nov 18.

134. van der Lelie D. The metagenome of an anaerobic microbial community decomposing poplar wood chips. *PLoS One*. 2012;7(5):e36740. Epub 2012 May 21.

135. van de Vossenberg J et al. The metagenome of the marine anammox bacterium 'Candidatus *Scalindua profunda*' illustrates the versatility of this globally important nitrogen cycle bacterium. *Environ Microbiol*. 2012 May 9. doi: 10.1111/j.1462-2920.2012.02774.x. [Epub ahead of print]

136. Wang L et al. Mapping N-Linked glycosylation sites in the secretome and whole cells of *Aspergillus niger* using hydrazide chemistry and mass spectrometry. *J Proteome Res*. 2012 Jan 1;11(1):143-56. Epub 2011 Dec 2.

Appendix F continued


137. Weissgerber T et al. Complete genome sequence of *Allochromatium vinosum* DSM 180T. *Stand Genomic Sci.* 2011 5:3 doi:10.4056/sigs.2335270 Epub 2011 Dec 22.

138. Wilson A et al. The essential role of the N-terminal domain of the orange carotenoid protein in cyanobacterial photoprotection: importance of a positive charge for phycobilisome binding. *Plant Cell.* 2012 May;24(5):1972-83. Epub 2012 May 25.

139. Worden AZ et al. Global distribution of a wild alga revealed by targeted metagenomics. *Curr Biol.* 2012 September 11;22(17):R675-7. doi:10.1016/j.cub.2012.07.054

140. Woyke T et al. Decontamination of MDA reagents for single cell whole genome amplification. *PLoS One.* 2011;6(10):e26161. Epub 2011 Oct 20.

141. Zeytun A et al. Complete genome sequence of *Francisella philomiragia* ATCC 25017. *J Bacteriol.* 2012 Jun;194(12):3266.

142. Zhang X. Improving genome assemblies by sequencing PCR products with PacBio. *Biotechniques.* 2012 Jul;53(1):61-2.

143. Zhaxybayeva O et al. Genome sequence of the mesophilic *Thermotogales* bacterium *Mesotoga prima* MesG1.Ag.4.2 reveals the largest *Thermotogales* genome to date. *Genome Biol Evol.* 2012 Jan;4(8):700-8. doi: 10.1093/gbe/evs059. Epub 2012 Jul 12.


U.S. DEPARTMENT OF ENERGY
Office of Science

BERKELEY LAB

## Comments?


David Gilbert
DOE JGI Public Affairs Manager
degilbert@lbl.gov


## Follow Us


The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.


All photography, unless otherwise noted, by Roy Kaltschmidt, Berkeley Lab.

CSO25609


JGI
DOE JOINT GENOME INSTITUTE