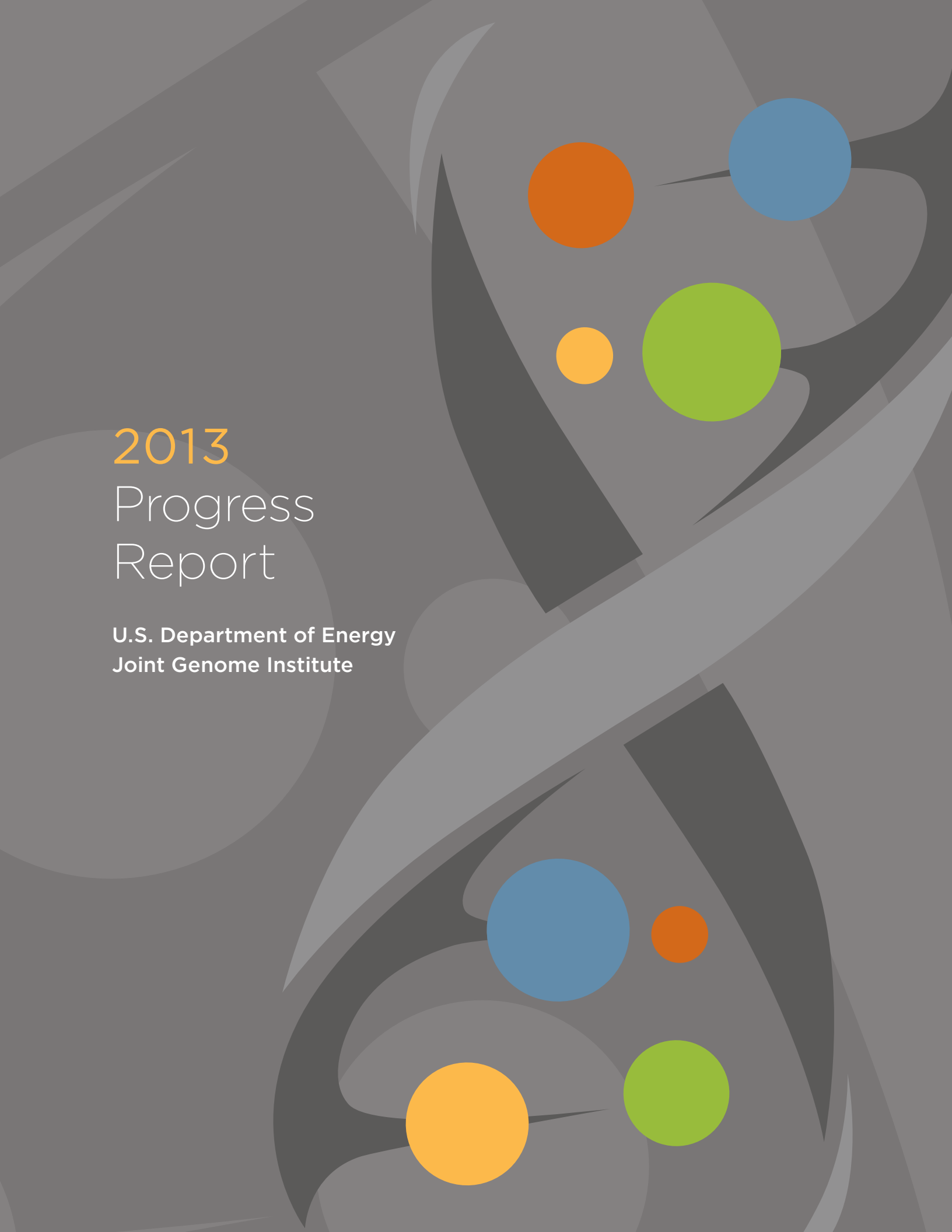


2013 Progress Report

U.S. Department of Energy
Joint Genome Institute

Inspired by the double-helix sculpture in the courtyard of the DOE Joint Genome Institute's Walnut Creek, California, headquarters, we have "refreshed" our logo. This new look reflects the DOE JGI's expanded capabilities as a next-generation genomic science center.

Table of Contents

DOE JGI
Mission

Francis Martin

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**

Implementing Our 10-Year Strategic Vision

In October 2012, we introduced a 10-Year Strategic Vision [http://bit.ly/JGI-Vision] for the Institute. A central focus of this Strategic Vision is to bridge the gap between sequenced genomes and an understanding of biological functions at the organism and ecosystem level. This involves the continued massive-scale generation of sequence data, complemented by orthogonal new capabilities to functionally annotate these large sequence data sets. Our Strategic Vision lays out a path to guide our decisions and ensure that the evolving set of experimental and computational capabilities available to DOE JGI users will continue to enable groundbreaking science.

An important early step toward implementation of our plan was the launch of the Emerging Technologies Opportunity Program (ETOP). The ETOP establishes partnerships between the DOE JGI and an external set of researchers who are developing cutting-edge technologies that will enhance our portfolio of capabilities. A particular emphasis is in the area of high-throughput technologies to link sequence data to functional information. Our core philosophy is that continued internal and external development of new technical and analytical assets will be essential in order to remain state of the art and maximize our users' scientific achievements. Through the first round of ETOP proposals, six exciting new partnerships are taking shape that include teams from Stanford University; the University of Washington; the Massachusetts Institute of Technology and the University of Vienna; the Arizona Genomics Institute; the University of California at Berkeley and Oak Ridge National Laboratory; and Pacific Northwest National Laboratory (PNNL). More details about these exciting projects can be found on page 51.

In 2013, we made a subtle but important change to the name of our primary user program — the Community Sequencing Program, which is now called the Community Science Program. This distinction more accurately reflects the work done by and at the DOE JGI. In addition to high-throughput DNA sequencing and analysis of DNA sequence, we offer a diverse panel of complementary capabilities for users to carry out DOE mission-relevant science. These capabilities include, for example, single-cell genomics, DNA synthesis for building genes and pathways, and transposon-mediated mutagenesis coupled with high-throughput sequencing to assist in the annotation of microbes and other genomes. These capabilities align with our Strategic Vision — to transition from a production sequencing facility to a full-fledged genomic-analysis resource, allowing researchers to convert sequence data into biological insights.

In keeping with this focus, we joined forces with the Environmental Molecular Sciences Laboratory (EMSL) at PNNL, another specialized DOE user facility, to issue the first joint JGI-EMSL call for proposals and a first round of projects has been approved. EMSL, also operated by the DOE Office of Biological and Environmental Research, provides scientific resources that directly complement and significantly expand our capacity to generate data that illuminate the complex pathways of cellular function. EMSL's mass spectrometry and other analytical capabilities allow researchers to measure with extreme precision minute changes in the entire complement of proteins expressed in cells. This ongoing partnership between EMSL and the DOE JGI will also provide access to nuclear magnetic-resonance spectrometers for specialized measurements that will be directed to metabolomics studies. This joint call allows users efficient access to a greater range of capabilities and encourages users of each facility to combine these capabilities in new and interesting ways. The projects selected from the initial round of JGI-EMSL submissions reflect the synergy of capabilities that these user facilities offer. They can be seen on page 50.

The DOE JGI dedicates a significant portion of its capabilities to enabling science at the DOE Bioenergy Research Centers (BRCs). The BRCs are accelerating the development of cellulosic biofuels, advancing the federal initiative that seeks to reduce U.S. gasoline consumption through the sustainable generation of clean biofuels. The JGI interacts extensively with the three Centers (BioEnergy Science Center led by Oak Ridge National Laboratory, DOE Great Lakes Bioenergy Research Center led by the University of Wisconsin, and Joint BioEnergy Institute led by Lawrence Berkeley National Laboratory), each targeting different plants both for model systems research and for improving actual feedstock crops.

In 2013, we worked with 1,155 users, generated more than 70 trillion bases of DNA sequence, and completed more than 4,300 user-initiated projects — a 60% increase over 2012. Along with our collaborators, we produced more than 125 papers, including 18 in high-profile journals such as *Nature*, *Science*, and *Proceedings of the National Academy of Sciences*, illustrating the diversity of high-impact DOE JGI contributions.

Selected highlights include:

- **Illuminating microbial dark matter.** To explore uncharted branches of the Tree of Life — so-called "microbial dark matter" — we employed single-cell genomics to target and sequence more than 200 uncultivated archaeal and bacterial cells from nine diverse habitats around the planet. With this additional genomic information, we were able to resolve ambiguous relationships between branches of the tree and uncovered unexpected features of microbial metabolism, challenging established boundaries among the three domains of life. This study expands the genomic representation of previously uncharacterized microbial branches and helps to illuminate the biological evolution on our planet. (*Nature*, July 14, 2013)

- **Marine blooming alga fills a gap in the Tree of Life.** We reported the first reference genome of a ubiquitous type of marine phytoplankton (*Emiliania huxleyi*) and conducted a comparison with 13 other isolates. This analysis revealed a "pan genome," a set of core genes shared among all strains, as well as another set of genes that were present in only some strains. This variability within species seems to support the organism's capacity both to thrive in habitats ranging from the equator to the subarctic and to form large-scale algal blooms under a wide variety of environmental conditions. Moreover, the study sheds light on the complex influence these

algae have on the global carbon cycle, driving CO_2 production or uptake, sequestration, and export to the deep ocean. (*Nature*, June 12, 2013)

- **The peach genome informs breeding strategies for biofuel crops.** We described the genome of peach, *Prunus persica*, one of the genetically best-characterized deciduous trees and a member of the same superfamily as poplar, a DOE JGI flagship biofuel feedstock species. The analyses suggested major genetic bottlenecks that have substantially shaped peach genome diversity. These findings may facilitate the breeding of candidate biofuel crops, particularly rapidly growing trees such as poplar, providing higher-energy-content fuels that can be efficiently extracted. (*Nature Genetics*, March 24, 2013)

- **The cotton genome, domestication of an important crop.** We showed that an abrupt five- to sixfold increase in ploidy (the number of sets of chromosomes in the cell nucleus) occurred some 60 million years ago in the ancestral cotton genome. We also described a more recent polyploidy event 1 million to 2 million years ago that brought together Old World and New World strains to produce the predecessor of modern cotton, *Gossypium raimondii*. This cotton was domesticated to produce the strain at the core of one of the world's largest industries — textiles — and the source of cotton oilseed. (*Nature*, December 20, 2012)

- **Biomass-degrading genes from fungi critical to managing the planet's carbon stores.** We described two *Agaricus bisporus* (button mushroom) genomes, their collections of genes, and the proteins they encode when grown on compost and during mushroom formation. The genomes encode a full repertoire of carbohydrate-degrading enzymes similar to that of wood-decayers, but different from the related symbiont *Laccaria bicolor* (also sequenced by the DOE JGI). These observations revealed genetic and enzymatic mechanisms governing adaptation to soils enriched with organic matter released during plant degradation. The sequenced genomes will expedite mushroom breeding for improved agronomic characteristics and highlight the critical contributions of these fungi toward soil structure and carbon sequestration in terrestrial ecosystems. (*PNAS*, October 8, 2012)

- **A tale of two algal genomes reveals chloroplast origins.** We sequenced two different algal species, *Bigelowiella natans* and *Guillardia theta*, to better understand the process of secondary endosymbiosis, in which a bacterium or other prokaryote is engulfed by a eukaryotic host. In organisms that acquired the ability to conduct photosynthesis by this mechanism, the nucleus from the ingested algal cell has disappeared. Our analysis revealed extensive genetic and biochemical mosaicism, in which genes from the host and the introduced symbiont come together for functions in the organelles and the cytoplasm of both algae. This conclusion has important implications for building ancestral trees of eukaryotes and more generally for our understanding of the evolution of the eukaryotic cell. (*Nature*, November 29, 2012)

You can read more about these and other achievements in the *Science* section that starts on page 20.

Finally, I cannot overstate the importance of the DOE JGI's computational infrastructure, which underpins all aspects of our activities. The torrent of information we and our collaborators generate has achieved "big data" status. Despite continued exponential growth in data production, the DOE JGI remains committed to making both raw and processed data available to the public through our

Genome Portal [http://genome.jgi.doe.gov]. The DOE JGI has established a robust and integral alliance with the DOE's National Energy Research Scientific Computing Center [NERSC; http://www.nersc.gov/] and the Energy Sciences Network [ESNet; http://www.es.net/] at Lawrence Berkeley National Laboratory. These partnerships permit more efficient and reliable access for users using our large suite of comparative genomic resources: the Integrated Microbial Genomes (IMG) data-management system for microbial genomes [http://img.jgi.doe.gov/], IMG/M [https://img.jgi.doe.gov/m] for metagenomes, MycoCosm [http://genome.jgi.doe.gov/programs/fungi/index.jsf] for fungal genomes, and Phytozome [http://www.phytozome.net/] for plant genomes. Statistics on the growing interest and utility of these systems can be found in the *Impact* section that begins on page 12.

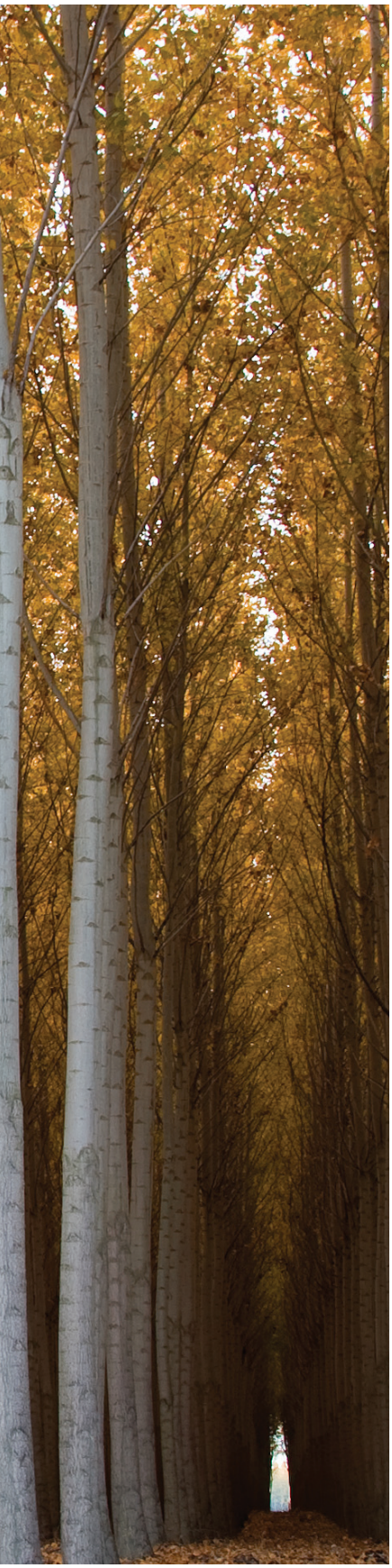
The DOE JGI is also depositing its various gene annotations and genomic data into DOE's new Systems Biology Knowledgebase [KBase; http://kbase.science.energy.gov/]. KBase is an emerging software and data environment designed to enable researchers to collaboratively generate, test, and share new hypotheses about gene and protein functions; perform large-scale analyses on a scalable computing infrastructure; and model interactions in microbes, plants, and their communities. Because the DOE JGI is the single largest producer of data relevant to DOE's science mission, it is essential that these data be made available seamlessly to KBase users — and that is the focus of ongoing collaboration between the groups. Further, the DOE JGI's effort to improve microbial genome annotation through transposon mutagenesis will be crucial for improving functional predictions that are at the core of KBase.

The DOE JGI continues to evolve as a genomic science user facility. Our 10-Year Strategic Vision sets clear priorities for us to ensure the uniqueness and value of the DOE JGI as an important resource for DOE researchers. With this, our organization remains motivated and committed yet agile enough to respond to rapidly changing technologies required to address global energy and environmental challenges. Our progress over the past year clearly reinforces our position as the premier genome science user facility dedicated to energy and the environment. I encourage you to stay engaged with us as we realize the promise of our Strategic Vision.

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

DOE Mission Areas

Andy Simonds



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 microorganisms that can photosynthesize or 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.

