2014

Progress Report

U.S. Department of Energy
Joint Genome Institute

On the cover: Near Rifle, Colorado lies the primary field site for Phase I of the Subsurface Systems Scientific Focus Area 2.0 (SFA 2.0), sponsored by the DOE Office of Biological and Environmental Research. (More on page 39 and at http://bit.ly/ESD_SFA2.)

Table of Contents

DOE JGI
Mission

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

Director's **Perspective**

On Track and According to Plan

The DOE JGI strives to fill a unique scientific niche. As a DOE Biological and Environmental Research (BER) user facility, we provide access to state-of-the-art genomic technologies supported by scientific subject matter expert staff committed to helping our users exploit these capabilities to solve complex problems related to bioenergy, carbon cycling and biogeochemistry. By focusing experimental and analysis capabilities on the best peer-reviewed ideas drawn from a broad community of scientists, we continuously encourage creative and important science relevant to the DOE's mission to advance the frontiers of energy and environmental research.

We measure our success as a user facility in several ways. First, the DOE JGI supported 1,263 primary users and collaborators in 2014. These collaborations led to 176 peer-reviewed papers, including 18 in *Science* and *Nature* journals alone. During the same time frame, the DOE JGI produced 11 plant, 100 fungal, and 845 prokaryotic *de novo* genomes, as well as 720 metagenomic data sets. Data from these projects are rapidly served to the broader scientific community via the DOE JGI data portal and three comparative genomics platforms. Together these websites were visited by over 700,000 unique IP addresses in fiscal year 2014 alone, leveraging the work done for DOE JGI's primary users into essential infrastructure for the broader energy and environmental research community.

The DOE JGI's influence spreads beyond the primary user scientists who gain access to sequencing, analysis and various other experimental capabilities through peer-reviewed user programs. Our resources extend to communities of scientists and bioinformaticians who draw upon the data and metadata we generate, analyze, and then serve through our portals. In addition, scientists and those following in their footsteps are mentored by the DOE JGI staff in the use of genomic tools to solve energy and environmental challenges through a variety of regular workshops and our Annual Genomics of Energy and Environment Meeting.

Our cadre of scientists on staff at the DOE JGI drive new approaches to sequence-based science and enable world-class contributions to the literature. For instance, during a single week in June 2014, three plant genome publications led by researchers at the DOE JGI were published in *Nature*, *Nature Biotechnology*, and *Nature Genetics*:

- **Common Bean Genome for Crop Improvements.** A DOE JGI team compared sequences of common bean, *Phaseolus vulgaris*, from pooled populations domesticated separately at distinct locations in Mesoamerica and the Andes. They found only a small fraction of shared genes indicating that different events had been involved in the domestication process at each location. They also analyzed regions of the genome associated with traits such as low diversity, flowering time, and nitrogen metabolism and identified dense clusters of genes related to disease resistance. The data advance the improvement of the common bean, though additional investigations into the genetic basis of how bean responds to biotic and abiotic stress will be required. (*Nature Genetics*, June 8, 2014)

- **Eucalyptus Genome for Biofuels and Specialty Chemicals.** A DOE JGI team identified genes encoding 18 final enzymatic steps for the production of cellulose and the hemicellulose xylan, both cell wall carbohydrates that can be used for biofuel production. In doing so, they defined a core set of genes as well as novel lignin-building candidates that are highly expressed in the development of the woody tissue called xylem that helps channel water throughout the tree. They also showed that eucalyptus has the highest diversity of genes for secondary metabolites such as terpenes, hydrocarbons that serve as chemical self-defenses against pests. This could provide an opportunity to exploit a particular biochemical pathway so that eucalyptus could become a viable feedstock for jet fuel. (*Nature*, June 12, 2014)

- **Citrus Genomes for Understanding Stress Response.** Part of a major international consortium, DOE JGI researchers analyzed and compared the genome sequences of 10 diverse citrus varieties. They found that these fruits are derived from two wild citrus species that diverged in Southeast Asia over five million years ago. By inferring the past hybridization events that gave rise to these common citrus varieties, these findings will help develop strategies for improving citrus. The compendium of data now available in the public portals will also help researchers worldwide to apply genomic tools to better understand how citrus varieties arose and how they respond to disease and other stresses. (*Nature Biotechnology*, June 2014)

Other high-profile publications from the DOE JGI and our collaborators in fiscal year 2014 include:

- **Identifying Stop Codon Reassignments in the Wild.** The tools of metagenomics and single-cell genomics, which illuminate the genetic blueprints of microbes without the need to grow them in the laboratory, reveal an unexplored, uncultured microbial world. We scanned 5.6 trillion base pairs of metagenomic data for stop codon reassignment events — whereby the canonical genetic code is no longer conserved — and detected recoding of stop codons in a substantial fraction of the >1,700 environmental samples examined. Phages can exploit slight changes in the codon table to suppress the host cell's protective mechanisms and conduct a "hostile takeover" of the cell. These observations are helping us get an unbiased view of how nature operates and how microbes manage our planet. (*Science*, May 23, 2014)

- **White/Brown Rot Paradigm for Wood Decay.** In analyzing 33 fungal genomes, many of which had previously been sequenced, we found that some wood-decaying fungi cannot easily be categorized as brown rots, which are capable of breaking down cellulose and hemicellulose, or white rots, which are also capable of breaking down lignin. Having so many fungal genomes as inputs for this analysis enabled us to see this trend and determine that there are other genetic markers we can use to identify potential white rots. The findings also broadened the range of fungal decay strategies that could prove useful for commercializing the biofuels production process. (*PNAS*, June 23, 2014)

- **Identifying Livestock Gut Microbes That Contribute to Greenhouse Gas Emissions.** Ruminants such as cattle and sheep are the single largest source of methane emissions, and methane is a greenhouse gas nearly 30 times more potent that carbon dioxide. To learn more about mitigating the amount of methane produced by livestock, we teamed with researchers in New Zealand to study the gut microbiomes in sheep, which outnumber humans seven to one, and identify the low-methane emitting animals. The deep sequencing we employed could help our collaborators in potentially breeding livestock with reduced-emissions traits, without impacting other desirable traits such as wool and meat quality. (*Genome Research*, June 6, 2014)

- **Signatures of Selection Inscribed on Poplar Genomes.** Building off of the reference poplar genome sequence we generated nearly a decade ago, we have been working on a long-term study that looks at the processes involved in shaping the genetic variation of natural poplar. This large-scale, population-based approach that looks at selection on a genome level allows us to look at which genotypes might thrive better under certain conditions compared to others, data that could help develop more accurate climate change models. The computational work yielded nearly 18 million Single Nucleotide Polymorphisms (SNPs), data that are of immediate use to tree breeding programs. (*Nature Genetics*, August 24, 2014)

- **Largest Soil DNA Sequence Analysis Effort to Date.** Soil is one of the most diverse microbiomes, and efforts to compare the soils in various Midwestern fields yielded the first in a series of publications. By analyzing the microbial composition of cornfields under continuous cultivation for the last century against pristine prairie lands, we hope to learn more about this ecosystem that traps the most carbon of any soil system in the country. What we found is that despite the nearly 400 billion bases of sequence generated, the data still aren't enough to characterize the microbial players involved. (*PNAS*, March 14, 2014)

- **Duckweed Genome Biofuel Feedstock Potential.** Despite its size, the sheer abundance of the tiny yet fast-growing duckweed plant could provide a viable candidate biofuel feedstock due to its ability to double its population within a couple of days. Additionally, unlike other candidate feedstock plants, duckweed has tiny amounts of lignin and cellulose, which have been challenging and cost-inefficient to remove for commercial biofuel production. Sequencing its simple genome — it turns out to have one of the smallest known plant genomes, only 158 Mb — revealed it has several genes that repress the switch from juvenile to mature leaf growth, as well as fewer genes related to cell wall and root growth. (*Nature Communications*, February 19, 2014)

These are just a small sample of this year's contributions to the scientific literature. A more extensive description of our achievements can be found on pages 22-39.

We published our Ten-Year Strategic Vision in 2012 after extensive engagement with a diverse selection of DOE JGI users and external experts in genomics, computing, and energy and environmental sciences. A central theme to the vision for DOE JGI's future is the recognition that a fundamental unsolved problem in genomics is the need for "high-throughput" approaches to bridge the gap between the availability of sequence and our ability to assign biological function. Since then, the major challenge in genome-scale investigations has shifted from sequence data generation to the robust isolation of specific material for sequencing and the functional and computational analysis of the data once they are generated. Accordingly, this Vision continues to guide us as a roadmap for the development of a suite of large-scale experimental and computational capabilities, as well as organizational structures to meet the needs of the Institute's users in the future.

New capabilities that have been established include epigenomics and methylation analyses; transposon-mediated mutagenesis paired with sequencing, transcriptomics, metatranscriptomics; and exome and whole genome resequencing. In each case the capability is coupled with appropriate analysis tools. A DNA Synthesis Science Program has also been developed to enable users to combine mining of the DOE JGI databases for hypothesis generation with design and synthesis of genes and gene pathways that can be used to study functional features of the sequence in appropriate hosts. These capabilities, which we expect to evolve with time, align with the DOE JGI's transition from a production-sequencing center to a genomic science analysis resource, allowing researchers to convert sequence data into biological insights.

One major initiative associated with implementing the Vision that is in full swing in 2014 is the Emerging Technologies Opportunity Program (ETOP), a program developed to facilitate the availability of cutting-edge sequence-to-function capabilities developed with collaborators to DOE JGI users. Through the ETOP, the DOE JGI has developed partnerships with world leaders in microfluidics (Stanford/Broad Institute), Raman spectroscopy and single-cell function-driven genomics (MIT/University of Vienna), advanced DNA synthesis technologies (University of Washington), metagenomic assembly (UC Berkeley/ORNL), and plant (Arizona State University) and fungal (Pacific Northwest National Laboratory — PNNL) material acquisition.

Another change that has gained momentum in 2014 in support of the goals of the strategic plan was the implementation of the JGI-EMSL Collaborative Science Initiative, a joint user program with the Environmental Molecular Sciences Laboratory (EMSL), a DOE user facility at PNNL. EMSL provides energy and environmental researchers with a broad array of sophisticated molecular characterization technologies, particularly in cellular imaging and sorting, and proteomics. Linking EMSL's molecular science capabilities with the DOE JGI's genomic capabilities in a single call provides users with easy access to both facilities to address questions that could not be answered by a either facility alone. This synergy is reflected in the large number of high-quality projects proposed by applicants to the first two joint JGI-EMSL user calls.

In addition to experimental capabilities, in 2014 we expanded our active partnership and renewed the memorandum of understanding with the National Energy Research Scientific Computing Center (NERSC) operated by Lawrence Berkeley National Laboratory. NERSC is one of the world's largest supercomputing facilities devoted to providing computational resources and expertise for basic scientific research. The DOE JGI computing infrastructure is now managed by NERSC with our IT staff joining NERSC groups with expertise in High Performance Computing (HPC), large-scale storage systems and networking. This has led to dramatically improved performance of our large compute cluster. The DOE JGI has also expanded its relationship with the LBNL Computational Research Division to focus on the identification of DOE JGI tasks that are scalable and can be more effectively performed in a HPC environment.

In the pages that follow, we highlight the scientific and technical advances achieved in 2014 aligned with our ongoing strategic planning efforts that are contributing to the DOE JGI's evolution as a "Next-Generation Genome Science User Facility." Based on the overall positive feedback we received from our Triennial Review for the DOE Office of Biological and Environmental Research, conducted in late 2014, we expect the upward trend in enabling groundbreaking science to continue.

I look forward to apprising you of our progress in sustaining the DOE JGI as the preeminent genome science user facility dedicated to solving urgent problems in energy and environmental research.

Edward ("Eddy") M. Rubin, MD, PhD
Director, DOE Joint Genome Institute

DOE Mission Areas

Bioenergy

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

One of the DOE JGI's Grand Challenge projects involves comparing microbial communities in untouched prairie and in prairie converted to farmland. Learn more on page 31. *(David Cornwell via Flickr CC BY-NC-ND 2.0)*

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

Yasuo Yoshikuni
DNA Synthesis Science

Chia-Lin Wei
Sequencing Technologies

Nikos Kyrpides
Prokaryote Super Program

Susannah Tringe
Metagenome Program

Tanja Woyke
Microbial Program

Atif Shahab
Institutional Informatics

Dan Rokhsar
Eukaryote Super Program

Igor Grigoriev
Fungal Program

Jeremy Schmutz
Plant Program

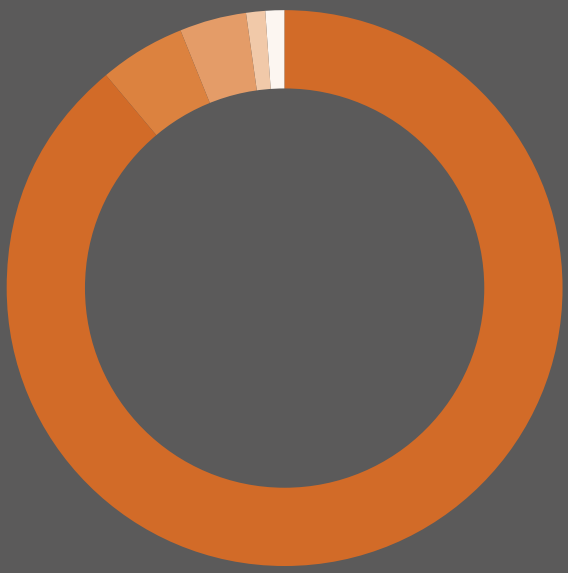
Impact **2014**

Primary Users **Fiscal Year 2014**

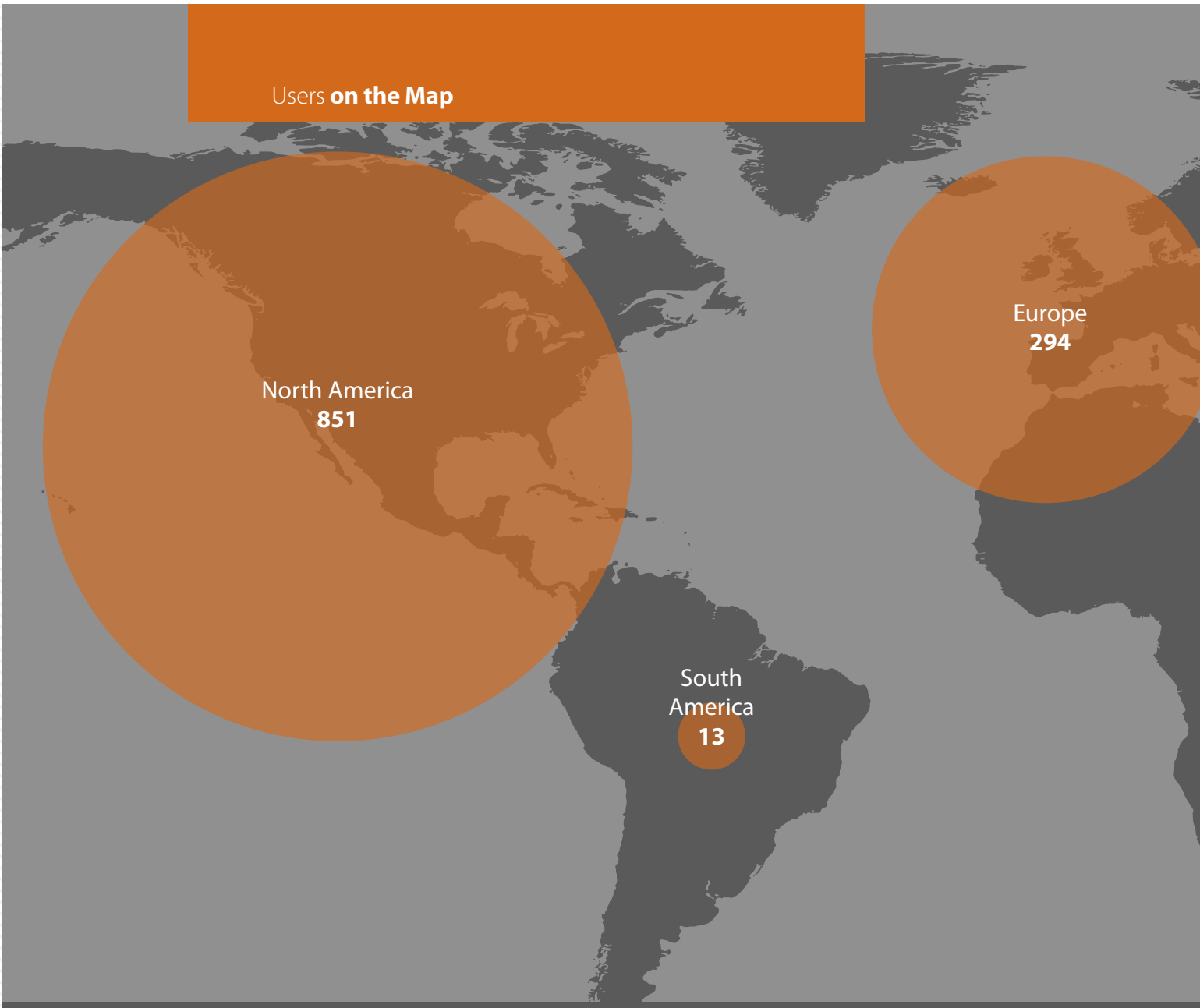
This category captures the primary users of the DOE JGI, which includes PIs and their collaborators on all user projects that were active during FY 2014. Each user is uniquely identifiable and is counted once per year regardless of the number of active projects in which he/she may be involved. This count does not include collaborators who are employed by the DOE JGI or funded through the DOE JGI's partner subcontracts.