Organizational Structure

Strategic Management

Eddy Rubin
JGI Director

Dan Rokhsar
Chief Informatics Officer

Axel Visel
User Programs and Strategic Planning

Ray Turner
Deputy Director, Operations

Len Pennacchio
Deputy Director, Genomic Technologies

Jim Bristow
Deputy Director, Science Programs

Chia-Lin Wei
Sequencing Technologies

Atif Shahab
Institutional Informatics

Dan Rokhsar
Eukaryote Super Program

Nikos Kyrpides
Prokaryote Super Program

Igor Grigoriev
Fungal Program

Jeremy Schmutz
Plant Program

Susannah Tringe
Metagenome Program

Tanja Woyke
Microbial Program

Impact **2013**

Primary Users **Fiscal Year 2013**

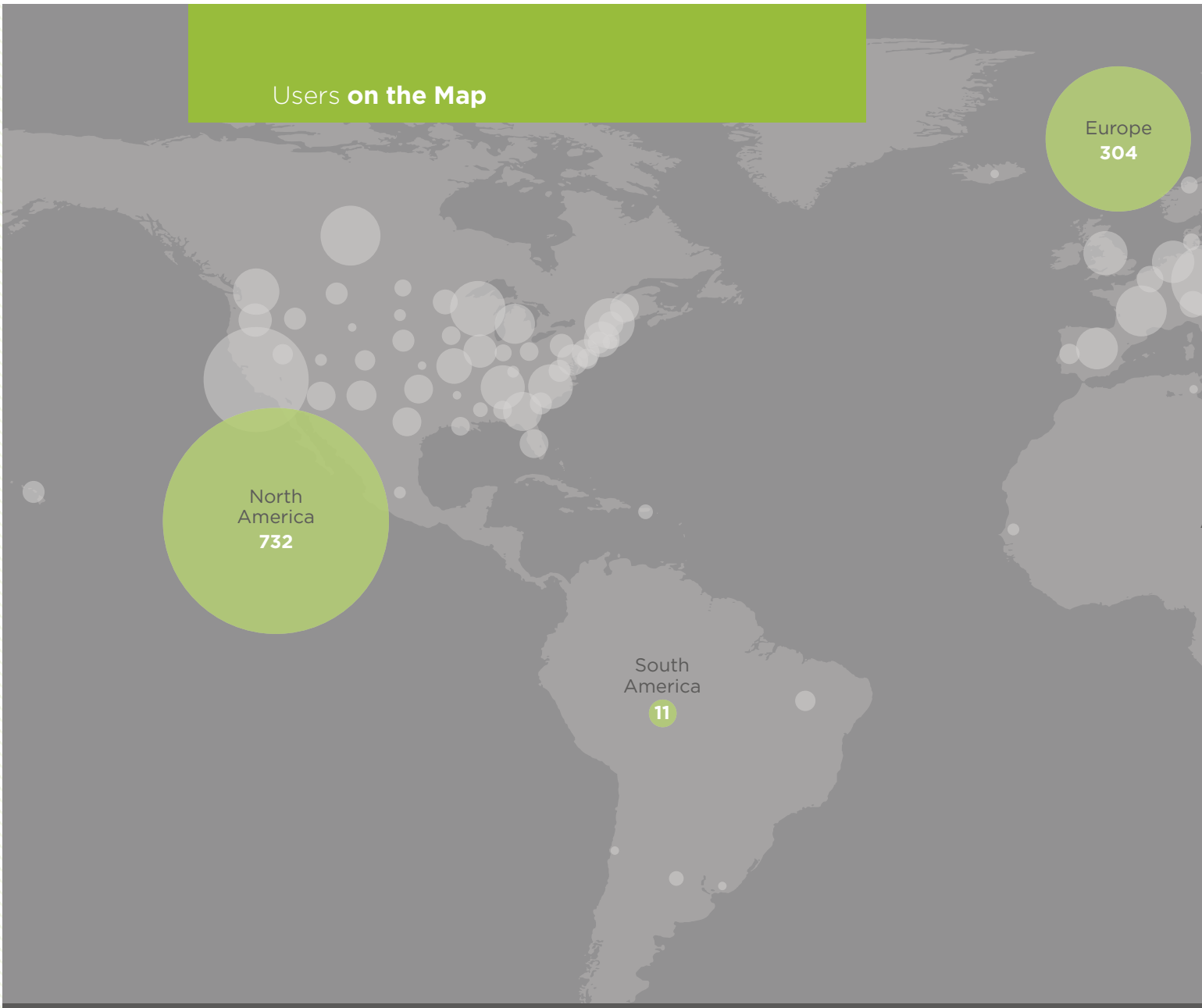
This category captures the primary users of the DOE JGI, which includes PIs and their collaborators on all user projects that were active during FY 2013. Each user is uniquely identifiable and is counted once per year regardless of the number of active projects in which he/she may be involved. This count does not include collaborators who are employed by the DOE JGI or funded through the DOE JGI's partner subcontracts.

Users **by Proposal Type**

Non-Bioenergy Research Centers (BRC)	1,042
DOE BRCs	113

Users **by Institution Type**

Academic	939
Government	94
DOE Laboratory	87
Company	35



North America	732	South America	11	Africa	6	Asia	53
Canada	52	Brazil	6	Senegal	2	Japan	19
Mexico	2	Argentina	3	South Africa	2	India	11
Puerto Rico	3	Chile	1	Egypt	1	Republic of Korea	7
United States	675	Uruguay	1	Tunisia	1	China	4
						Israel	4
						Hong Kong	2
						Indonesia	2
						Singapore	2
						Malaysia	1
						Philippines	1

Europe	304		
Germany	74	Greece	7
France	37	Portugal	6
United Kingdom	28	Slovenia	5
Netherlands	26	Denmark	4
Spain	25	Norway	4
Sweden	16	Czech Republic	2
Austria	13	Hungary	2
Belgium	10	Poland	2
Italy	10	Serbia	2
Switzerland	10	Turkey	2
Finland	9	Iceland	1
Russian Federation	9		

Australia & New Zealand	49
Australia	35
New Zealand	14

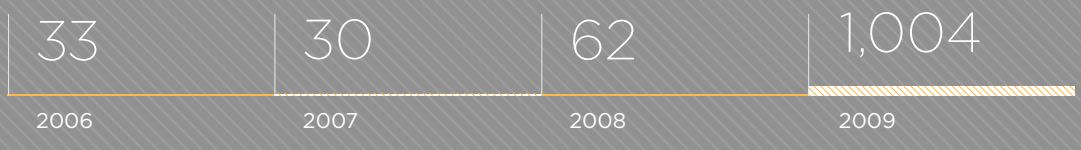
Data **Users**

DOE JGI systems also support investigators who have utilized computational, educational, and/or data resources located at the DOE JGI, but are not included in the primary user count because their projects were not conducted as part of DOE JGI's user programs.

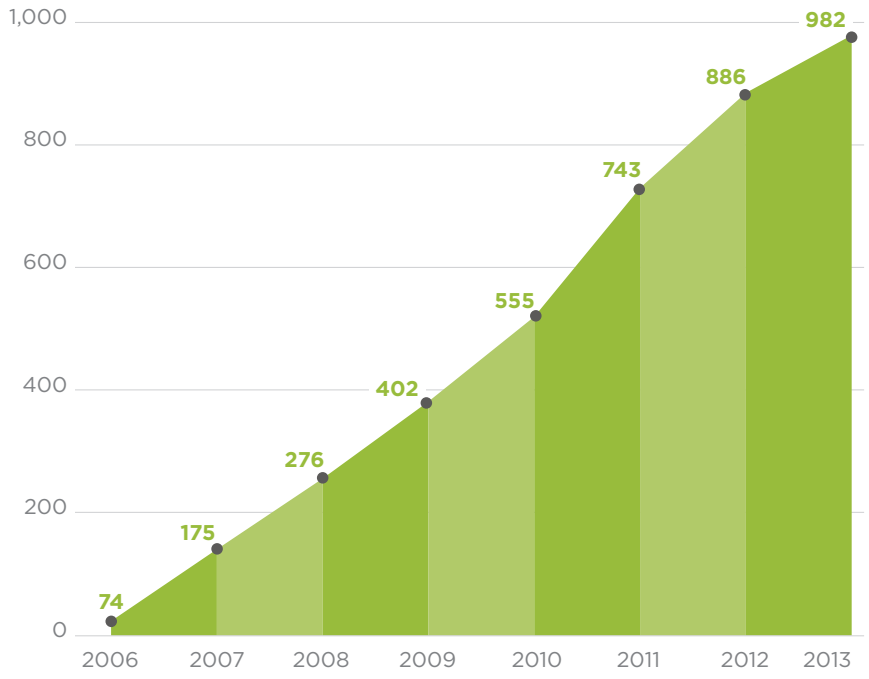
Registered Analysis/Tool *(new/active annotators)*	
MycoCosm	**2,871/65**
IMG and IMG/M	**1,523**
Educational	
IMG Annotation Collaboration Toolkit (IMG-ACT)	**366**
Workshops and Meetings	
Workshop Participants	**505**
User Meeting Attendees	**225**
Web Portal Visitors *(unique visits)*	
JGI Portals and MycoCosm	**279,987**
IMG Systems	**114,905**
GOLD	**45,301**
Phytozome	**104,828**
VISTA	**13,370**

Sequence Productivity

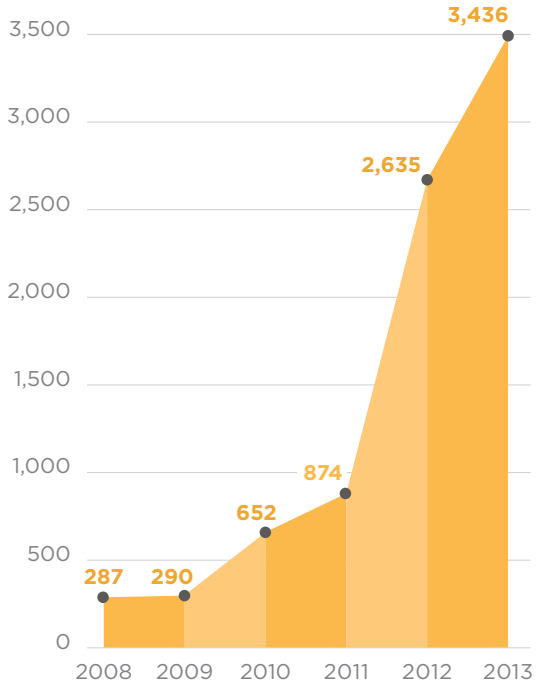
(in billions of bases or GB)

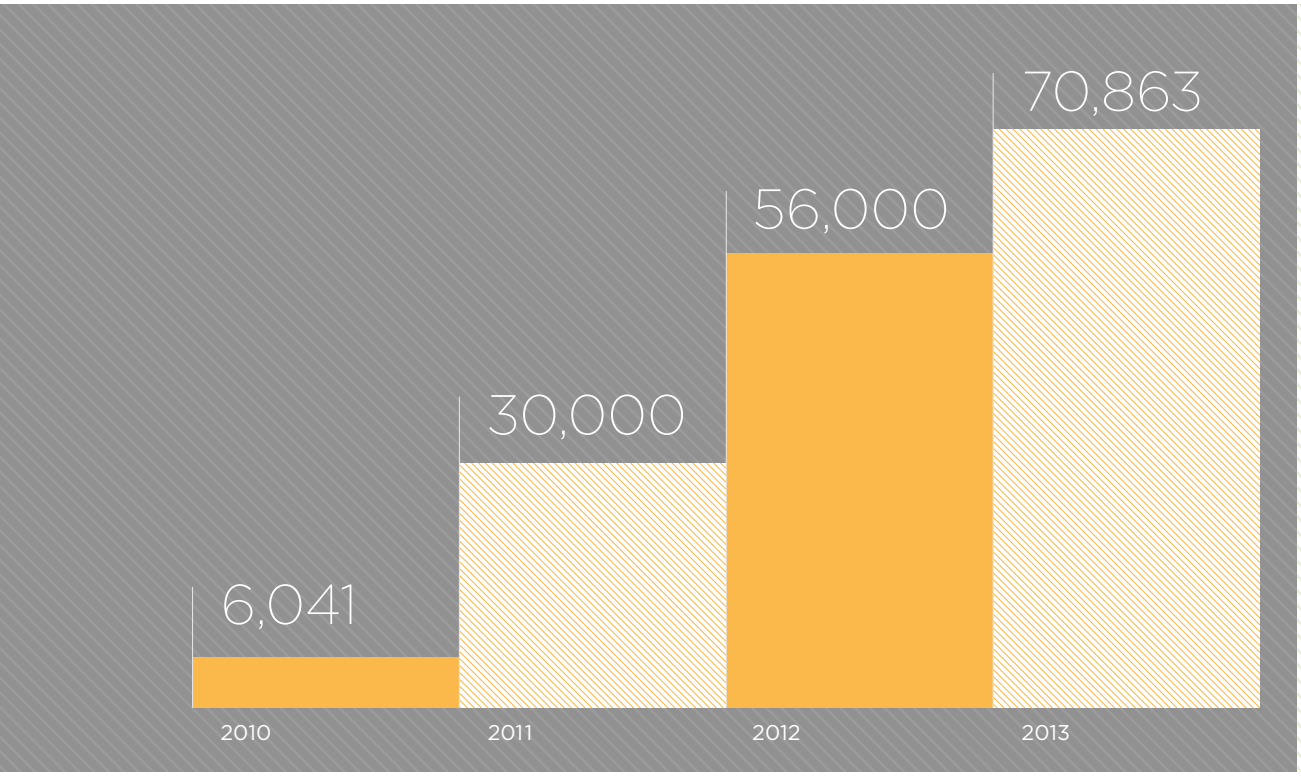


Cumulative Number of Scientific Publications

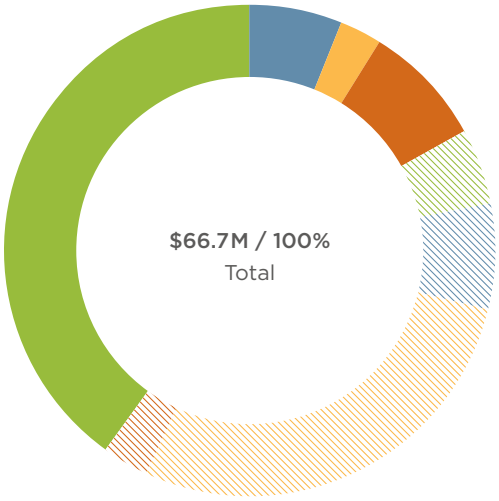


Projects Completed (per year)





Operating Expenses **FY13**



$26.4M / 40%
Science & Analysis

$3.5M / 5%
Facilities

4.1M / 6%
R&D

$4.6M / 7%
NERSC/IT

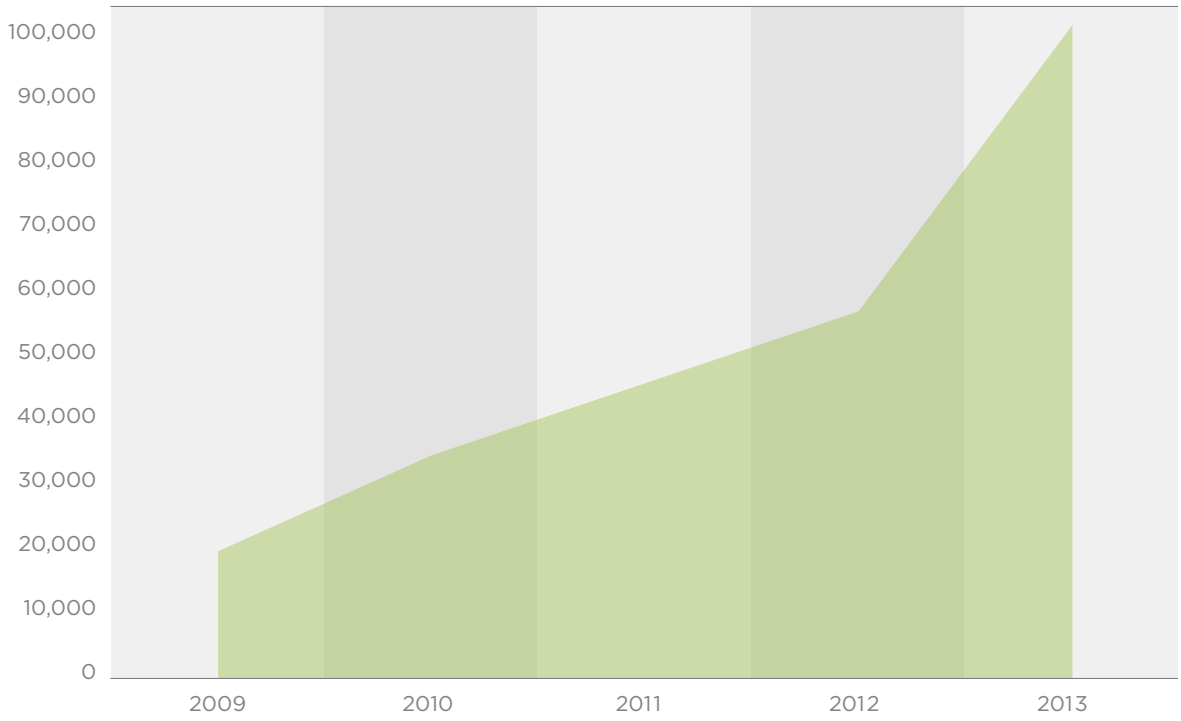
$1.8M / 3%
Management

$19.1M / 28%
Sequence/Data Generation

$5.3M / 8%
Operations

$1.9M / 3%
User Support

Citations



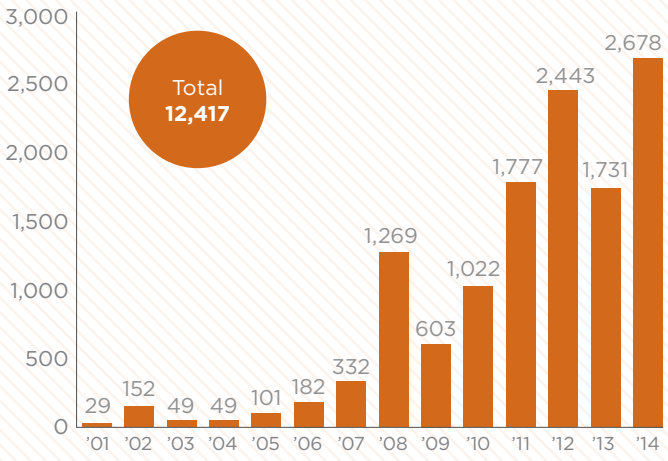
Brachy — little weed, huge impact. **A case study.**

While cellulosic energy crops — particularly grasses such as switchgrass and *Miscanthus*, and fast-growing trees including poplar — are poised to become a major source of renewable energy in the United States, we still know little about the genetic traits that affect their usefulness for biofuels production. The temperate wild grass species *Brachypodium distachyon* ("*Brachy*") is a model plant sequenced by the DOE JGI to better understand how to breed grasses into superior energy crops. *Brachy* is small in size, can be grown rapidly, is self-fertilizing, and has simple growth requirements.

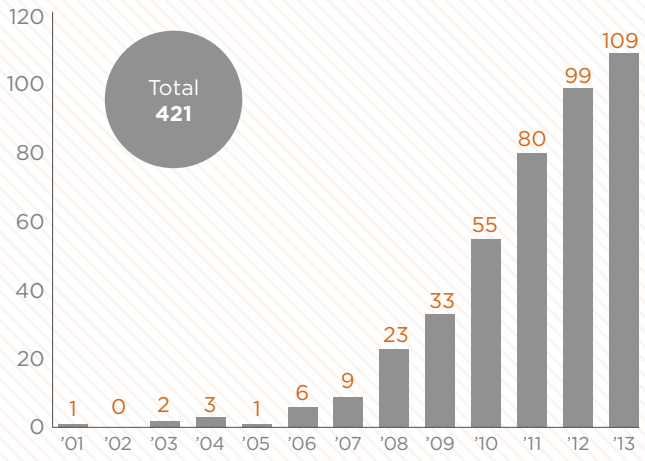
Using *Brachy* as an example, the diagram below depicts how the DOE JGI's advancement of our understanding of model organism genomes generates a dramatic surge of interest in the research community. From the time the project was announced, to the point when the sequence was released to the public and the DOE JGI and collaborators published the analysis of the *Brachy* genome in the February 2010 issue of *Nature*, success could be measured in seeds ordered and delivered to researchers and, importantly, a torrent of *Brachy* publications and citations of DOE JGI's analysis paper.

Longtime DOE JGI collaborator John Vogel, Ph.D., Research Molecular Biologist, Western Regional Research Center, U.S. Department of Agriculture, Albany, Calfornia.

Seed Packets Shipped



Publications *Number where genome paper is cited*



Science: Year in Review

The DOE JGI is on the front lines of many of the most pressing scientific inquiries of our day. Our researchers are filling in the branches in the Tree of Life as they track elusive “microbial dark matter” and probe the genomes of some of the most common trees and other plants — even weeds — to help us achieve efficient biofuel production. They delve underground and underwater to discover microbial novelty and important lineages and diversity, which will offer insights into the global carbon cycle.

Research at the DOE JGI can be organized along three lines: bioenergy, carbon cycling, and biogeochemistry. These fields are often intertwined, allowing our investigations to be applied in multiple ways. They cover the spectrum of organisms, from the infinitesimally small microbes to familiar farm crops.

Bioenergy

The targets of DOE JGI sequencing and analysis have included some common and agriculturally important crops such as soybeans, sorghum, cotton, and peaches. While these efforts will likely lead to crops better equipped to cope with environmental stress and other agronomic challenges, these species also hold deeper secrets in their tissues: They may lead us to improved biofuels production.

Cotton bolls are almost pure cellulose and, as such, represent a target biomass for next-generation biofuels. Growing, processing, and manufacturing cotton is a major global industry. In the United States, about 200,000 jobs are related to cotton production and processing, contributing about $27 billion to the economy. This includes cotton fiber for textiles and apparel, cottonseed oil, and meal byproducts.

In the December 20, 2012, edition of *Nature*, an international consortium of researchers from 31 institutions including a team from the DOE JGI presented a high-quality draft assembly of the simplest cotton (*Gossypium raimondii*) genome. The team also compared the genome from this ancestral species indigenous to the Americas to several other sets of cotton data contributed by the U.S. Department of Agriculture (USDA). The results have allowed researchers to trace the evolution of cotton over millions of years, from wild varieties to the domesticated species now associated with textile production. The DOE JGI's contribution of sequencing and assembling the 760 million base-pair genome stems from a Community Sequencing Program proposal led by University of Georgia Professor Andrew Paterson.

"This cotton data will help accelerate the study of gene function, particularly cellulose biosynthesis, the understanding of which is fundamental to improved biofuels production," said Jeremy Schmutz, head of the DOE JGI Plant Program and a faculty investigator at the HudsonAlpha Institute for Biotechnology, who led the effort to sequence and assemble the genome.

Don Jones, Director of Agricultural Research at Cotton Inc., said this *G. raimondii* gold-standard genome will be the foundation for sequencing upland cotton, *G. hirsutum*, which makes up most of the worldwide field crop. Another species, *G. barbadense*, produces pima cotton but accounts for less than 2 percent of the cotton crop. "This sequence is a cornerstone that will help advance our knowledge so we more thoroughly understand the biology that leads to enhanced yield, improved fiber quality, and better stress tolerance, all improvements that will benefit growers in the not-too-distant future," Jones said.

Schmutz talks about the cotton genome project on YouTube at http://bit.ly/JGI-cotton-video.

This Lovell peach tree at Clemson University provided the DNA used to determine the peach genome. *(Clemson University)*

Rapidly growing trees like poplars and willows are candidate "biofuel crops" from which researchers expect to be able to efficiently extract cellulosic ethanol and higher-energy-content fuels. Domesticating these crops requires a deep understanding of the physiology and genetics of trees, and scientists are turning to long-domesticated fruit trees for hints. Although peach and poplar do not seem to have many traits in common, both trees are part of the rosid superfamily, which includes not only fruit crops like apples, strawberries, cherries, and almonds, but many other plants, including the rose, which gives the superfamily its name.

"The close relationship between peach and poplar trees is evident from their DNA sequence," said Jeremy Schmutz, head of the Plant Program at the DOE JGI, and a faculty investigator at the HudsonAlpha Institute for Biotechnology.

In the March 24 edition of *Nature Genetics*, Schmutz and several colleagues were part of the International Peach Genome Initiative (IPGI) that published the 265-million base pair genome of the Lovell variety of *Prunus persica*.

"Using comparative genomics approaches, characterization of the peach sequence can be exploited not only for the improvement and sustainability of peach and other important tree species, but also to enhance our understanding of the basic biology of trees," the team wrote. They compared 141 peach gene families to those of six other fully sequenced diverse plant species to unravel unique metabolic pathways, such as those that lead to lignin biosynthesis — the molecular "glue" that holds the plant cells together — a key barrier to deconstructing biomass into fuels.

For bioenergy researchers, the size of the peach genome makes it an ideal plant model for studying genes found in related genomes such as poplar — one of the DOE JGI's Plant Flagship Genomes (http://bit.ly/JGI-Plants) — and for developing methods for improving plant biomass yield for biofuels. Learn more about poplar and DOE JGI Plant Flagship Genomes at http://genome.jgi.doe.gov/programs/plants/flagship_genomes.jsf.

"One gene we're interested in is the so-called 'evergreen' locus in peaches, which extends the growing season," said Daniel Rokhsar, DOE JGI Eukaryotic Program head under whose leadership sequencing of the peach genome began back in 2007. "In theory, it could be manipulated in poplar to increase the accumulation of biomass."

To better understand how life on Earth thrives and evolves, we probe processes that control all our natural systems. Some photosynthetic plants reveal a fascinating history — far back in time, they stole their photosynthetic abilities from algae. And yet the remnants of an original algal nucleus can still be found in the host plant. The DOE JGI collaborated to sequence and analyze genomes and transcriptomes from two algae. The results are filling in the picture of how photosynthetic organisms regulate and maintain the global carbon cycle. Similarly, we sequenced the plant *Capsella rubella*, which self-fertilizes, and found glimpses of genome expansion and shrinkage to help scientists understand the sequences in plant genomes that control activation of certain traits.

Widely used as a model for plant research, *Arabidopsis thaliana* was the first plant to have its genome completely sequenced. However, there is still much to learn about this plant, including the function of its many DNA conserved noncoding sequences (CNSs). These regions hold key roles in activating certain traits during plant development, but why or how they are conserved is not yet clear.

To help answer these questions, the genomes of multiple plants, including *A. thaliana* and two others sequenced by the DOE JGI, were compared in two studies published in *Nature Genetics* in June 2013.

The first study, published on June 9, focused on *Capsella rubella's* "selfing," or the ability to self-fertilize, and its effect on *C. rubella's* genome. The DOE JGI sequenced and compared it with *C. grandiflora*, from which *C. rubella* split about 200,000 years ago, and members of the closely related *Arabidopsis* genus.

Capsella rubella is related to *Capsella bursa-pastoris*, or shepherd's-purse (shown here), an invasive weed. *(Montse Poch)*

C. rubella showed a mass decline of the removal of harmful mutations without a naturally occurring alteration in the amount of genes present that can move between chromosomes. From these findings, it is theorized that a dramatic event left *C. rubella* in a situation where a need for pollinators outweighed the known negatives of inbreeding and caused the *C. rubella* to shift into selfing. Though this caused the *C. rubella* to face a bottleneck, its ancestral genome structure remained intact.

"The factors driving such contrasting modes of genome expansion and shrinkage are far from resolved, and it will be important to broaden future comparisons to larger phylogenetic scales to better understand the processes driving genome structure evolution," the team said on the study's conclusion and future plans.

The second study, published June 30, focused on factors that account for reduced diversity and gene mutation rate in a species. The team compared three newly sequenced genomes with six previously sequenced genomes including *C. rubella* and another brassica species, the salt-tolerant *Eutrema salsugineum*.

Scientists identified 90,000 noncoding base pairs, which account for 17 percent of *A. thaliana*'s genome. The study yielded the first high-resolution, genome-wide map of noncoding regions. They concluded that the plants retained these sequences because they are vital in the evolution of genome organization. This information, along with ongoing research, will enable scientists to further understand the sequences in plant genomes that control the activation of certain traits in the development of these plants.

Many of the world's most important photosynthetic eukaryotes — such as plants — did not by themselves develop the ability to combine carbon dioxide, water, and solar energy to produce organic compounds. Rather, they got their light-harnessing organelles — chloroplasts — indirectly by stealing them from other organisms. The evolution of chloroplasts, the photosynthetic compartments of plants and algal cells, is complex but has had a profound effect on our planet.

Chloroplasts originally evolved from photosynthetic bacteria through a process called primary endosymbiosis, in which a bacterium or other prokaryote is engulfed by a eukaryotic host. The chloroplasts of red and green algae have subsequently come to reside within other, previously nonphotosynthetic eukaryotes by secondary endosymbiosis. Such events have contributed to the global diversity of photosynthetic organisms that play a crucial role in regulating and maintaining the global carbon cycle. In most organisms that acquired photosynthesis by this mechanism, the nucleus from the ingested algal cell has disappeared, but in some cases it persists as a residual organelle known as a nucleomorph. Such organisms have four distinct genomes.

To better understand the process of secondary endosymbiosis and why nucleomorphs persist in some organisms, an international team of 73 researchers at 27 institutions, including the DOE JGI, collaborated to sequence and analyze the genomes and transcriptomes (the expressed genes) of two tiny algae. The team, led by John Archibald of Canada's Dalhousie University, published its findings on the algae *Bigelowiella natans* and *Guillardia theta* online November 28, 2012, in *Nature*.

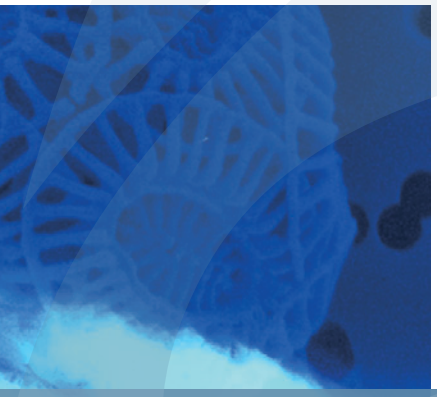
Archibald compared these algae to Russian nesting dolls with "sophisticated sub-cellular protein-targeting machinery" and four genomes derived from the two eukaryotes that merged over time. "Approximately 50 percent of the genes in both genomes are 'unique' with no obvious

counterpart in other organisms," he added. "This indicates just how different they are from characterized species."

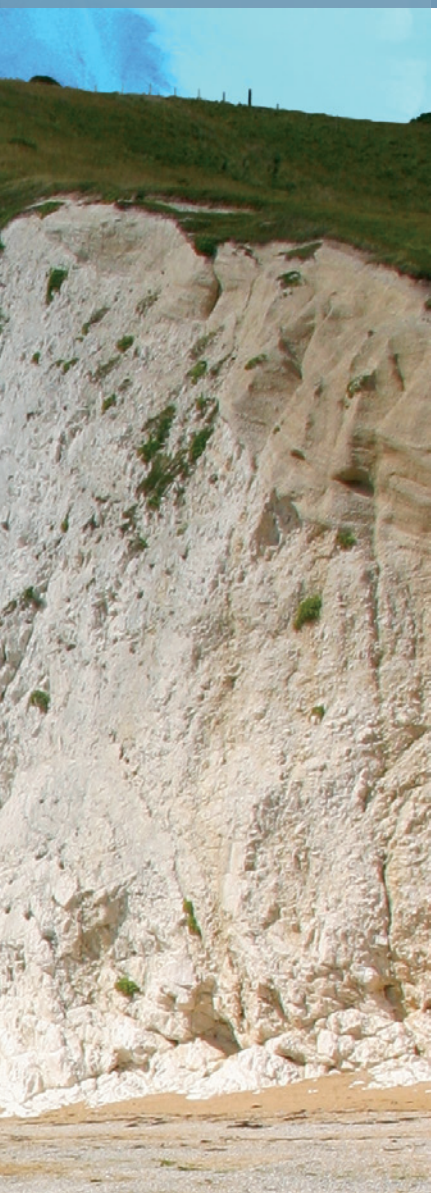
"The reason for the persistence of nucleomorphs in both organisms appears to be surprisingly simple: they are no longer able to transfer their DNA to the host cell nucleus by the process of endosymbiotic gene transfer," said Archibald. Unlike most other secondarily photosynthetic eukaryotes in which the symbiont's genetic matter has completely migrated over to the host, in cryptophytes and chlorarachniophytes the nucleus and chloroplast from the engulfed algae remain partitioned off from the host cell. "As a consequence," he said, "genetic and biochemical mosaicism — the presence of two or more populations of cells with different genotypes in one individual — is rampant in *G. theta* and *B. natans*."