Users **by Institution Type**

Academic	1,001
DOE Laboratory	118
Government	93
Company	28
Other Non-Profits	23



Users **on the Map**



North America	851
United States	802
Canada	44
Mexico	3
Puerto Rico	2

South America	13
Brazil	6
Argentina	3
Chile	1
Panama	1
Peru	1
Uruguay	1

Europe	294
Germany	68
France	38
United Kingdom	35
Spain	29
Netherlands	26
Austria	14
Sweden	12
Switzerland	12
Italy	9
Finland	8
Belgium	7
Russian Federation	7
Greece	6
Portugal	6
Slovenia	4
Norway	4
Denmark	3
Serbia	2
Turkey	2
Hungary	1
Iceland	1

Africa	6
Senegal	2
South Africa	2
Egypt	1
Tunisia	1

Asia	52
Japan	19
China	10
India	9
Israel	3
Hong Kong	2
Indonesia	2
Malaysia	2
Republic of Korea	2
Singapore	2
Philippines	1

Australia & New Zealand	47
Australia	34
New Zealand	13

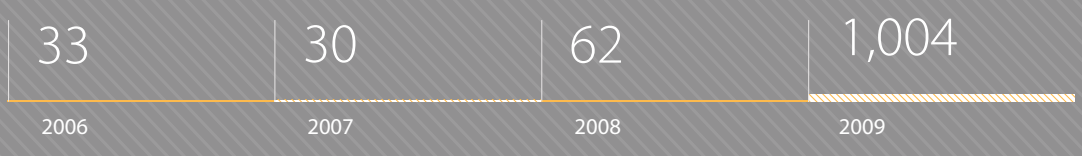
Users of **JGI Tools & Data**

DOE JGI systems also support investigators who have utilized computational 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.

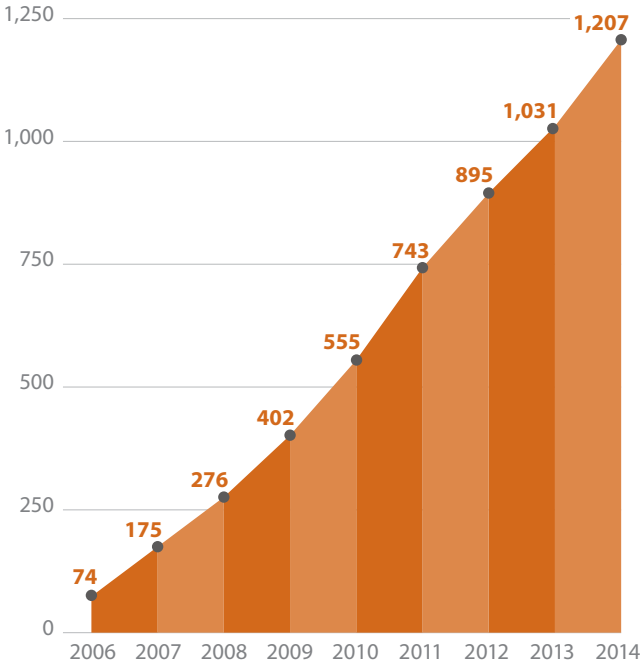
Workshops and Meetings	
Workshop Participants	**505**
Genomics of Energy & Environment	
Annual User Meeting Attendees *(total/non-JGI)*	**457/286**
Web Portal Visitors *(unique visits)*	
JGI Portal	**350,000**
IMG Systems	**342,993**
MycoCosm	**104,000**
GOLD	**107,537**
Phytozome	**120,000**
VISTA	**17,215**

Sequence **Output**

(in billions of bases or GB)