Nucleomorph genomes have probably persisted simply because the mechanism for transfer of genes to the nucleus was closed off, rather than because nucleomorphs had to be retained as separate entities.

Scanning electron microscopy image of the endosymbiotic algae *Guillardia theta* (with two flagella), the first cryptophyte to be sequenced, and *Bigelowiella natans* (with the single flagellum), the first chlorarachniophyte to be sequenced. *(Dr. Geoff McFadden, University of Melbourne, Australia)*



Carbon Cycle



The basis of most ocean food chains is made up of single-celled photosynthetic algae. One of these algae, *Emiliania huxleyi*, has the ability both to trap large amounts of organic carbon and to emit carbon dioxide — and this talent makes it an important subject for research. Using next-generation sequencing technologies, DOE JGI researchers revealed the first-ever algal pan-genome, in which individuals possess a shared core of genes, yet have different gene sets to allow survival in local environments. The research helped explain such phenomena as algal blooms. Other marine microbes have pared-down genetic codes, a strategy that helps them survive in some nutrient-poor environments. Using a special single-cell strategy, DOE JGI research provided a first look at the survival adaptations of bacteria in various challenging environments.

The towering white cliffs of Dover in England are made of the chalky white shells that envelop the single-celled photosynthetic alga, *E. huxleyi*. "*Ehux*" is a coccolithophore, with an exoskeleton made of calcium carbonate. Even though the process by which the alga's "armor" forms releases carbon dioxide, *Ehux* can trap as much as 20 percent of organic carbon, derived from CO_2, in some marine ecosystems.

Ehux and its brethren are the basis of most ocean food chains. Phytoplankton biomass exceeds that of all marine animals combined. Its versatility in either contributing to primary production of organic compounds from carbon dioxide or adding to CO_2 emissions makes *Ehux* a critical player in the marine carbon cycle. Sequenced by the DOE JGI, the *Ehux* genome was compared with sequences from other algal isolates and the results were reported in the June 12, 2013, edition of *Nature*.

Part of the third-most-abundant group of phytoplankton, behind the diatoms and dinoflagellates, the *Ehux* strain sequenced by the DOE JGI was isolated from the South Pacific and is the first reference genome for coccolithophores. The project took longer than expected due to the complexities and size of the *Ehux* genome. Originally estimated to be about 30 million bases, closer to a diatom, the genome ended up being closer to 141 million bases. Starting with an individual investigator working with the DOE JGI for more than 10 years, the team forged a strong community of users to bring the project to completion.

With the advent of next-generation sequencing technologies, the team was able to conduct a comparison of 13 *Ehux* strains, revealing the first-ever algal "pan-genome." *Ehux* does not exist as a clearly defined "species" with a uniform genome, but as a more diffuse community of genomes — a pan-genome — with different individuals possessing a shared "core" of genes, but supplemented by different gene sets thought to be useful in dealing with the particular challenges of their local environments.

The team found variability in the *Ehux* genome that helps explain the alga's ability to thrive in oceans from the equator to the subarctic and cause algal blooms in the spring and summer that can cover several hundred thousand square kilometers.

The researchers also found that the core gene sets include genes that allow *Ehux* to thrive in low levels of phosphorus and to assimilate and break down nitrogen-rich compounds. Additionally, the algal genome offers hints that *Ehux* may be involved in the global sulfur cycle as it is able to produce a compound that can influence cloud formation and therefore climate. The conditions under which *Ehux* produces the compound and the amounts synthesized are questions researchers are looking forward to investigating.

Among the challenges in cultivating microbes is a process known as genome streamlining — over time, some microbes have pared down their genetic codes to those genes necessary to survive in specific nutrient-poor environments. A marine microbial study published in the June 25, 2013, early edition of the *Proceedings of the National Academy of Sciences* suggests that as finicky as lab-cultured microbes are known to be, free-living microbes are even more so.

The researchers, including those from DOE JGI and the Bigelow Laboratory for Ocean Sciences in Maine, isolated single microbial cells from samples collected in the Gulf of Maine, the Mediterranean Sea, the North Pacific, and the South Atlantic. The team then sequenced and assembled draft genomes of 56 single amplified genomes (SAGs). "Using this large-scale single-cell strategy, our study provides the first glimpse into the geographic distribution and genomic features of some of the bacterial inhabitants, most of which have not yet been cultivated in the laboratory," said Tanja Woyke, head of the DOE JGI Microbial Program and one of the senior authors of the study.

Research under way in the new Bigelow Laboratory Single Cell Genomics Center in Maine. *(Dennis Griggs)*

Water samples were taken from the ocean photic zone, or sunlight zone, the region of the ocean penetrated by sufficient sunlight to allow for photosynthesis. “This is where most of the sun’s energy is captured in the ocean,” said Woyke. The SAGs were then compared with existing microbial cultures and metagenomic data sets. The team found evidence that unlike the free-living microbes, the cultured microbes thrive in more nutrient-rich environments. “These findings illustrate that bacterial adaptations to nutrient-poor environments are widespread and that dispersal is not limited, but rather water temperature and latitude are key drivers in geographic distribution of certain bacterial groups,” Woyke said. “To put it in other words — bacteria from the sunlight zone can travel large distances, but they then like to settle and thrive in regions of shared temperature and latitude.”

“Single-cell genomics continues to push research to a new level, previously inaccessible to scientists,” said Ramunas Stepanauskas, director of the Bigelow Laboratory Single Cell Genomics Center and senior author of this study. “While other cultivation-independent tools are available, these rarely allow us to assess who in a community does what, a central question in microbial ecology. The single-cell sequencing approach allowed us to link an organism’s identity and functional gene repertoire for more than fifty key community members of the ocean photic zone, shedding light on the biology of this ecosystem that covers 70% of the planet.”

At minus 20°C, in a saline lake in Antarctica, the water never freezes. Yet this forbidding environment presents rich opportunities for DOE JGI researchers. Special microbes live in Deep Lake, thriving in high-salt-level waters of extreme cold, and scientists partnering with our facility have completed an ecological portrait of this microbial community. The researchers’ results could help with costs associated with cleaning contaminated sites and oil recovery.

Sequestered in Antarctica’s Vestfold Hills, Deep Lake became isolated from the ocean 3,500 years ago by the Antarctic continent rising, resulting in a saltwater ecosystem that remains liquid in extreme cold, providing researchers a unique niche for studying the evolution of the microbes that now thrive under such conditions. Deep Lake’s microscopic inhabitants are dominated by haloarchaea, microbes that require high salt concentrations to grow and are naturally adapted to conditions that would prove lethally cold to other organisms. In a detailed analysis [http://www.pnas.org/content/early/2013/09/25/1307090110.abstract] published online the week of September 30, 2013, in the journal *Proceedings of the National Academy of Sciences*, researchers have, for the first time, achieved a complete ecological picture of the Deep Lake microbial community.

A team led by Rick Cavicchioli of the University of New South Wales, Australia, partnered with the DOE JGI to generate sequence data from DNA isolated from individual microbes, and compared them with metagenomic (microbial community) information sampled at various depths of Deep Lake.

“Understanding how haloarchaea can thrive in Deep Lake could be used to develop engineering concepts for reducing energy costs in a variety of situations, such as for cleaning up contaminated sites in permanently or seasonally cold regions,” Cavicchioli said. Owing to the ability of salt-loving enzymes to function under extremes, he suggests they could also be used as catalysts for peptide synthesis and enhanced oil recovery, and can function in water-organic solvent mixtures. “These enzymes will be especially useful for transforming contaminated sites with particularly high levels of petroleum-based products,” he added.

Deep Lake’s extremes have rendered the microbial neighborhood rather homogeneous. Four isolates in the study represented about 72 percent of the cells in the community. Though gene exchange across species boundaries is considered infrequent, the researchers observed that haloarchaea living in the lake’s hypersaline environment practice it comparatively often, like neighbors “chewing the fat” in a small-town coffee klatch. “It’s intriguing that while gene exchange is rampant, species lineages appear to be maintained by virtue of each species having a high level of specialization, enabling niche partitioning and peaceful coexistence,” said Cavicchioli of their findings. “Haloarchaea

Deep Lake in Antarctica as an expedition work site in November 2008, shown with mobile work shelters and equipment for sampling. *(Rick Cavicchioli)*

are known for being 'promiscuous,' that is, prone to exchange DNA between themselves. Our study demonstrated that this exchange occurs at a much higher level than has previously been documented in nature. They communicate, share, specialize, and coexist."

What distinguishes this "conversation" is that the haloarchaea of Deep Lake exchange the information of DNA not just between species but among distinct genera, and moreover in huge tranches, some 35,000 letters of code, with not a letter out of place.

Cavicchioli noted that, "as the content being shifted around lack core genes, it speaks to these microbes' ability to be flexible and collaborative. This shuttled gene content could confer such benefits as resistance to viruses or bolster their ability to respond to specific environmental factors. Moreover, the markers that we analyzed indicated that a high level of gene exchange occurs throughout the Deep Lake community."

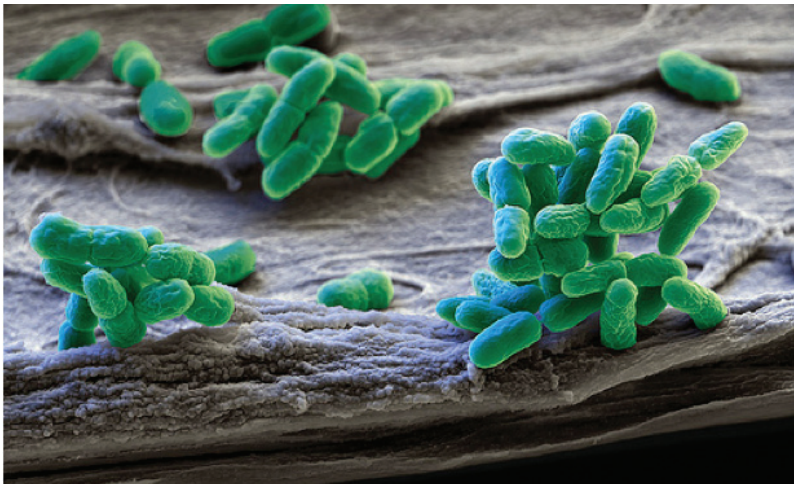
From the cold depths to an excess of sunshine — a team led by DOE JGI researchers is on the trail of cyanobacteria that can protect themselves from excess sunlight using special antennae. Light is crucial for photosynthetic organisms, but one can have too much of a good thing. Excess light can harm organisms when the amount of energy absorbed exceeds the rate of carbon fixation.

To protect themselves from excess light, cyanobacteria rely on light-harvesting antennae called phycobilisomes that can sense light conditions in order to efficiently collect this energy. A protein known as OCP helps regulate the phycobilisomes, and is in turn regulated by a protein known as FRP.

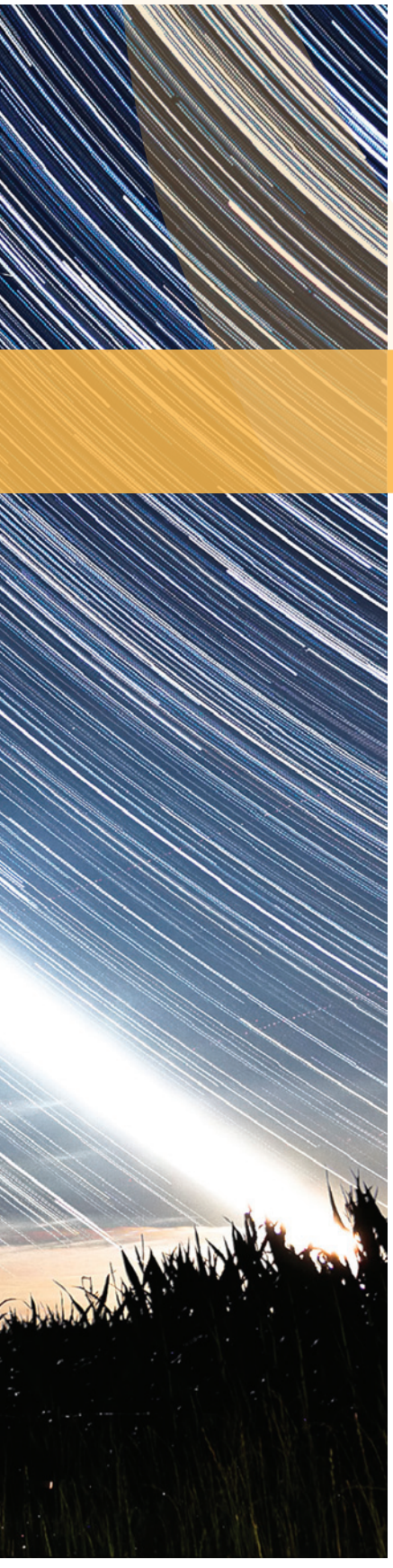
In an article published ahead online May 28, 2013, in *Proceedings of the National Academy of Sciences*, a team led by DOE JGI researchers focused on the crystal structure of the FRP protein in a species of *Synechocystis* bacteria to learn more about how cyanobacteria protect themselves from excess light.

They were able to look at the crystal structure using beamlines at Lawrence Berkeley National Laboratory's Advanced Light Source. They found that FRP has two forms: It forms a dimer — a structure formed from two similar sub-units — in the active state and otherwise appears as a tetramer (with four sub-units). The researchers also found a region of highly conserved residues on the dimer structure of FRP that are essential to the protein's activity.

Their findings have led the team to propose a model of how the FRP and OCP proteins interact with phycobilisomes under both high-light and low-light conditions in order to ensure the cyanobacteria can safely and efficiently harness light to conduct photosynthesis.



SEM of *Synechocystis* cyanobacteria, which helped researchers study the crystal structure of the FRP protein. *(BASF)*



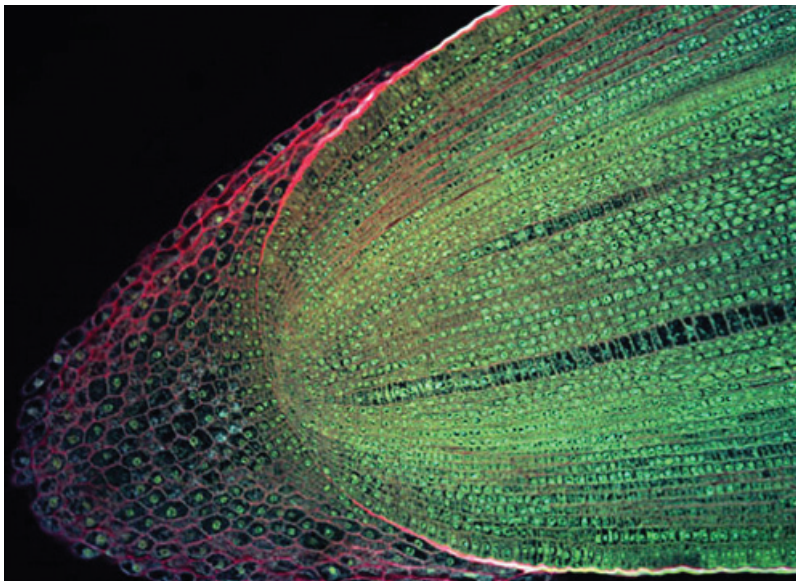
Biogeochemistry

Some frontiers in DOE JGI research are so elusive they can only be inferred by indirect effects. You have heard about dark matter, which is said to account for the majority of mass in the universe. Here on Earth, DOE JGI researchers are seeking "microbial dark matter," an all-but-invisible yet vital infrastructure of life on our planet. Dramatic advances in this area by an international collaboration led by the DOE JGI have led to discoveries that fill out formerly uncharted areas in the bacterial and archaeal Tree of Life.

In cosmology, dark matter is said to account for the majority of mass in the universe; its presence, however, is inferred by indirect effects rather than detected through telescopes. The biological equivalent is "microbial dark matter," a pervasive yet practically invisible infrastructure of life on the planet that can have profound influences on the most significant environmental processes, from plant growth and health, to nutrient cycles in terrestrial and marine environments, to the global carbon cycle, and possibly even to climate processes. By employing next-generation DNA sequencing of genomes isolated from single cells, great strides are being made in the monumental task of systematically bringing to light and filling in uncharted branches in the bacterial and archaeal Tree of Life. In an international collaboration led by the DOE JGI, the most recent findings from exploring microbial dark matter were published online July 14, 2013, in the journal *Nature*. The study builds on a DOE JGI pilot project, the Genomic Encyclopedia of Bacteria and Archaea (GEBA: http://www.jgi.doe.gov/programs/GEBA/).

This microbial dark-matter campaign targeted uncultivated microbial cells from nine diverse habitats: Sakinaw Lake in British Columbia; the Etoliko Lagoon of western Greece; a sludge reactor in Mexico; the Gulf of Maine; off the north coast of Oahu, Hawaii; the Tropical Gyre in the south Atlantic; the East Pacific Rise; the Homestake Mine in South Dakota; and the Great Boiling Spring in Nevada. From these samples, the team laser-sorted 9,000 cells, from which they were able to reassemble and identify 201 distinct genomes that represent 28 major previously uncharted branches of the Tree of Life.

What takes place underground and underwater has enormous impact on the health of plants and the planet. DOE JGI researchers were part of a team that sequenced bacterial diversity in fields where maize is grown. The results challenged assumptions about soil microbial diversity. And investigating organisms living in water, a team that included the DOE JGI assembled two nearly complete genomes from a candidate phylum found in a pool at Yellowstone and a hot spring in Nevada.



Longitudinal section of a root tip of maize (*Zea mays*). To the left of the image, the large, loosely packed cells of the root cap can be seen. These cells protect the actively dividing undifferentiated plant tissue as the root grows down through the soil and encounters the microbial communities that colonize the rhizosphere.

To get around the difficulty of growing most microbes in the laboratory, recent efforts have focused on conducting surveys based on sequencing marker or 16S ribosomal RNA genes, which are conserved across microbial lineages because of their essential role as "housekeeping" genes —critical for the organism's survival. Genome sequencing of the rest of the genomes of most of these lineages is, however, proceeding much more slowly. "Microbial genome representation in the databases is quite skewed," said Chris Rinke, DOE JGI postdoctoral fellow and first author of the study. "More than three-quarters of all sequenced genomes fall into three taxonomic groups or phyla but there are over 60 phyla we know of." For the majority of them, however, there are no cultivated members available.

The team's findings fell into three main areas. The first was the discovery of unexpected metabolic features. They observed certain traits in archaea that previously only were seen in bacteria and vice-versa. The second contribution arising from the work was the correct reassignment, or binning, of data of some 340 million DNA fragments from other habitats to the proper lineage. This course correction provides insights into how organisms function in the context of a particular ecosystem as well as a much-improved and more accurate understanding of the associations of newly discovered genes with resident life forms. The third finding was the resolution of relationships within and between microbial phyla—the taxonomic ranking between domain and class—which led the team to propose two new superphyla, which are highly stable associations between phyla. Tonya Woyke, head of the DOE JGI Microbial Genomics and her colleagues are pursuing a more accurate characterization of these relationships so that they can better predict metabolic properties and other useful traits that can be expressed by different groups of microbes.

The rhizosphere describes the soil surrounding the plant root where microbe-plant interactions occur. These interactions can be critical in determining plant health and yield, and could be exploited for developing bioenergy feedstocks. To learn more about these influences, the DOE JGI initiated a Rhizosphere Grand Challenge project involving the model plant *Arabidopsis*. A similar project was initiated with maize.

In the study that appeared April 16, 2013, in *Proceedings of the National Academy of Sciences*, DOE JGI researchers were part of a team that characterized the rhizosphere bacterial diversity of maize plants in order to identify genes involved in plant-microbe interactions and how the environments might influence such interactions.

The team sequenced and analyzed rhizosphere microbial DNA taken from soil samples of maize fields in two climatic regions: Illinois and Missouri; and northeastern New York. They noted in their findings that despite the geography, the microbes from the Midwestern fields were less similar to one another; the Missouri soil microbes had greater similarity with the New York microbes. Other findings suggest that the microbial community around maize is strongly influenced by a few key genes rather than several alleles in concert.

At the phylum level, the number and diversity of unknown microbes still far outnumber those being studied. Metagenomics and single-cell genomics are tools that help researchers learn more about the "biological dark matter" that has not been cultivated and studied in the laboratory.

In an article published May 14, 2013, in *Nature Communications*, a team including DOE JGI researchers and collaborators used both techniques to assemble two nearly complete genomes belonging to the candidate phylum OP9, which was first discovered in the Obsidian Pool at Yellowstone National Park. The findings have led the team to suggest the phylum be called "Atribacteria."

The OP9 genomes were assembled from single cells collected at a hot spring in California and a metagenomic data set from cellulosic biomass communities being cultivated in Nevadan hot-spring sediments. Based on the team's analyses of the microbial genomes' cell structures and central metabolisms, they proposed naming the genomes "*Candidatus Caldatribacterium californiense*" and "*Ca. Caldatribacterium saccharofermentans*," respectively. Their findings revealed that OP9 may play a role in breaking down and using (hemi)cellulose.

"This study also illustrates the utility of combining single-cell and metagenomics approaches, even in cases where data sets originate in distinct species or environments," they said.

Cellulolytic communities were incubated in sediments at Great Boiling Spring, Nevada. A metagenome isolate from one site was used to generate an OP9 genome. *(Peacock JP et al., PLoSONE.doi:10.1371/journal.pone.0059927)*

Despite their miniscule size, cyanobacteria play significant roles in the global carbon and nitrogen cycles. Researchers are also looking at utilizing them for biofuel and biotechnology applications.

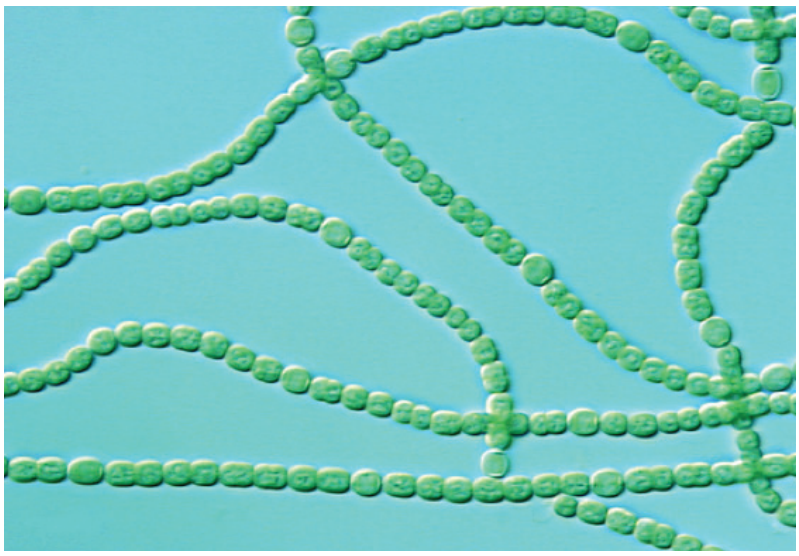
To learn more about the diverse strains in this phylum, the DOE JGI is leading an effort called CyanoGEBA to increase the number of available cyanobacterial genomes for study. The project's name is a nod to the DOE JGI's pioneering Genomic Encyclopedia of Bacteria and Archaea (GEBA) program, sequencing genomes from under-represented branches in the Tree of Life.

In an article published online December 31, 2012, in the *Proceedings of the National Academy of Sciences*, DOE JGI's Cheryl Kerfeld and her colleagues reported on a comparative analysis of several dozen CyanoGEBA genomes. A total of 54 cyanobacterial genomes were sequenced; 29 complete genomes were assembled while the rest are in draft status. The team's contributions have doubled the amount and diversity of the extant genomes.

"With the exponentially growing capacity for sequencing genomes it is becoming increasingly important to focus sequencing efforts so as to obtain a high-value return," the team wrote. "The extensive phylogenetically based survey … demonstrates the benefits gained from a more balanced representation of sequenced genomes within a phylum."

One of the questions the team hopes to answer through a continued increase of the phylogenetic diversity of cyanobacterial genomes is the identity of the closest relatives to the original lineage whose chloroplasts were engulfed in that first endosymbiotic event that resulted in the diverse range of photosynthetic organisms.

The DOE JGI's research horizons are broad as well as deep. In sediment buried in a contaminated aquifer in Colorado, a research team led by a DOE JGI collaborator used metagenomics to reconstruct a dominant organism and a member of a new phylum-level lineage. The analysis found bacteria and archaea from classes and orders previously unrecognized or unsampled. These findings are considered the "dark matter" of carbon and other biogeochemical cycles, and contribute significantly to our understanding of Earth's biogeochemistry.



Chains of cyanobacteria. *(James Golden, UC San Diego)*

The 1,450-mile-long Colorado River flows from the southwestern United States to northwest Mexico *(Wolfgang Staudt)*

Despite the efforts made to learn more about the microbial diversity in, on, and around the planet in the past decade, the microbes located below the Earth's surface remain difficult to characterize, in part due to their locations. However, these microbes are known to play significant roles in biogeochemical cycles. To help fill in the gaps of knowledge about these microbes, a team of researchers led by DOE JGI collaborator Jillian Banfield of UC Berkeley and Lawrence Berkeley National Laboratory sequenced samples from a contaminated aquifer at the Rifle Integrated Field Research Challenge site adjacent to the Colorado River in western Colorado. The two microbial communities found in the samples were studied as part of a 2012 DOE JGI Community Sequencing Program (now Community Science Program) project.

Through metagenomics, researchers reconstructed a dominant organism and member of a new phylum-level lineage from an aquifer sediment in Colorado.

As reported in the study published August 27, 2013, in *Nature Communications*, analyses of the data sets indicated that the subsurface microbial communities consisted of many bacteria and archaea from classes and orders not previously recognized or sampled. Additionally, the researchers were able to completely reconstruct the genome of a dominant organism called RBG-1 in a microbial community, one previously unknown and which turned out to be a member of a new phylum lineage.

"We document extraordinary microbial novelty and the importance of previously unknown lineages in sediment biogeochemical transformations," the researchers reported. "Many bacteria and archaea in these communities are novel at the phylum level or belong to phyla lacking a sequenced representative."

Analysis of the complete microbial genome led to a detailed metabolic model with evidence for multiple new enzymes and pathways. The findings serve to emphasize the unexplored diversity of subsurface microbes, considered by the researchers to be the "dark matter of the carbon and other biogeochemical cycles."

Computational Infrastructure

The DOE JGI's capacity as a next-generation genomics user facility has generated petabytes of data and analysis. In 2013, our genome sequence data alone consisted of 70 trillion nucleotides. In order to keep pace, the various portals available to access this information need to be robust and nimble. Over the past two years, DOE JGI has invested considerable time and energy to upgrading Genepool, the JGI's 8,000+ core computing cluster, as well as the DOE JGI's many Web services including Integrated Microbial Genomes (IMG); IMG's metagenome-focused counterpart, IMG/M; and the Genome Portal. The computing infrastructure and user interfaces have been enhanced to make data access faster and easier for the DOE JGI's user community.

In August, 2013, the JGI Archive and Metadata Organizer (JAMO) came online. This hierarchical data management system helps JGI staff locate and restore data or analysis in a matter of minutes instead of hours. JAMO began as a collaboration among the Sequence Data Management (SDM), Quality Assurance and Quality Control, and Genome Assembly groups at the DOE JGI. It has been a cross-program effort at the DOE JGI and cross-divisional effort between the DOE JGI and Lawrence Berkeley National Laboratory's National Energy Research Scientific Computing Center (NERSC).

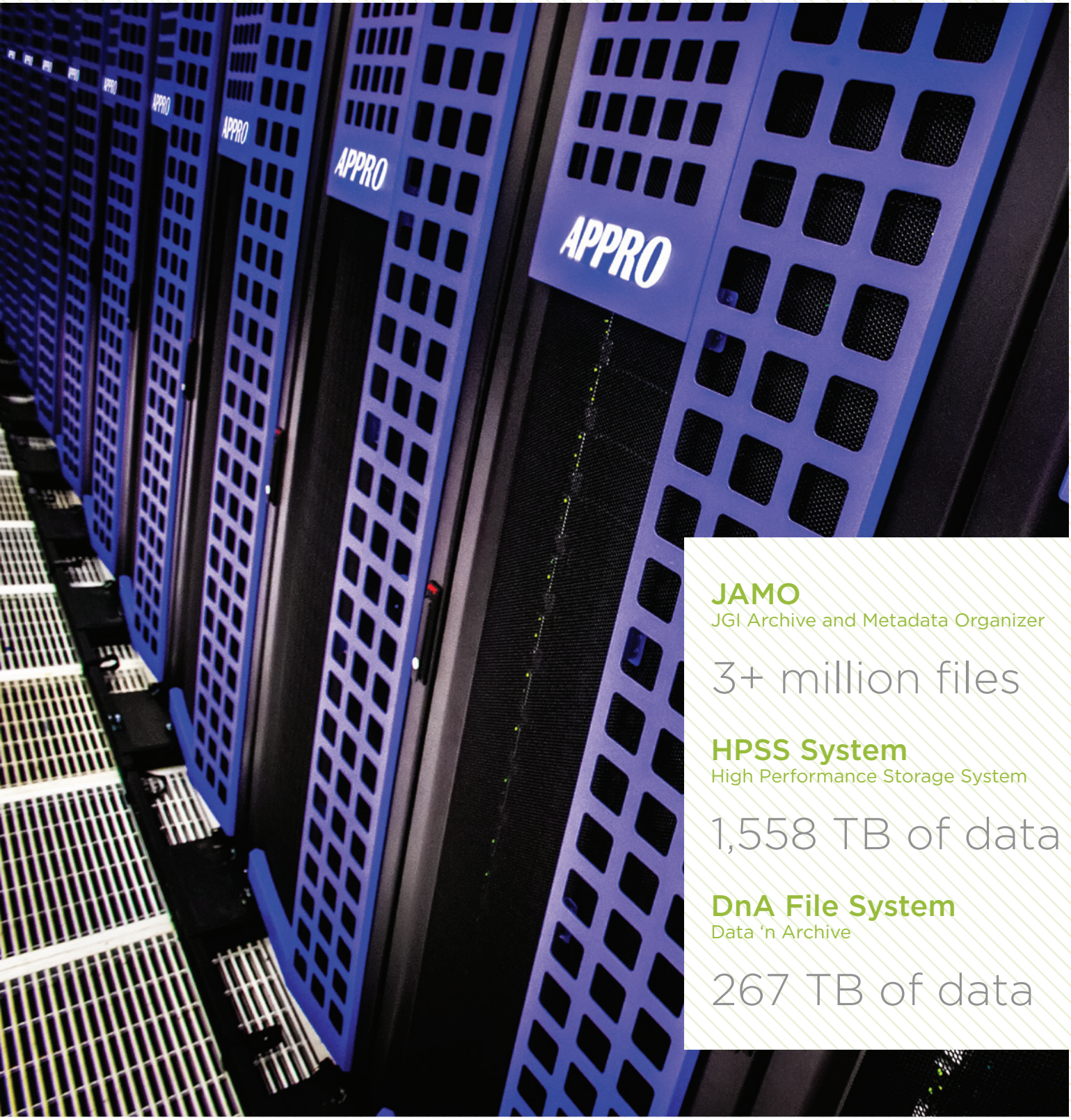
JAMO's architects, Chris Beecroft and Alex Boyd (SDM) of the DOE JGI, worked closely with NERSC's Kjiersten Fagnan (consulting), Nick Balthaser, and Wayne Hurlbert (Storage Systems).