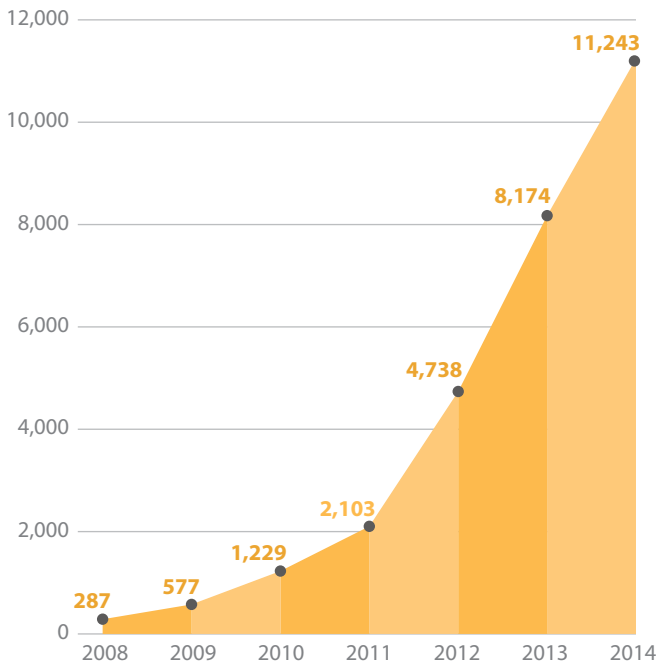


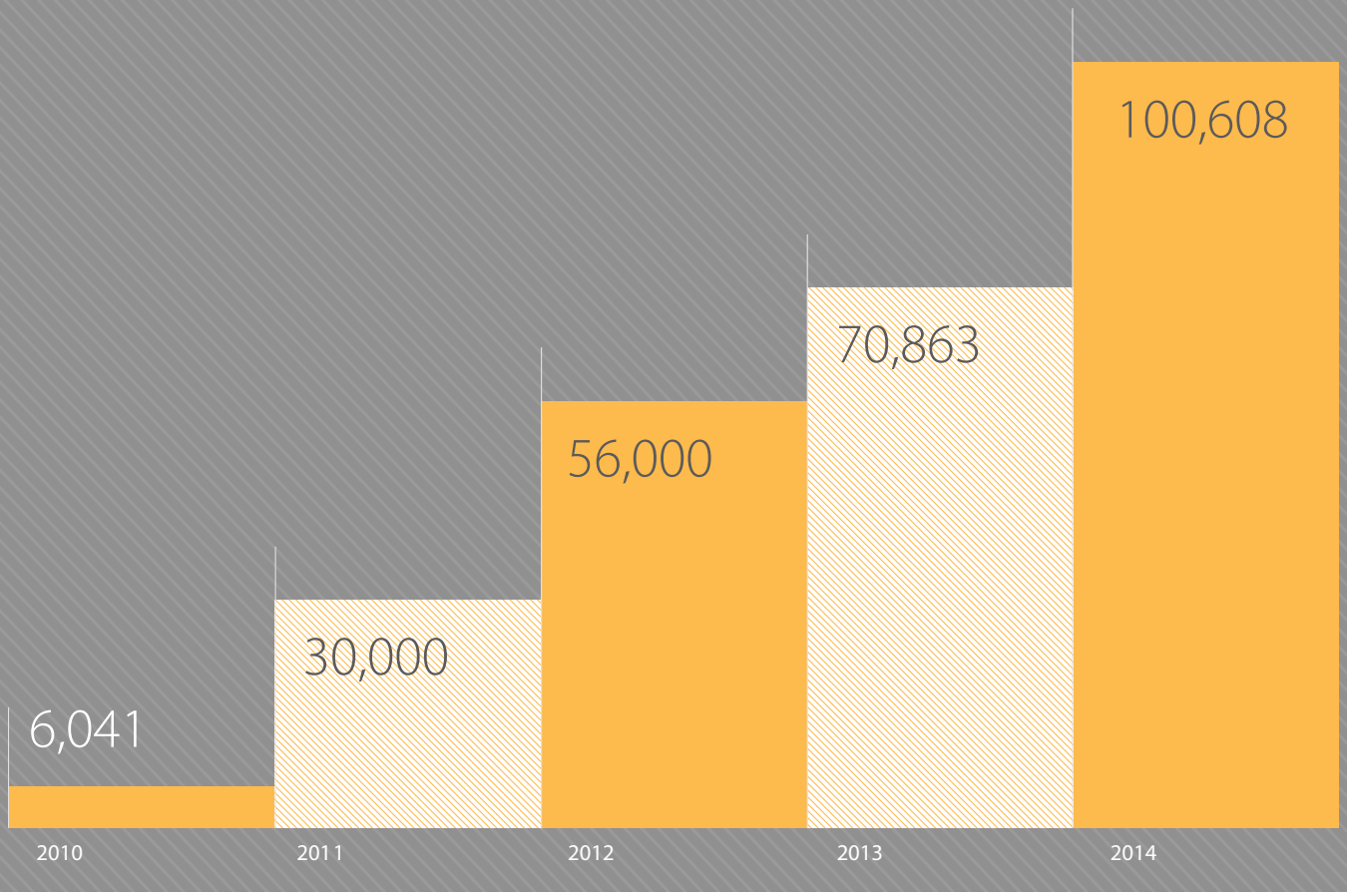
Cumulative Number of **Scientific Publications**



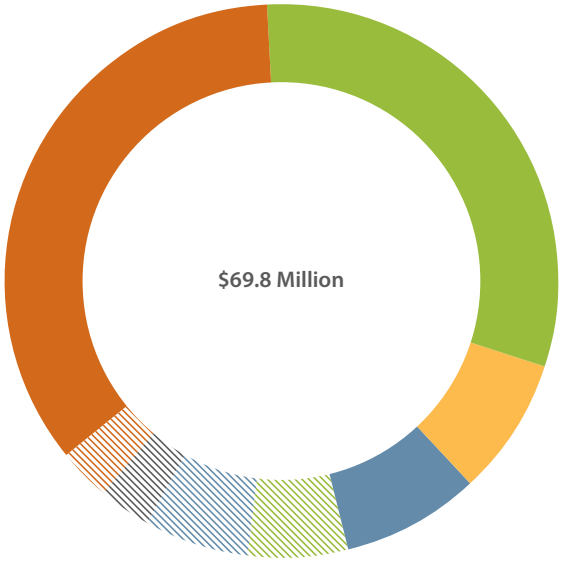
Papers published in 2011-2013 were cited nearly 5,000 times by researchers in 2014.

Cumulative Number of **Projects Completed**



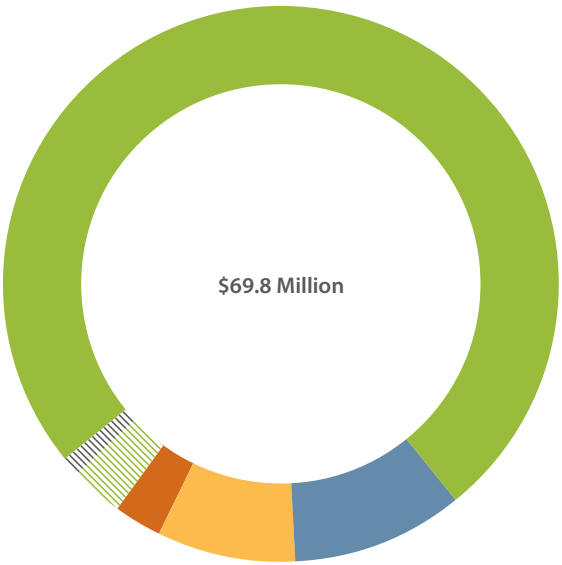


JGI Funding **FY2014**



- **35%** Science & Analysis
- **31%** Sequence/Data Generation
- **8%** R&D
- **8%** IT Infrastructure
- **6%** Operations
- **6%** Facilities & Lease
- **3%** User Support
- **3%** Management

Partner Institution Funding Profile **FY2014**



- **76%** JGI Walnut Creek

24% Partner Institutions

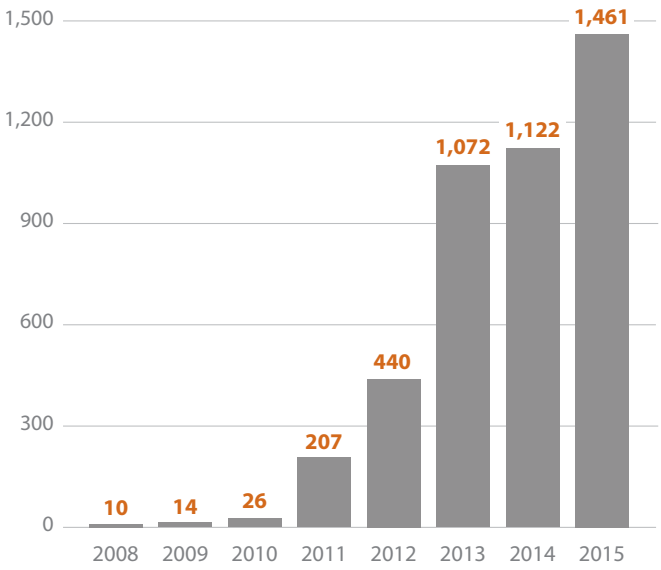
- **10%** HudsonAlpha Institute for Biotechnology
- **8%** National Energy Research Scientific Computing Center
- **3%** Lawrence Livermore National Laboratory
- **3%** Emerging Technologies Opportunity Program (2–National Laboratory & 6–Academic Institution)
- **< 1%** Oak Ridge National Laboratory