"NERSC has a strong interest in supporting our user's data management needs, and the Archival Storage Team was pleased to help the SDM group innovate in their development of software to improve the management of JGI data," said Jason Hick, NERSC Storage Systems Group Lead.

"We are also looking forward to using the JAMO system to generate reproducible pipelines and work flows by storing detailed metadata with each project analysts complete," said Fagnan. "This project is extremely collaborative and generated a lot of positive momentum in software development at the JGI."

Also in 2013, the DOE JGI used several million CPU hours on NERSC's first petascale supercomputer, Hopper. These calculations could not have been completed on the Genepool cluster. DOE JGI's Rob Egan and Kristin Tennessen and their colleagues in the R&D and Omics groups made headway in using massively parallel resources to complete these large-scale calculations more efficiently. Other members of the DOE JGI staff were also involved in the early testing of bioinformatics algorithms like MPI-based metagenome assembly on NERSC's newest supercomputer, Edison. In 2014, the DOE JGI will expand these efforts through several high-performance computing initiatives.

The success of these computing projects is in part due to the DOE JGI's ongoing partnership with NERSC, one of the nation's foremost centers for high-performance computing. In 2010, all of the DOE JGI's computational resources were moved to NERSC and both sides have learned a great deal through this partnership. The infrastructure advancements to Genepool and other DOE JGI portals mean rapid and smooth access for users across the globe. Our partnership with NERSC enables the DOE JGI researchers and user community to devote more of their time to game-changing research.



JAMO
JGI Archive and Metadata Organizer

3+ million files

HPSS System
High Performance Storage System

1,558 TB of data

DnA File System
Data 'n Archive

267 TB of data

Appendices

Appendix A
Acronyms at a Glance

BER	DOE Office of Biological and Environmental Research
BERAC	Biological and Environmental Research Advisory Committee
BESC	BioEnergy Sequencing Center (at ORNL)
BRC	Bioenergy Research Center (i.e., BESC, GLBRC, JBEI)
CSP	Community Science Program
DOE	Department of Energy
EMSL	Environmental Molecular Sciences Laboratory (at PNNL)
ETOP	Emerging Technologies Opportunity Program
GEBA	Genomic Encyclopedia of Bacteria and Archaea
GLBRC	Great Lakes Bioenergy Research Center
GOLD	Genomes OnLine Database
HPC	High Performance Computing
HPPS	High Performance Storage System
IMG	Integrated Microbial Genomes data-management system
ISM	Integrated Safety Management
ITS	Integrated Tracking System
JAMO	JGI Archive and Metadata Organizer
JBEI	Joint BioEnergy Institute
LANL	Los Alamos National Laboratory
LBNL	Lawrence Berkeley National Laboratory
LLNL	Lawrence Livermore National Laboratory
NERSC	National Energy Research Scientific Computing Center
NREL	National Renewable Energy Laboratory
ORNL	Oak Ridge National Laboratory
PNNL	Pacific Northwestern National Laboratory
SAC	Scientific Advisory Committee

Appendix B
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: Aligning and merging fragments of a much longer DNA sequence in order to reconstruct the original sequence. This is required, as DNA sequencing technology cannot read whole genomes at once, but rather reads small pieces of between 20 and 1,000 bases, depending on the technology used.

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.

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.

Biomass: Material derived from living or recently living organisms, usually referring to plants or plant-derived material (lignocellulosic biomass). Biomass can serve as an energy source directly by burning or indirectly, after conversion into biofuels.

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

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

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.

Cellulose: An organic compound made of a long chain of several hundred to over 10,000 glucose units. It is a critical part of the cell wall of plants and many algae.

Cellulosic biofuel: A type of liquid transportation fuel produced from lignocellulose, a structural material that makes up much of the mass of plants. Lignocellulose is composed mainly of cellulose, hemicellulose, and lignin.

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.

Cyanobacteria: A phylum of bacteria that obtain their energy through photosynthesis, and named for the color of the bacteria. Although often called blue-green algae, that name is a misnomer as cyanobacteria are prokaryotic and algae are eukaryotic.

Draft genome: (also called a draft assembly) 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.

Hemicellulose: An organic compound that is part of most plant cell walls. Unlike cellulose, which is crystalline, strong, and resistant to being broken down, hemicellulose is much more fragile and has a random structure. As such, it is easier to break down than cellulose.

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.

Lignin: A complex polymer of aromatic alcohols known as monolignols, usually derived from wood. It is a critical part of the cell wall of plants and many algae.

Lignocellulosic biomass: Biomass derived from plants, the most abundant raw material for the production of biofuels.

Locus: (plural loci) The specific location of a gene or DNA sequence or position on a chromosome.

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.

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

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.

Peptide: Short chains of amino acids, the same compounds that make up proteins. Peptide chains are much shorter than the chains of amino acids that make up proteins.

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

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

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.

Selfing: Self-pollination or self-fertilization.

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. Also used as a verb to describe the process of determining the nucleotide order.

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.

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

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 C
2014 User Programs Supported Projects

Community Science Program Projects

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, Jeff	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)
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

PROPOSER	AFFILIATION	PROJECT DESCRIPTION
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)

JGI-EMSL Collaborative Science Initiative Projects

In 2013, the DOE JGI joined forces with the Environmental Molecular Sciences Laboratory (EMSL) at PNNL, another specialized user facility, to issue a joint call for proposals and have selected the first round of projects. EMSL, operated by the DOE Office of Biology and Environmental Research, provides scientific resources that directly complement and significantly expand our capacity to generate data illuminating the complex pathways of cellular function. The FY2014 JGI-EMSL project approvals reflect the synergy of the DOE JGI and EMSL's capabilities:

PROPOSER	AFFILIATION	PROJECT DESCRIPTION
Firestone, Mary Kathryn	UC Berkeley	Mapping soil carbon from cradle to grave: Using comparative transcriptomics, proteomics and metabolite analysis to identify the microbial blueprint for root-enhanced decomposition of organic matter
Hansel, Colleen	Woods Hole Oceanographic Institution	Genome-enabled investigations of the role of secreted proteins and reactive metabolites in carbon degradation by pure and mixed Ascomycete fungal communities
Harris, Steven	University of Nebraska-Lincoln	Engineering morphology and secretion to enhance the productivity of fungal fermentations
Hess, Matthias	Washington State University Tri-Cities	FECB: A Functional Encyclopedia of Cyanobacteria: Building the knowledge framework for an enhanced understanding of carbon and nitrogen cycling
Hofmockel, Kirsten	Argonne National Laboratory	Development of novel approaches to target microbial drivers of C cycling in soil aggregates
Kistler, Harold	USDA ARS Cereal Disease Laboratory	Organelles promoting high level terpenoid biosynthesis in filamentous fungi
O'Malley, Michelle Ann	UC Santa Barbara	Identification and regulation of cellulases within novel anaerobic gut fungi
Weyman, Philip	J. Craig Venter Institute	Functional genomics of moss-cyanobacteria interactions in boreal forest ecosystems

Emerging Technologies Opportunity Program (ETOP) Projects

The objectives of the ETOP are to identify and fund new and existing DOE JGI partners to develop promising projects focused on new technical capabilities that could be provided to users. Successful pilot-scale projects may be expanded as needed to meet future user demand. This will establish a process for ETOP partners to develop or provide specialized or advanced versions of needed capabilities, obviating the need for them to be developed at the DOE JGI.

PROPOSER	AFFILIATION	PROJECT DESCRIPTION
Banfield, Jillian Pan, Chongle Thomas, Brian	UC Berkeley/Lawrence Berkeley National Laboratory Oak Ridge National Laboratory UC Berkeley	Development of a pipeline for high-throughput recovery of near-complete and complete microbial genomes from complex metagenomic datasets
Quake, Stephen	Stanford University	New methods to isolate single cells for characterizing complex environmental samples
Magnuson, Jon	Pacific Northwest National Laboratory	Development and implementation of high-throughput methods for fungal culturing and nucleic acid isolation
Shendure, Jay	University of Washington	Accurate gene synthesis with tag-directed retrieval of sequence-verified DNA molecules
Stocker, Roman Wagner, Michael	MIT University of Vienna (Austria	Accelerated cell sorting by combining labeling with heavy water, Raman microspectroscopy, microfluidics and flow cytometry
Wing, Rod	Arizona Genomics Institute (AGI)	Generation of high-quality genomic DNA from plants and other organisms, large insert libraries and high-quality physical maps for improved physical map and sequence level-assemblies

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 Science 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*

Steve Briggs, *University of California*

Jeff Dangl, *University of North Carolina*

David Dooling, *Monsanto*

Joe Ecker, *Salk Institute*

Glenn Kubiak, *Lawrence Berkeley National Laboratory*

Nancy Moran, *Yale University*

Rick Myers, *HudsonAlpha Institute for Biotechnology*

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*

DOE Representatives

Dan Drell, *U.S. Department of Energy*

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*

Samuel Hazen, *UMass Amherst*

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*

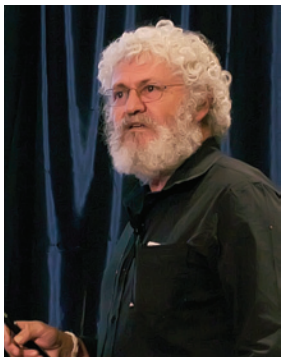
Ramunas Stepanauskas, *Bigelow Laboratory for Ocean Sciences*

Niels van der Lelie, *RTI*

Appendix E
2013 Genomics of Energy and Environment Meeting

Keynote Speakers

Chris Voigt, co-director of the Synthetic Biology Center at the Massachusetts Institute of Technology, delivered the opening keynote, exploring how the DOE JGI might play a key role in the field of synthetic biology. His talk focused on two big ideas: the intersection of sequencing, synthesis, and design; and optimizing the cycle of designing, debugging, and correcting. He noted that, "If we're going to scale up, we need to take a fundamentally different approach to design."



Eric Karsenti from the European Molecular Biology Laboratory, Heidelberg, closed the annual meeting with a look back at the three-year TARA Oceans expedition he mounted. The scientific goal was to learn more about the plankton gene network by sampling the major oceans in order to characterize the environments. The second objective was as a floating science education and outreach center to the general public. When at port, the research team's 36-foot schooner was opened to the public and scientists were available to answer questions. According to the expedition statistics, 5,000 children boarded the boat over the course of the voyage.

Other Featured Speakers

(in order of appearance):

Eric Allen, *University of California at San Diego*

Eoin Brodie, *Lawrence Berkeley National Laboratory, Earth Sciences Division*

Joe DeRisi, *University of California at San Francisco*

Laura A. Hug, *University of California at Berkeley*

Wayne Reeve, *Murdoch University*

Eldredge Bermingham, *Smithsonian Tropical Research Institute*

Christopher W. Schadt, *Oak Ridge National Laboratory*

Jeremy Schmutz, *HudsonAlpha Institute for Biotechnology*

Sam Hazen, *University of Massachusetts*

Rick Amasino, *University of Wisconsin*

Sean P. Gordon, *U.S. Department of Agriculture*

Zander Myburg, *University of Pretoria*

Jose Pruneda-Paz, *University of California at San Diego*

Len Pennacchio, *DOE Joint Genome Institute*

David Weston, *Oak Ridge National Laboratory*

Tanja Woyke, *DOE Joint Genome Institute*

Paul Blainey, *Broad Institute*

Sarah L. Lebeis, *University of North Carolina*

Tom Gilbert, *University of Copenhagen*

Greg Bell, *DOE ESnet*, *Lawrence Berkeley National Laboratory, Scientific Networking Division*

Peter E. Larsen, *Argonne National Laboratory*

Bob Schmidt, *SG Biofuels*; *University of CaliforniaSan Diego*

Jack Newman, *Amyris Inc.*

Adam Guss, *Oak Ridge National Laboratory*

Sam Deutsch, *DOE Joint Genome Institute*

Jane Lau, *Lawrence Berkeley National Laboratory*

Learn more about the meeting talks at
http://www.jgi.doe.gov/meetings/usermeeting/2013/agenda.html

Videos of the talks are available on DOE JGI's YouTube channel at
http://bit.ly/JGI13_UM8 and on our SciVee channel at http://www.scivee.tv/node/58289

Appendix F
2013 Publications

October 2012 –September 2013

Abt B et al. Genome sequence of the thermophilic fresh-water bacterium *Spirochaeta caldaria* type strain (H1T), reclassification of *Spirochaeta caldaria* and *Spirochaeta stenostrepta* in the genus Treponema as ... *Stand Genomic Sci.* 2013;8(1). doi: 10.4056/sigs.3096473. Epub 2013 Apr 15.

Aklujkar M et al. The genome of *Pelobacter carbinolicus* reveals surprising metabolic capabilities and physiological features. *BMC Genomics*. 2012 Dec 10; 13:690. doi: 10.1186/1471-2164-13-690.

Allers E et al. Diversity and population structure of Marine Group A bacteria in the northeast subarctic Pacific Ocean. *The ISME Journal* 2013;7(2):256–268. doi:10.1038/ismej.2012.108. Epub 15 November 2012.

Anderson IJ et al. Genome sequence of the flexirubin-pigmented soil bacterium *Niabella soli* type strain (JS13-8T). *Stand. Genomic Sci.* 4 Dec 2012;7(2). doi:10.4056/sigs.3117229

Appel AM et al. Frontiers, opportunities, and challenges in biochemical and chemical catalysis of CO_2 fixation. *Chem Rev.* 2013 Jun 14. [Epub ahead of print.]

Aylward FO et al. Comparison of 26 sphingomonad genomes reveals diverse environmental adaptations and biodegradative capabilities. *Appl Environ Microbiol.* 2013 Apr 5. [Epub ahead of print.]

Aylward FO et al. Complete genome of *Serratia* sp. Strain FGI 94, a strain associated with leaf-cutter ant fungus gardens. *Genome Announc.* 2013 Mar 14;1(2):e0023912. doi: 10.1128/genomeA.00239-12

Aylward FO et al. *Leucoagaricus gongylophorus* produces a diversity of enzymes for recalcitrant plant polymer degradation in leaf-cutter ant fungus gardens. *Appl Environ Microbiol.* 2013 Jun;79(12):3770-8. doi: 10.1128/AEM.03833-12. 2013 Apr 12.

Aylward FO et al. Complete genome of *Enterobacteriaceae* bacterium Strain FGI 57, a strain associated with leaf-cutter ant fungus gardens. *Genome Announc.* 2013 Jan;1(1). doi:pii: e00238-12. 10.1128/genomeA.00238-12. Epub 2013 Feb 21.

Baran R et al. Functional genomics of novel secondary metabolites from diverse Cyanobacteria using untargeted metabolomics. *Mar. Drugs.* 2013 Sep 30. 11(10):3617-3631. doi:10.3390/md11103617

Beck DA et al. A metagenomic insight into freshwater methane-utilizing communities and evidence for cooperation between the *Methylococcaceae* and the *Methylophilaceae*. *PeerJ.* 2013 Feb 19;1:e23. doi:10.7717/peerj.23. Print 2013.

Bendall ML et al. Exploring the roles of DNA methylation in the metal-reducing bacterium *Shewanella oneidensis* MR-1. *J Bacteriol.* 2013 Aug 30. [Epub ahead of print.]

Beyersmann P et al. Genome sequence of *Phaeobacter caeruleus* type strain (DSM 24564T), a surface-associated member of the marine *Roseobacter* clade. *Stand Genomic Sci.* 2013;8(3). doi:10.4056/sigs.3927623. Epub 2013 Jul 30.