Single-Cell Genomics: **A Case Study**

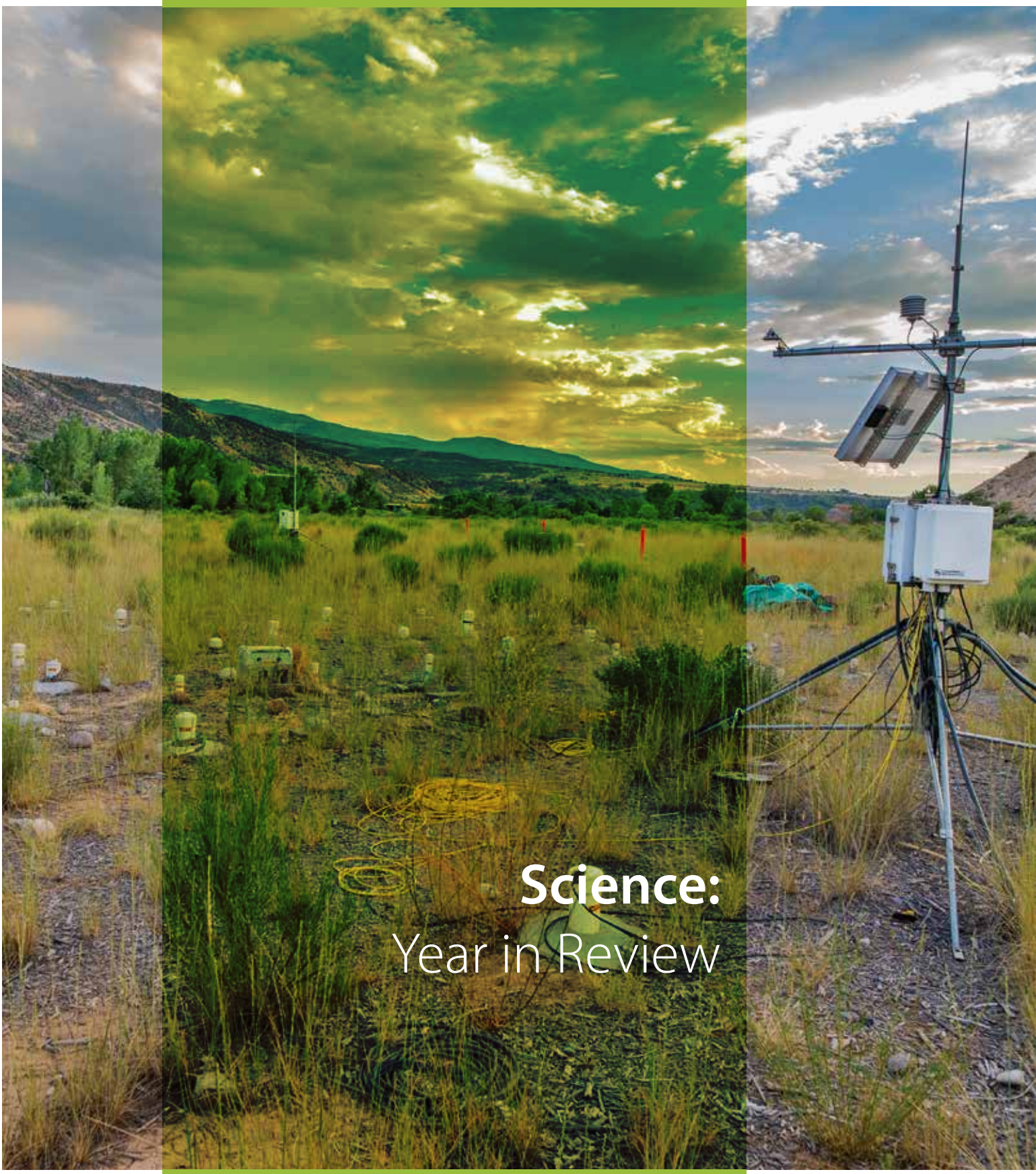
In addition to our external users, the research efforts of a small cadre of scientists based at the DOE JGI also serve as an important scientific driver for the Institute. The DOE JGI scientific staff are encouraged to explore new approaches to sequence-based science, as well as develop and apply novel experimental and computational strategies. This has resulted in important scientific outcomes in addition to developing technologies and approaches that have been later embraced by DOE JGI users. One example of this process is illustrated by the study of uncultivated environmental microbes. DOE JGI scientists pioneered both the development of computational strategies for interpreting massive metagenomic datasets, and the use of single-cell genomics to study the genetic make-up of individual uncultivated microbial cells from complex communities. Presently the DOE JGI's metagenomic analysis pipelines and single-cell capabilities are in heavy demand by users. Single-cell sequencing at the DOE JGI is offering a new window into the study of microbial ecology, evolution and ecosystem function, with a strong emphasis on exploring microbial dark matter and filling in the branches in the tree of life.

Meeting User Demand: **Single-Cell Genomics**

CSP Single-Cell Genome Requests

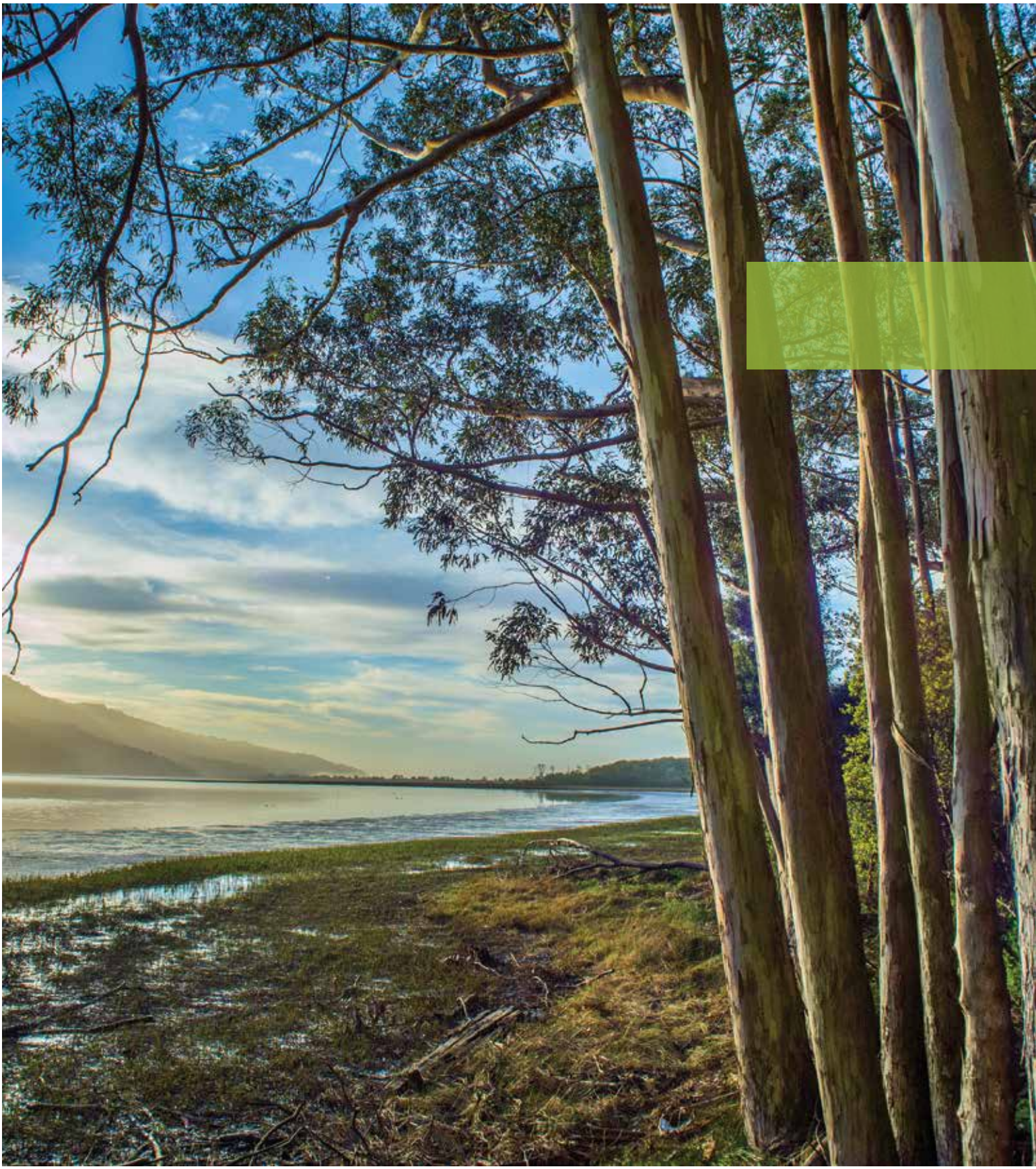


Crystal Geyser in Utah near Green River was created in the 1930s from an 800-meter deep abandoned oil exploration well. It is a cold CO_2-driven geyser that can erupt up to 40 m in height and has been shown to contain unusual microbial diversity rich in candidate phyla for characterization for their role in carbon and other element cycling. To decipher microbial community structure and function, the DOE JGI is working with collaborators to investigate this model aquatic system with an array of different technologies, including deep metagenomic sequencing including long read technologies, metatranscriptomics, and single-cell genomics.



Science:
Year in Review

The DOE Joint Genome Institute is at the forefront of many of the most pressing scientific inquiries. Its research endeavors can be organized by three DOE-mission areas: bioenergy, carbon cycling, and biogeochemistry; however, this can also involve, as all science does, "boldly going where no one has gone before" into explorations of previously unexplored areas of biology and making surprising discoveries. DOE JGI's fields of study often intersect, allowing our investigations to be applied toward multiple solutions for the most significant impact. Our researchers focus on organisms that run the gamut from minute microbes to massive trees.



Bioenergy

Eucalyptus: A Principal Candidate Biomass Energy Crop

While native to Australia, eucalyptus trees are planted worldwide mostly for the value of its wood; for the Department of Energy, their energy-rich cellulosic biomass makes them one of the principal candidate biomass energy crops. Cellulose and hemicellulose make up approximately 80 percent of the woody biomass in a eucalyptus, with the remaining biomass primarily comprised of lignin, the tough "glue" that holds it all together. Reported in the June 12, 2014, edition of *Nature*, the international effort to sequence and analyze the 640-million-base-pair genome of *Eucalyptus grandis* engaged more than 80 researchers from 30 institutions, representing 18 countries. Researchers at the University of Pretoria in South Africa, the Brazilian Agricultural Research Corporation (EMBRAPA), Oak Ridge National Laboratory (ORNL), the BioEnergy Science Center, DOE JGI, and the HudsonAlpha Institute for Biotechnology led the project.

Because of its wide adaptability, extremely fast growth rate and excellent wood and fiber properties, eucalyptus trees are grown in 100 countries across six continents and account for over 40 million acres. Combing through the 36,000-plus genes found in eucalyptus (nearly twice as many as in the human genome), the researchers homed in on those that may influence the production of secondary cell wall material that can be processed for pulp, paper, biomaterials and bioenergy applications. The eucalyptus team identified genes encoding 18 final enzymatic steps for the production of cellulose and the hemicellulose xylan, both cell wall carbohydrates that can be used for biofuel production.

An additional finding by the team was that among sequenced plants to date, eucalyptus showed the highest diversity of genes for specialized metabolites such as terpenes. These hydrocarbons serve as chemical self-defenses against pests, as well as providing the familiar aromatic essential oils used in both medicinal cough drops and for industrial processes. "By having a library of these genes that control the synthesis of terpenes we are able to dissect which genes produce specific terpenes; then we can modify this biochemical pathway in the leaves so that we can develop the potential of eucalyptus as an alternative source feedstock for jet fuel," noted DOE JGI and ORNL researcher Jerry Tuskan.

The extensive catalog of genes contributed by the team will allow breeders to adapt eucalyptus trees for sustainable energy production in regions, such as the Southeastern United States, where it cannot currently be grown.

A video of Jerry Tuskan on the implications of the team's eucalyptus genome analysis can be viewed at http://bit.ly/eucalyptusTuskan.

Duckweed has also been used to clean contaminated water. *(Benjamin Lewis via Flickr CC BY-NC-SA 2.0)*

Duckweed Genome's Biofuel Feedstock Potential

Despite its size, the sheer abundance of the tiny yet fast-growing Greater Duckweed (*Spirodela polyrhiza*) plant makes it a viable candidate biofuel feedstock. It grows almost everywhere at nearly all altitudes, and can double its population within a couple of days. Additionally, unlike other candidate feedstock plants, duckweed has tiny amounts of lignin and cellulose, which have been challenging and cost-inefficient to remove for commercial biofuel production. In the February 19, 2014, issue of *Nature Communications*, researchers from Rutgers University, the DOE JGI, and several other facilities detailed the complete genome sequence of *S. polyrhiza*.

Sequencing its simple genome — it turns out to have one of the smallest known plant genomes, about 158 million bases with fewer than 20,000 genes — revealed it has several genes that repress the switch from juvenile to mature leaf growth, as well as fewer genes related to cell wall and root growth. This switch repression is known as neoteny.

"The most surprising find was insight into the molecular basis for genes involved in maturation — a forever-young lifestyle," said senior author Joachim Messing, director of the Waksman Institute of Microbiology at Rutgers University. The leaves of *S. polyrhiza* resemble cotyledons, embryonic leaves inside plant seeds that become the first leaves after germination. Unlike other plants however, duckweed continuously produces cotyledon leaves through neoteny, in which juvenile traits are retained.

"Because of the reduction in neoteny, there is an arrest in development and differentiation of organs. So this arrest allowed us to uncover regulatory networks that are required for differentiation and development," Messing said. Researchers also found that many of the genes responsible for cellulose and lignin production in land-dwelling plants were missing in duckweed, and there were fewer copies of those genes that were present. Genes for another compound related to cell walls called "expansins," which are involved with cell wall and root growth, were also reduced.

Tracking Monkey Flower's Gene-shuffling Hotspots

Since Darwin's days, the genus *Mimulus* has been a model system for studying ecological and evolutionary genetics in nature. In the November 26, 2013, issue of the *Proceedings of the National Academy of Sciences*, a team led by DOE JGI researchers provided a reference genome for the monkey flower (*Mimulus guttatus*), a cousin of the snapdragon. The researchers were also able to track more than 400,000 DNA recombination events within a wild *Mimulus* population. These events contribute to genetic variation, and help explain how different individuals in the population may harbor a diverse set of traits. The information adds to centuries of *Mimulus'* role as a leading model system for studying ecological and evolutionary genetics in nature, and contributes insights to the plant community that could translate into crop improvement strategies for potential candidate biofuel feedstock crops.

"We were able to accomplish this by a novel method of analysis developed at the JGI involving sequencing of pooled DNA from the population and aligning these sequences to a *Mimulus* reference assembly," said the study's first author, DOE JGI computational scientist Uffe Hellsten. "This analysis allows us to pinpoint very accurately, down to a few letters of genetic code, the variation of recombination across the genome, and demonstrate that recombination is enhanced near starts of genes." Regions with high recombination activity are dubbed "hotspots."

"While cold spots appear to be entirely devoid of recombination, hot spots display a spectrum of 'temperatures,' ranging from 'lukewarm,' which are common in the genome, to increasingly hotter in less common regions," Hellsten added. "We found that recombination events occur much more commonly close to the beginnings of genes and that more than 25 percent of the genome consists of cold spots, which don't participate in recombination at all, are inherited as unshuffled blocks from parents."

A well-known example of evolution involves copper-tolerant populations of monkey flower found around an abandoned California copper mine. *(James Gaither)*

Signatures of Selection Inscribed on Poplar Genomes

Researchers have identified genomic regions that contribute to adaptive traits for wild poplar populations, such as how regional temperatures can affect the timing of spring bud flush. *(David Gilbert)*

Forests creeping steadily north and becoming established in the thawing Arctic is just one of the predicted effects of rising global temperatures. In a study published ahead online on August 24, 2014, in *Nature Genetics*, a team led by Jerry Tuskan of the DOE JGI and Oak Ridge National Laboratory and by Stephen DiFazio of West Virginia University used a combination of genome-wide selection scans and analyses to understand the processes involved in shaping the genetic variation of natural poplar (*Populus trichocarpa*) populations. The data generated could help develop more accurate climate change models.

As part of this long-term study that builds off of the reference poplar genome sequence generated by the DOE JGI a decade ago, the team took samples from 1,100 poplar trees growing in wild populations in California, Oregon, Washington and British Columbia. They then clonally propagated (through cuttings) these trees in three plantations in California and Oregon. For their analyses, they pared the group down to 544 unrelated individuals whose genotypes could be accurately determined so as to characterize the genetic basis for variation in adaptation. "This is the first time that deep genomics resources have ever been applied to an ecological question, in this case: 'What does selection do at the genome level?'" said Tuskan. "In the past, people looked at adaptation to factors such as temperature and light levels, and they examined variation in those genes as they vary across environmental gradients. There was a preconceived notion and a very narrow view of what was causing the response. Here, we took five major approaches, applied them blindly to the whole genome, and let the analysis show us where the fingerprints of selection are and what genes fall under those fingerprints."