Binder M et al. Phylogenetic and phylogenomic overview of the Polyporales. *Mycologia*. 2013 Aug 11. [Epub ahead of print.]

Bollmann A et al. Complete genome sequence of *Nitrosomonas* sp. Is79 – an ammonia oxidizing bacterium adapted to low ammonium concentrations. *Stand Genomic Sci.* 2013 Feb 25; 7(3). doi:10.4056/sigs.3517166.

Bradnam KR et al. Assemblathon 2: Evaluating de novo methods of genome assembly in three vertebrate species. *Gigascience.* 2013 Jul 22;2(1):10. [Epub ahead of print.]

Brown SD et al. Genome sequences of industrially relevant *Saccharomyces cerevisiae* Strain M3707, isolated from a sample of distillers yeast and four haploid derivatives. *Genome Announc.* 2013 Jun 27;1(3). doi:pii: e00323-13. 10.1128/genomeA.00323-13.

Burow LC et al. Anoxic carbon flux in photosynthetic microbial mats as revealed by metatranscriptomics. *ISME J.* 2012 Nov 29. doi: 10.1038/ismej.2012.150. [Epub ahead of print.]

Cai F et al. Evidence for the widespread distribution of CRISPR-Cas system in the phylum Cyanobacteria. *RNA Biol.* 2013 May 1;10(5):687-93. doi: 10.4161/rna.24571. Epub 2013 Apr 11.

Cai F et al. The structure of CcmP, a tandem bacterial microcompartment domain protein from the β-carboxysome, forms a subcompartment within a microcompartment. *J Biol Chem.* 2013 May 31;288(22):16055-63. doi: 10.1074/jbc.M113.456897. Epub 2013 Apr 9.

Campbell JH et al. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *Proc Natl Acad Sci U S A.* 2013 Apr 2;110(14):5540-5. doi: 10.1073/pnas.1303090110. Epub 2013 Mar 18.

Castelle CJ et al. Extraordinary phylogenetic diversity and metabolic versatility in aquifer sediment. *Nat Commun.* 2013 Aug 27;4:2120. doi: 10.1038/ncomms3120.

Challacombe JF et al. Complete genome sequence of *Halorhodospira halophila* SL[1]. *Stand Genomic Sci.* 2013;8(2). doi:10.4056/sigs.3677284. Epub 2013 Apr 15.

Chen IM et al. Improving microbial genome annotations in an integrated database context. *PLoS One.* 2013;8(2):e54859. doi: 10.1371/journal.pone.0054859. Epub 2013 Feb 12.

Chin CS et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods.* 2013 Jun;10(6):563-9. doi: 10.1038/nmeth.2474. Epub 2013 May 5.

Cho Y et al. Fungal-specific transcription factor AbPf2 activates pathogenicity in *Alternaria brassicicola*. *Plant J.* 2013 Aug;75(3):498-514. doi: 10.1111/tpj.12217. Epub 2013 May 24.

Cho Y et al. Transcription factor *Amr1* induces melanin biosynthesis and suppresses virulence in *Alternaria brassicicola*. *PLoS Pathog.* 2012;8(10):e1002974. doi: 10.1371/journal.ppat.1002974. Epub 2012 Oct 25.

Cipriano MJ et al. RegTransBase - a database of regulatory sequences and interactions based on literature: a resource for investigating transcriptional regulation in prokaryotes. *BMC Genomics.* 2013 Apr 2;14:213. doi: 10.1186/1471-2164-14-213.

Clark SC et al. ALE: A generic assembly likelihood evaluation framework for assessing the accuracy of genome and metagenome assemblies. *Bioinformatics.* 2013 Jan 21. [Epub ahead of print.]

Clingenpeel S K et al. Yellowstone lake nanoarchaeota. *Front Microbiol.* 4:274. doi:10.3389/fmicb.2013.00274 (2013). Epub 2013 Sept 26.

Cole JK et al. Sediment microbial communities in Great Boiling Spring are controlled by temperature and distinct from water communities. *ISME J.* 2013 Apr;7(4):718-29. doi: 10.1038/ismej.2012.157.

Condon BJ et al. Comparative genome structure, secondary metabolite, and effector coding capacity across *Cochliobolus* pathogens. *PLoS Genet.* 2013 Jan;9(1):e1003233. doi: 10.1371/journal.pgen.1003233. Epub 2013 Jan 24.

Curtis BA et al. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature.* 2012 Dec 6;492(7427):59-65. doi: 10.1038/nature11681. Epub 2012 Nov 28.

D'haeseleer P et al. Proteogenomic analysis of a thermophilic bacterial consortium adapted to deconstruct switchgrass. *PLoS One.* 2013 Jul 19;8(7):e68465. doi: 10.1371/journal.pone.0068465. Print 2013 Jul 22.

Davis JR et al. Genome sequence of *Streptomyces viridosporus* Strain T7A ATCC 39115, a lignin-degrading Actinomycete. *Genome Announc.* 2013 Jul 5;1(4). pii: e00416-13. doi: 10.1128/genomeA.00416-13.

de Wit PJ et al. The genomes of the fungal plant pathogens *Cladosporium fulvum* and *Dothistroma septosporum* reveal adaptation to different hosts and lifestyles but also signatures of common ancestry. *PLoS Genet.* 2012 Nov;8(11):e1003088. doi:10.1371/journal.pgen.1003088. Epub 2012 Nov 29.

Demaere MZ et al. High level of intergenera gene exchange shapes the evolution of haloarchaea in an isolated Antarctic lake. *Proc Natl Acad Sci U S A.* 2013 Sep 30. [Epub ahead of print.]

DeAngelis KM et al. Metagenomes of tropical soil-derived anaerobic switchgrass-adapted consortia with and without iron. *Stand. Genomic Sci.* 2013 Feb 25;7(3). doi:10.4056/sigs.3377516.

Deng L et al. Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria, as revealed by viral tagging. *MBio.* 2012 Oct 30;3(6). doi:pii: e00373-12. 10.1128/*MBio.*00373-12.

Desiniotis A et al. Complete genome sequence of the ethanol-producing *Zymomonas mobilis* subsp. *mobilis* Centrotype ATCC 29191. *J Bacteriol.* 2012 Nov;194(21):5966-7. doi: 10.1128/JB.01398-12.

Dodsworth JA et al. Single-cell and metagenomic analyses indicate a fermentative and saccharolytic lifestyle for members of the OP9 lineage. *Nat Commun.* 2013;4:1854. doi: 10.1038/ncomms2884.

Dubchak I et al. Whole-genome rVISTA: A tool to determine enrichment of transcription factor binding sites in gene promoters from transcriptomic data. *Bioinformatics.* 2013 Aug 15;29(16):2059-61. doi: 10.1093/bioinformatics/btt318. Epub 2013 Jun 4.

Elkins JG et al. Complete genome sequence of the hyperthermophilic sulfate-reducing bacterium *Thermodesulfobacterium geofontis* OPF15T. *Genome Announc.* 2013 Apr 11;1(2). doi:pii: e00162-13. 10.1128/genomeA.00162-13.

Emerson D et al. Comparative genomics of freshwater Fe-oxidizing bacteria: Implications for physiology, ecology, and systematics. *Front Microbiol.* 2013 Sep 12;4:254. doi: 10.3389/fmicb.2013.00254.

Everroad RC et al. Draft genome sequence of an oscillatorian cyanobacterium, Strain ESFC-1. *Genome Announc.* 2013 Aug 1;1(4). doi:pii: e00527-13. 10.1128/genomeA.00527-13.

Flowers JJ et al. Comparative genomics of two 'Candidatus Accumulibacter' clades performing biological phosphorus removal. *ISME J.* 2013 Jul 25. doi: 10.1038/ismej.2013.117. [Epub ahead of print.]

Freese HM et al. Genome sequence of the phage-gene rich marine *Phaeobacter arcticus* type Strain DSM 23566T. *Stand Genomic Sci.* 2013;8(3). doi:10.4056/sigs.383362. Epub 2013 Jul 30.

Geraldes A et al. A 34K SNP genotyping array for *Populus trichocarpa*: Design, application to the study of natural populations and transferability to other *Populus* species. *Mol Ecol Resour.* 2013 Mar;13(2):306-23. doi: 10.1111/1755-0998.12056. Epub 2013 Jan 11.

Ghodhbane-Gtari F et al. Draft genome sequence of *Frankia* sp. strain CN3, an atypical, noninfective (Nod-) ineffective (Fix-) isolate from *Coriaria nepalensis*. *Genome Announc.* 2013 Mar 14;1(2):e0008513. doi: 10.1128/genomeA.00085-13.

Gobler CJ et al. The central role of selenium in the biochemistry and ecology of the harmful pelagophyte, *Aureococcus anophagefferens*. *ISME J.* 2013 Mar 7. doi: 10.1038/ismej.2013.25. [Epub ahead of print.]

Göker M T et al. Genome sequence of *Phaeobacter daeponensis* type strain (DSM 23529 T), a facultatively anaerobic bacterium isolated from marine sediment, and emendation of *Phaeobacter daeponensis*. *Stand Genomic Sci.* 2013 Sept. Epub 2013 Sept.

Gross SM et al. *De novo* transcriptome assembly of drought tolerant CAM plants, *Agave deserti* and *Agave tequilana*. *BMC Genomics.* 2013 Aug 19;14(1):563. [Epub ahead of print.]

Harris JK et al. Phylogenetic stratigraphy in the Guerrero Negro hypersaline microbial mat. *ISME J.* 2013 Jan;7(1):50-60. doi: 10.1038/ismej.2012.79.

Haudry A et al. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nature*. 2013 Aug;45(8):891-8. doi: 10.1038/ng.2684. Epub 2013 Jun 30.

He S et al. Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PLoS One.* 2013 Apr 12;8(4):e61126. doi:10.1371/journal.pone.0061126. Print 2013.

Hemschemeier A et al. Hypoxic survival requires a 2-on-2 hemoglobin in a process involving nitric oxide. *Proc Natl Acad Sci*. doi: 10.1073/pnas.1302592110. Epub 2013 June 10.

Hirsch AM et al. Complete genome sequence of Micromonospora Strain L5, a potential plant-growth-regulating Actinomycete, originally isolated from *Casuarina equisetifolia* root nodules. *Genome Announc.* 2013 Sep 26;1(5). doi:pii: e00759-13. 10.1128/genomeA.00759-13.

Hunt DE et al. The relationship between abundance and specific activity of bacterioplankton in open ocean surface waters. *Appl Environ Microbiol.* 2013 Jan;79(1):177-84. doi: 10.1128/AEM.02155-12. Epub 2012 Oct 19.

Huntemann M et al. Genome sequence of the phylogenetically isolated spirochete *Leptonema illini* type strain (3055T). *Stand Genomic Sci.* 2013;8(2). doi:10.4056/sigs.3637201. Epub 2013 May 25.

Hurwitz B et al. Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biology* 2013;14:R123 doi:10.1186/gb-2013-14-11-r123.

Inskeep WP et al. Phylogenetic and functional analysis of metagenome sequence from high-temperature archaeal habitats demonstrate linkages between metabolic potential and geochemistry. *Front Microbiol.* 2013 May 15;4:95. doi: 10.3389/fmicb.2013.00095. Print 2013.

Inskeep WP et al. The YNP Metagenome Project: Environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. *Front Microbiol.* 2013 May 6;4:67. doi: 10.3389/fmicb.2013.00067. Print 2013.

International Peach Genome Initiative et al. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet.* 2013 Apr 26;45(5):487-94. doi: 10.1038/ng.2586. Epub 2013 Mar 24.

Iwata A et al. Identification and characterization of functional centromeres of the common bean. *Plant J.* 2013 Oct;76(1):47-60. doi: 10.1111/tpj.12269. Epub 2013 Aug 5.

Kamke J et al. Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *ISME J.* 2013 Jul 11. doi: 10.1038/ismej.2013.111. [Epub ahead of print.]

Kappler U et al. Complete genome sequence of the facultatively chemolithoautotrophic and methylotrophic alpha Proteobacterium *Starkeya novella* type strain (ATCC 8093(T)). *Stand Genomic Sci.* 2012 Oct 10;7(1):44-58. doi: 10.4056/sigs.3006378.

Kazakov AE et al. Transcription factor family-based reconstruction of singleton regulons: Study of the Crp/Fnr, ArsR and GntR families in *Desulfovibrionales* genomes. *J Bacteriol.* 2013 Jan;195(1):29-38. doi: 10.1128/JB.01977-12. Epub 2012 Oct 19.

Kaznadzey A et al. PSimScan: Algorithm and utility for fast protein similarity search. *PLoS One.* 2013;8(3):e58505. doi: 10.1371/journal.pone.0058505. Epub 2013 Mar 7.

Kerfeld Cheryl et al. Structural, mechanistic and genomic insights into OCP-mediated photoprotection. *Genomics of Cyanobacteria: Advances in Botanical Research*. 2013;65:1-26.

Kerfeld CA. Introduction: Sequences and consequences. *Biochem Mol Biol Educ*. 2013 Jan-Feb;41(1):12-15. doi: 10.1002/bmb.20660.

Kirilovsky D et al. The orange carotenoid protein: A blue-green light photoactive protein. *Photochem Photobiol Sci.* 2013 Feb 8. [Epub ahead of print.]

Kits KD et al. Genome sequence of the obligate gammaproteobacterial methanotroph *Methylomicrobium album* Strain BG8. *Genome Announc.* 2013 Apr 11;1(2):e0017013. doi: 10.1128/genomeA.00170-13.

Klatt CG et al. Community structure and function of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal environments. *Front Microbiol.* 2013 Jun 03;4:106. doi: 10.3389/fmicb.2013.00106. Print 2013.

Kontur, WS et al. Revised sequence and annotation of the Rhodobacter sphaeroides 2.4.1 genome. *J Bacteriol.* 2012 Dec;194(24):7016-7. doi: 10.1128/JB.01214-12.

Kozubal MA et al. Geoarchaeota: A new candidate phylum in the Archaea from high-temperature acidic iron mats in Yellowstone National Park. *ISME J.* 2012 Nov 15. doi: 10.1038/ismej.2012.132. [Epub ahead of print.]

Kruse T et al. Complete genome sequence of *Dehalobacter restrictus* PER-K23T. *Stand Genomic Sci.* 2013;8(3). doi:10.4056/sigs.3787426. Epub 2013 Jul 30.

Lamour KH 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.

Liolios K et al. Complete genome sequence of the halophilic bacterium *Spirochaeta africana* type strain (Z-7692^{T}) from the alkaline Lake Magadi in the East African Rift. *Stand Genomic Sci.* 2013;8(2). doi:10.4056/sigs.3607108. Epub 2013 May 20.

Lorenz, WW et al. Conifer DBMagic: A database housing multiple de novo transcriptome assemblies for 12 diverse conifer species. *Tree Genetics & Genomes* December 2012; 8,(6):1477-1485

Lucas S et al. Complete genome sequence of the thermophilic, piezophilic, heterotrophic bacterium *Marinitoga piezophila KA3*. *J Bacteriol.* 2012 Nov;194(21):5974-5. doi: 10.1128/JB.01430-12. PubMed PMID: 23045491.

Manning VA et al. Comparative genomics of a plant-pathogenic fungus, *Pyrenophora tritici-repentis*, reveals transduplication and the impact of repeat elements on pathogenicity and population divergence. *G3* (Bethesda). 2013 Jan;3(1):41-63. doi: 10.1534/g3.112.004044. Epub 2013 Jan 1.

Mascher M et al. Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J.* 2013 Sep 2. doi: 10.1111/tpj.12319. [Epub ahead of print.]

Mavromatis K et al. The fast changing landscape of sequencing technologies and their impact on microbial genome assemblies and annotation. *PLoS One.* 2012;7(12):e48837. doi: 10.1371/journal.pone.0048837. Epub 2012 Dec 12.

Mavromatis K et al. Complete genome sequence of the moderate thermophile *Anaerobaculum* mobile type strain (NGA^T). *Stand Genomic Sci.* 2013;8(1). doi:10.4056/sigs.3547050. Epub 2013 Apr 15.