Going from 1,000 genotypes and 45,000 genes "to figuring out what's not just statistically significant but biologically meaningful" wasn't easy. DiFazio highlighted the significance of the computational work that yielded nearly 18 million SNPs, data that is of immediate use to tree breeding programs. "Our approach is particularly powerful because we are mining standing natural variation resulting from tens of thousands of years of evolution and selection such that the alleles or gene variants that we have identified have great promise to provide robust, long-term improvements to biofuel feedstocks."

A video of Tuskan discussing the importance of genetic selection in trees can be viewed at http://bit.ly/Tuskan14fingerprints.

Symbiotic Root Fungi: A Plant's Best Friend

In the December 10, 2013, issue of the *Proceedings of the National Academy of Sciences*, a team led by the French National Institute for Agricultural Research (INRA) and including DOE JGI researchers decribed the genome of *Rhizophagus irregularis*, the first arbuscular mycorrhizae fungus (AMF) to be sequenced. The fungus is a member of the Glomeromycota family, an ancient lineage of fungi that has a symbiotic relationship with roots going back nearly 420 million years to the earliest plants. More than two-thirds of the world's plants depend on this soil-dwelling symbiotic fungus to survive, including critical agricultural crops such as wheat, cassava, and rice.

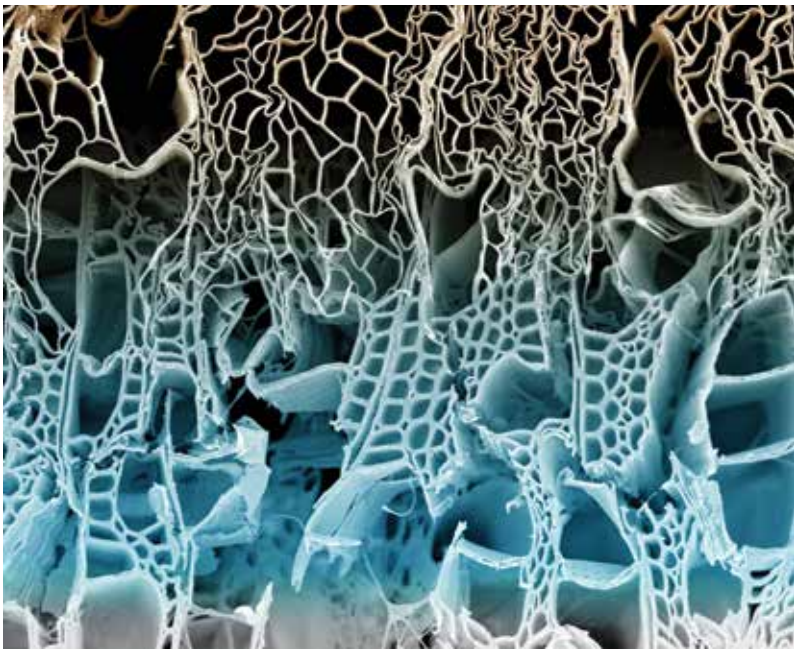
Though AMF play a vital role in how phosphorus and carbon cycle through the atmosphere and land-based ecosystems, the fungi's role in this vital job is poorly understood. A relic of fungal evolution, AMF diverged early on from other forms of fungi. In root cells, they form dense clusters of branched structures called arbuscules that serve as the main route of nutrient exchange between plants and fungi. Unable to live on their own, AMF maintain symbiotic relationships with their plant hosts, keeping them alive in exchange for needed sugars. Scientists theorize that the benefits these fungi provided enabled ancient plants to evolve during the Paleozoic era, about 250 to 500 million years ago.

The analysis of the *R. irregularis* genome has revealed that this asexual fungus doesn't shuffle its genes the way researchers expected. Moreover, rather than having lost much of its metabolic genes, as observed in many mutualistic organisms, it has expanded its range of cell-to-cell communication genes and phosphorus-capturing genes. Phosphorus, a critical element for cellular function, is otherwise difficult to extract from the soil and is often the limiting factor for how quickly a plant grows.

Teasing apart the complex relationship between soil fungi and plants is likely to have an impact on improving biofuel production from plant biomass. "Through analysis of this and other mycorrhizal genomes, we can help to better understand interactions and conditions critical for a sustainable growth of bioenergy plants, but also staple crops, a prerequisite to help feeding the world," said INRA's Francis Martin, a study senior author and longtime DOE JGI collaborator.

White/Brown Rot Paradigm for Wood Decay

If a fungus can break down all the components — cellulose, hemicellulose and lignin — of plant cell walls, it is considered a white rot fungus. If a fungus can only break down cellulose and hemicellulose but not lignin, it is considered a brown rot fungus. In the July 8, 2014, issue of the *Proceedings of the National Academy of Sciences*, a team led by DOE JGI researchers reported that categorizing wood-decaying fungi as either white rot or brown rot may not be as clear-cut as previously thought. The discovery complicates but also broadens the range of fungal decay strategies to be explored for commercializing the process of biofuels production.



Micrographs of transverse sections of *Populus* wood decayed by brown rot (top) and white rot (bottom) fungi. *(Images by Benjamin Held, University of Minnesota. Artist interpretation by Zosia Rostomian, Berkeley Lab Public Affairs.)*

This finding emerged after researchers analyzed 33 basidiomycete fungal genomes, 22 of which are wood decayers, four of which had been recently sequenced by the DOE JGI. Based on previously sequenced genomes, the team observed that two of the new fungi, *Botryobasidium botryosum* and *Jaapia argillacea*, had the cellulose-attacking enzymes characteristic of white rot fungi, but lacked certain lignin-degrading enzymes called class II peroxidases (PODs) that are typically associated with this group. Applying a statistical process called Principal Components Analysis (PCA) to find similarities in fungi based on their plant biomass degrading genes, they found that the two new fungi grouped close to *Phanerochaete chrysosporium*, the first white rot species sequenced. This was a curious finding because the new fungi were phylogenetically distant from *P. chrysosporium*, and, moreover, didn't have PODs.

The team then grew isolates of *B. botryosum* and *J. argillacea* on pine and aspen wood. They found that the fungi superficially degraded the wood surfaces, but in localized areas, went further and broke down the cell walls and removed cellulose, hemicellulose and lignin.

DOE JGI Fungal Genomics head Igor Grigoriev noted this wasn't the first time they'd seen a genome that appeared to blur the definitions between white rots and brown rots. "We thought we saw an anomaly with a previously sequenced white rot fungus, *Schizophyllum commune,*" he said. "Now we see a trend. This is the value of having multiple data points and so many fungal genome sequences. This is the whole point of doing fungal genomics at scale. Now that it's clear that [POD] is not the only player, we should broaden our search for enzymes that have bioenergy applications. It is important to identify a whole range of enzymes sourced from nature that can be used to develop second-generation biofuels in terms of breaking down lignin and other components in plant cell walls."

Pinpointing bacterial drug candidates in sponges

The kidney-red coral reef sponge, *Theonella swinhoei*, is a source of several anti-fungal and anti-cancer drug candidates, but it isn't the producer. Symbiotic bacteria that live in the sponge produce these compounds, and they were identified in a collaborative effort led by the Institute of Microbiology at ETH Zurich, Switzerland, and including DOE JGI researchers. The work was described online on January 29, 2014, in *Nature*.

Scientists have been harvesting medically important compounds from bacteria grown in labs for decades, but it's estimated that these "cultivated bacteria" represent only one percent of known microbes on the planet,

and it's unclear how many of the uncultivated 99 percent are capable of producing other useful natural compounds. This is the first time researchers have pinpointed an uncultivated bacterium as a prolific source of bioactive compounds.

"In this study we bypass the traditional culturing step and demonstrate the single-cell- and metagenomics-based discovery of such producers," said DOE JGI computational biologist Christian Rinke. For the study, he provided a phylogenetic analysis, which deciphers the evolutionary history of bacterial lineages. "Metagenomics and single cell genomics, which enable us to obtain genome sequences from uncultivated microorganisms, highlight the potential of both techniques and provide new enticements to apply to environmental samples for drug discovery as well as for sourcing enzymes for such industrial applications as biofuels."