Mavromatis K et al. Complete genome sequence of the bile-resistant pigment-producing anaerobe *Alistipes finegoldii* type strain (AHN2437(T)). *Stand Genomic Sci.* 2013 Apr 15;8(1):26-36. doi: 10.4056/sigs.3527032.

Meier-Kolthoff, JP et al. Genome sequence of the chemoheterotrophic soil bacterium *Saccharomonospora cyanea* type strain (NA-134(T)). *Stand Genomic Sci.* 2013;9:28-41. doi:10.4056/sigs.4207886.

Morin E et al. Genome sequence of the button mushroom *Agaricus bisporus* reveals mechanisms governing adaptation to a humic-rich ecological niche. *Proc Natl Acad Sci U S A.* 2012 Oct 23;109(43):17501-6. doi: 10.1073/pnas.1206847109. Epub 2012 Oct 8.

Murugapiran SK et al. *Thermus oshimai* JL-2 and *T. thermophilus* JL-18 genome analysis illuminates pathways for carbon, nitrogen, and sulfur cycling. *Stand Genomic Sci.* 2013 Feb 25; 7(3). doi:10.4056/sigs.3667269.

Nakatsu CH et al. Complete genome sequence of *Arthrobacter* sp strain FB24. *Stand Genomic Sci.* 2013;9:106-116. doi:10.4056/sigs.4438185.

Neupane S et al. Non-contiguous finished genome sequence of plant-growth promoting *Serratia proteamaculans* S4. *Stand Genomic Sci.* 2013;8(3). doi:10.4056/sigs.3927623. Epub 2013 Jul 30.

Nordberg H et al. BioPig: A Hadoop-based analytic toolkit for large-scale sequence data. *Bioinformatics.* 2013 Sep 10. [Epub ahead of print.]

Nouioui I et al. Draft genome sequence of Frankia sp. Strain BMG5.12, a nitrogen-fixing Actinobacterium isolated from Tunisian soils. *Genome Announc.* 2013 Jul 11;1(4). pii: e00468-13. doi: 10.1128/genomeA.00468-13.

Ohm RA et al. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *PLoS Pathog.* 2012 Dec;8(12):e1003037. doi: 10.1371/journal.ppat.1003037. Epub 2012 Dec 6.

Palaniappan K M-K et al. Genome sequence of the moderately thermophilic sulfur-reducing bacterium *Thermanaerovibrio velox* type strain (Z-9701 T) and emended description of the genus Thermanaerovibrio. *Stand Genomic Sci.* 2013;9(1). Epub 2013 Sept.

Paterson AH et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature*. 2012 Dec 20;492(7429):423-7. doi: 10.1038/nature11798.

Peacock JP et al. Pyrosequencing reveals high-temperature cellulolytic microbial consortia in Great Boiling Spring after *in situ* lignocellulose enrichment. *PLoS One.* 2013;8(3):e59927. doi: 10.1371/journal.pone.0059927. Epub 2013 Mar 29.

Peiffer JA, Spor A, Koren O, Jin Z, Tringe SG, Dangl JL, Buckler ES, Ley RE. Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proc Natl Acad Sci U S A.* 2013 Apr 16;110(16):6548-53. doi: 10.1073/pnas.1302837110. Epub 2013 Apr 1.

Penz T et al. Comparative genomics suggests an independent origin of cytoplasmic incompatibility in *Cardinium hertigii*. *PLoS Genet.* 2012 Oct;8(10):e1003012. doi: 10.1371/journal.pgen.1003012. Epub 2012 Oct 25.

Pester M et al. Complete genome sequences of *Desulfosporosinus orientis* DSM765T, *Desulfosporosinus youngiae* DSM17734T, *Desulfosporosinus meridiei* DSM13257T, and *Desulfosporosinus acidiphilus* DSM22704T. *J Bacteriol.* 2012 Nov;194(22):6300-1. doi: 10.1128/JB.01392-12.

Polívka T et al. Carotenoid-protein interaction alters the S(1) energy of hydroxyechinenone in the orange carotenoid protein. *Biochim Biophys Acta*. 2012 Oct 16;1827(3):248-254. doi: 10.1016/j.bbabio.2012.10.005. [Epub ahead of print.]

Rajeev L et al. Dynamic cyanobacterial response to hydration and dehydration in a desert biological soil crust. *ISME J.* 2013 Jun 6. doi: 10.1038/ismej.2013.83. [Epub ahead of print.]

Read B et al. Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature*. 2013 Jul 11;499(7457):209-13. doi: 10.1038/nature12221. Epub 2013 Jun 12.

Riedel T et al. Genome sequence of the orange-pigmented seawater bacterium *Owenweeksia hongkongensis* type strain (UST20020801(T)). *Stand Genomic Sci.* 2012 Oct 10;7(1):120-30. doi: 10.4056/sigs.3296896.

Riedel T et al. Genome sequence of the Antarctic rhodopsins-containing flavobacterium *Gillisia limnaea* type strain (R-8282(T)). *Stand Genomic Sci.* 2012 Oct 10;7(1):107-19. doi: 10.4056/sigs.3216895.

Riedel T et al. Genome sequence of the *Leisingera aquimarina* type strain (DSM 24565^T), a member of the marine *Roseobacter* clade rich in extrachromosomal elements. *Stand Genomic Sci.* 2013 Jul 30; 8(3). doi:10.4056/sigs.3858183

Rinke C et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature*. 2013 Jul 25;499(7459):431-7. doi: 10.1038/nature12352. Epub 2013 Jul 14.

Rivers AR et al. Transcriptional response of bathypelagic marine bacterioplankton to the Deepwater Horizon oil spill. *ISME J.* 2013 Aug 1. doi: 10.1038/ismej.2013.129. [Epub ahead of print.]

Romano C et al. Comparative genomic analysis of phylogenetically closely related *Hydrogenobaculum* sp. isolates from Yellowstone National Park. *Appl Environ Microbiol.* 2013 May;79(9):2932-43. doi: 10.1128/AEM.03591-12. Epub 2013 Feb 22.

Scully ED et al. Metagenomic profiling reveals lignocellulose degrading system in a microbial community associated with a wood-feeding beetle. *PLoS One.* 2013 Sep 4;8(9):e73827. doi: 10.1371/journal.pone.0073827.

Sen A et al. Draft genome sequence of *Frankia* sp. strain QA3, a nitrogen-fixing actinobacterium isolated from the root nodule of *Alnus nitida*. *Genome Announc.* 2013 Mar 21;1(2):e0010313. doi: 10.1128/genomeA.00103-13.

Sentchilo V et al. Community-wide plasmid gene mobilization and selection. *ISME J.* 2013 Jun;7(6):1173-86. doi: 10.1038/ismej.2013.13. Epub 2013 Feb 14.

Shi W et al. Comparative genomic analysis of the endosymbionts of herbivorous insects reveals eco-environmental adaptations: biotechnology applications. *PLoS Genet.* 2013 Jan;9(1):e1003131. doi: 10.1371/journal.pgen.1003131. Epub 2013 Jan 10.

Shih PM et al. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci U S A.* 2013 Jan 15;110(3):1053-8. doi: 10.1073/pnas.1217107110. Epub 2012 Dec 31.

Simakov O et al. Insights into bilaterian evolution from three spiralian genomes. *Nature*. 2013 Jan 24;493(7433):526-31. doi: 10.1038/nature11696. Epub 2012 Dec 19.

Slotte T et al. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat Genet.* 2013 Jul;45(7):831-5. doi: 10.1038/ng.2669. Epub 2013 Jun 9.

Spring S et al. Complete genome sequence of the sulfate-reducing firmicute *Desulfotomaculum ruminis* type strain (DL(T)). *Stand Genomic Sci.* 2012 Dec 19;7(2):304-19. doi: 10.4056/sigs.3226659. Epub 2012 Dec 11.

Stackebrandt E et al. Complete genome sequence of *Coriobacterium glomerans* type strain (PW2^T) from the midgut of *Pyrrhocoris apterus L*. (red soldier bug). *Stand Genomic Sci.* 2013;8(1). doi:10.4056/sigs.3507020. Epub 2013 Apr 15.

Stackebrandt E et al. Complete genome sequence of the free-living aerobic spirochete *Turneriella parva* type strain (H^T). *Stand Genomic Sci.* 2013;8(2). doi:10.4056/sigs.3617113. Epub 2013 May 25.

Stackebrandt E et al. High-quality-draft genome sequence of the yellow-pigmented flavobacterium *Joostella marina* type strain (En5^T). *Stand Genomic Sci.* 2013;8(1). doi:10.4056/sigs.3537045. Epub 2013 Apr 15.

Stanley DN et al. Comparative Analysis of 126 cyanobacterial genomes reveals evidence of functional diversity among homologs of the redox-regulated CP12 protein. *Plant Physiol*. 2013 Feb;161(2):824-35. doi: 10.1104/pp.112.210542. Epub 2012 Nov 26.

Susanti D et al. Complete genome sequence of *Desulfurococcus fermentans*, a hyperthermophilic cellulolytic crenarchaeon isolated from a freshwater hot spring in Kamchatka, Russia. *J Bacteriol.* 2012 Oct;194(20):5703-4.

Sutter M et al. Two new high-resolution crystal structures of carboxysome pentamer proteins reveal high structural conservation of CcmL orthologs among distantly related cyanobacterial species. *Photosynth Res*. 2013 Aug 15. [Epub ahead of print.]

Sutter M et al. Crystal structure of the FRP and identification of the active site for modulation of OCP-mediated photoprotection in cyanobacteria. *Proc Natl Acad Sci U S A.* 2013 Jun 11;110(24):10022-7. doi: 10.1073/pnas.1303673110. Epub 2013 May 28.

Swan BK et al. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc Natl Acad Sci U S A.* 2013 Jul 9;110(28):11463-8. doi: 10.1073/pnas.1304246110. Epub 2013 Jun 25.

Takacs-Vesbach CD et al. Metagenome Sequence Analysis of filamentous microbial communities obtained from geochemically distinct geothermal channels reveals specialization of three *Aquificales* lineages. *Front Microbiol.* 2013 May 29;4:84. doi: 10.3389/fmicb.2013.00084. Print 2013.

Tamano K et al. Increased production of fatty acids and triglycerides in *Aspergillus oryzae* by enhancing expressions of fatty acid synthesis-related genes. *Appl Microbiol Biotechnol*. 2013 Jan;97(1):269-81. doi: 10.1007/s00253-012-4193-y. Epub 2012 Jun 26.

Teshima H et al. Genome sequence of *Frateuria aurantia* type strain (Kondô 67 T), a xanthomonade isolated from *Lilium auratium* Lindl. *Stand Genomic Sci.* 2013;9(1). Epub 2013 Sept.

Tripp HJ. The unique metabolism of SAR11 aquatic bacteria. *J Microbiol.* 2013 Apr;51(2): 147-53. doi: 10.1007/s12275-013-2671-2. Epub 2013 Apr 27.

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*. 2013 May;15(5):1275-89. doi: 10.1111/j.1462-2920.2012.02774.x. Epub 2012 May 9.

Visser M et al. Genome analysis of *Desulfotomaculum kuznetsovii* strain 17^{T} reveals a physiological similarity with *Pelotomaculum thermopropionicum* SI^{T}. *Stand Genomic Sci.* 2013;8(1). doi:10.4056/sigs.3627141. Epub 2013 Apr 15.

Wall LG et al. Draft genome sequence of *Frankia* sp. Strain BCU110501, a nitrogen-fixing Actinobacterium isolated from nodules of *Discaria trinevis*. *Genome Announc.* 2013 Jul 11;1(4). pii: e00503-13. doi: 10.1128/genomeA.00503-13.

Weiss M et al. Permanent draft genome sequence of *Comamonas testosteroni* KF-1. *Stand Genomic Sci.* 2013;8(2). doi:10.4056/sigs.3847890. Epub 2013 May 30.

Wright J et al. Genomic properties of Marine Group A bacteria indicate a role in the marine sulfur cycle *The ISME Journal*. Epub 12 September 2013; doi:10.1038/ismej.2013.152

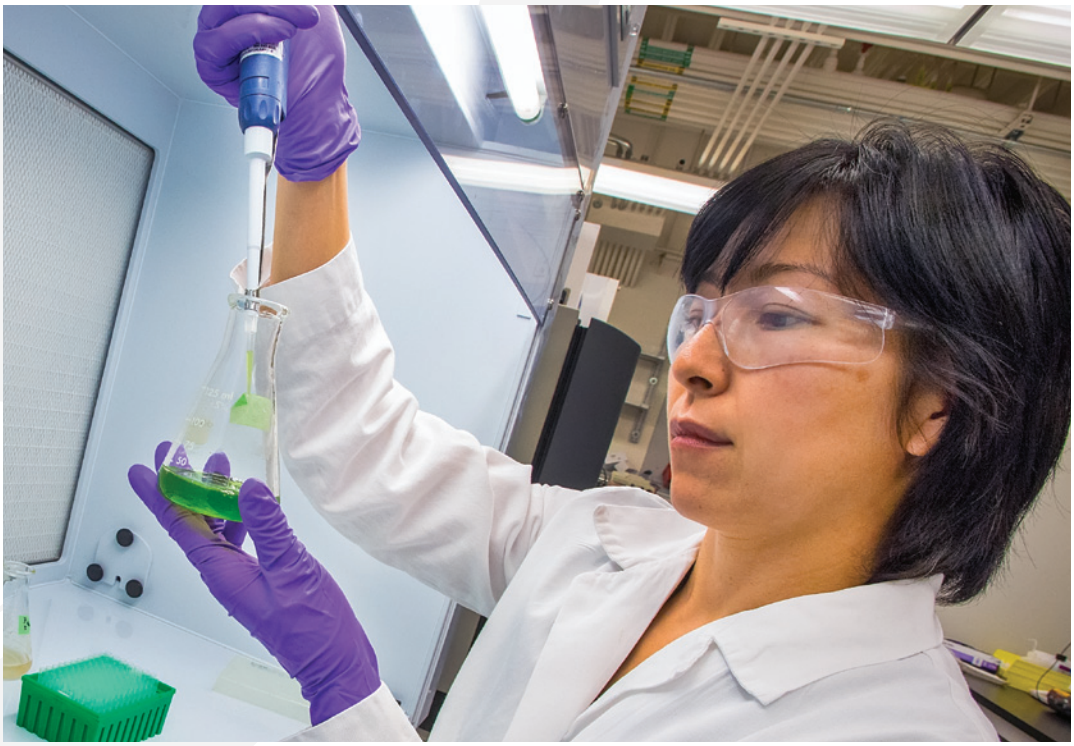
Wullschleger SD et al. Revisiting the sequencing of the first tree genome: *Populus trichocarpa*. *Tree Physiol*. 2012 Oct 25. [Epub ahead of print.]

Xie G et al. Genome sequences of two *Klebsiella pneumoniae* isolates from different geographical regions, Argentina (Strain JHCK1) and the United States (Strain VA360). *Genome Announc.* 2013 May 2;1(2). pii: e00168-13. doi: 10.1128/genomeA.00168-13.

Yang R et al. The reference genome of the halophytic plant *Eutrema salsugineum*. *Front Plant Sci.* 2013;4:46. doi: 10.3389/fpls.2013.00046. Epub 2013 Mar 21.

Zarzycki J et al. Cyanobacterial-based approaches to improving photosynthesis in plants. *J Exp Bot*. 2013 Jan;64(3):787-98. doi: 10.1093/jxb/ers294. Epub 2012 Oct 23.

Zhang JY et al. Development of an integrated transcript sequence database and a gene expression atlas for gene discovery and analysis in switchgrass (*Panicum virgatum* L.). *Plant J.* 2013 Apr;74(1):160-73. doi: 10.1111/tpj.12104. Epub 2013 Feb 18.




PYREX
Made in Germany
No. 4980
STOPPER No. 5

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.

CSO27271

U.S. DEPARTMENT OF
ENERGY
Office of Science

BERKELEY LAB


JGI
JOINT GENOME INSTITUTE
UNITED STATES DEPARTMENT OF ENERGY