Salt-Tolerance Lessons from a Dead Sea Fungus

With an increasing global population and changing global climate, understanding how organisms called "halophiles" have adapted to thrive in salty environments could prove useful in developing fuel and food crops that can tolerate cultivation in dry and salty areas. Some organisms thrive in salty environments by lying dormant when salt concentrations are very high. Other organisms need salt to grow. To learn which survival strategy the filamentous fungus *Eurotium rubrum* uses to thrive in Israel's Dead Sea, a team including DOE JGI researchers studied its genome, describing their findings in the May 9, 2014, issue of *Nature Communications*.

The DOE JGI team first sequenced, assembled and annotated the 26.2-million base genome of *E. rubrum*. When the team compared *E. rubrum's* gene families against those in two other halophilic species, they found that high acidic residues were common in all three species, a general trait all salt-tolerant microbes share.

To learn more about the fungus' tolerance for salt, samples were grown in liquid and solid media at the University of Haifa, at salinities from zero up to 90 percent of Dead Sea water. The researchers found that it had viable spores when grown in 70 percent diluted Dead Sea water. A study of *E. rubrum's* transcriptome conducted at the University of Bayreuth found that in high salinity conditions, the fungal cells need to keep cell membrane transport under tight control, suggesting that the fungus does not go dormant as might have been expected.

Additionally, the *E. rubrum* genome sequence could provide insights into identifying enzymes that can tolerate ionic liquids, environmentally benign organic salts often used as green chemistry substitutes for more cost-effective biomass pretreatment processes.

Despite its name, some algae, bacteria, and fungi thrive in and around the Dead Sea. *(Itamar Grinberg for the Israeli Ministry of Tourism via Flickr CC BY-SA 2.0)*

Carbon Cycle

Largest Soil Sequence and Analysis Effort to Date

"It's one of the most diverse microbial habitats on Earth, yet we know surprisingly little about the identities and functions of the microbes inhabiting soil," said Jim Tiedje, Distinguished Professor at the Center for Microbial Ecology at Michigan State University. With MSU colleagues and collaborators from the DOE JGI and Lawrence Berkeley National Laboratory (Berkeley Lab), Tiedje published the largest soil DNA sequencing effort to date in the March 10, 2014, issue of the *Proceedings of the National Academy of Sciences* (PNAS).

In this ambitious pilot study launched by the DOE JGI, which provided the raw sequencing power to actually do it, MSU researchers sought to compare the microbial populations of different soils sampled from Midwestern corn fields, under continuous cultivation for 100 years, with those sourced from pristine expanses of the Great Prairie. "The Great Prairie sequesters the most carbon of any soil system in the U.S. and produces large amounts of biomass annually, which is key for biofuels, food security, and carbon sequestration," said Tiedje. "It's an ecosystem that parallels the large ocean gyres in its importance in the world's primary productivity and biogeochemical cycles."

For the Great Prairie soil experiment, the team generated nearly 400 billion letters of code, which amounts to more than 130 human genome equivalents, or 88,000 *E. coli* genomes. To parse the data, MSU's C. Titus Brown and Adina Chuang Howe deployed a compression method, common with large computer files such as JPEG images conveyed through the Internet, that allows a substantial amount of data to be discarded without the actual data content being degraded. Brown calls the technique "digital normalization," resulting in a 2- to 200-fold decrease in computational requirements for the actual biological analysis. The technique is expected to enable significant improvement in genome assembly and provide the critical references to advance future investigations of soils and other complex environments. In the meantime, the results of the PNAS study indicated that despite 400 billion bases of data, much more data are needed to study the content of soil metagenomes comprehensively. *(Prairie image on left by Rachel Gardner via Flickr CC BY-NC-ND 2.0)*

Capturing the Carboxysome at Work

Found in almost every ecosystem on the planet, cyanobacteria can create their own energy through photosynthesis. Their abundance means cyanobacteria play a major role in the Earth's carbon cycle, the movement of carbon between the air, sea and land.

Inside the cyanobacteria are microcompartments called "carboxysomes" that look a lot like multi-faceted envelopes of viruses. Carboxysomes are icosahedral, having about 20 triangle-shaped sides or facets. They contain copious amounts of Ribulose 1,5 Biphosphate Carboxylase Oxygenase (commonly known as RuBisCo), an extremely abundant but slow enzyme required to fix carbon, inside their protein shells. The microcompartment also helps concentrate carbon dioxide and corral it near RuBisCo, while locking out oxygen, which otherwise tends to inhibit the chemical reactions involved in carbon fixation.

Using a pioneering visualization technique, researchers from the University of California, Berkeley and the DOE JGI established a timeline by which cyanobacteria build carboxysomes. They described their work in the November 21, 2013, issue of *Cell*. Lead author Jeffrey Cameron of UC Berkeley developed mutant strains of a *Synechococcus* cyanobacterium in which the genes for building carboxysomes were intentionally broken. He then introduced the products of each of the knocked-out genes, which had been tagged with a fluorescent marker. Through time-lapse microscopy, he captured time-lapse digital images of the bacteria as they used the glowing building blocks and incorporated them into their new carboxysomes.

The research team also painstakingly took high-resolution still photographs using a transmission electron microscope of the intermediate stages of carboxysome construction. With these detailed images, they were able to provide a specific role for each product of each knocked-out gene, along with a timeline for how the bacteria built its carboxysomes. "The results provide clues to the organization of the enzymes encapsulated in the carboxysome and how this enhances CO_2 fixation," said senior author Cheryl Kerfeld, formerly of the DOE JGI and now at Michigan State University. The team also suggested that other bacteria might build different types of microcompartments the same way carboxysomes are built, from the inside out

Mitigating Methane Emissions from Ruminant Gut Microbes

Methane is a greenhouse gas some 28 times more potent than carbon dioxide. According to the U.S. Environmental Protection Agency, ruminant livestock such as cattle and sheep are the single largest source of methane emissions, and in a country like New Zealand, where the sheep outnumber people 7 to 1, that's a big deal.

A team of researchers led by the DOE JGI deployed high throughput DNA sequencing and specialized analysis techniques to explore the contents of the rumens of sheep in collaboration with New Zealand's AgResearch Limited to see what role the microbiome (composed of microbes living in the rumen) play in this process. The researchers took advantage of a large sheep screening and breeding program in New Zealand that aims to breed low methane-emitting ruminants without impacting other traits such as reproduction and wool and meat quality. The study was published ahead online on June 6, 2014, in *Genome Research*.

The team measured the methane yields from a cohort of 22 sheep, and from this group, they selected four sheep with the lowest methane emissions, four sheep with the highest emissions and two sheep with intermediate emission levels. Rumen metagenome DNA samples collected on two occasions from the 10 sheep were sequenced at the DOE JGI, generating 50 billion bases of data each.

The team then checked to see if there was a correlation between the proportions of methanogens in the eight sheep with the highest and lowest recorded methane emissions. In sheep with low methane emissions, they found elevated levels of one particular species of methanogen while sheep with high methane emissions had elevated levels of another group of methanogens. Exploring further, the team then identified a methane-producing pathway and three variants of a gene encoding an important methane-forming reaction that were involved in elevated methane yields.

"It's not so much the actual composition of the microbiome that determines emission — which conventional wisdom would suggest — but mostly transcriptional regulation within the existing microbes that makes the difference, which is a concept that is relatively new in metagenomic studies," DOE JGI Director Eddy Rubin said. The team's findings suggest new possible targets for mitigating methane emissions at the microbiome level.

The EPA attributes one-fifth of methane emissions to livestock. *(Sheep image courtesy of AgResearch – Gerry le Roux, Sciencelens. Art by Wayne Keefe, Berkeley Lab Public Affairs.)*

Screening and breeding for low-methane producing sheep is still underway, and importantly, low-methane lines then need to be tested for stability of the trait, as well as the absence of any impacts on fertility or meat or wool production.

Prochlorococcus Single-Cell Genomics

By some estimates, up to 100 million cells of the cyanobacterium *Prochlorococcus* can be found in a single liter of water. These important organisms are thought to be responsible for providing about 20 percent of the oxygen produced by the planet each year.

Though it is considered a single species, this unicellular organism can be classified into several distinct major populations, within which *Prochlorococcus* cells still display a wide range of genomic diversity. To learn more about the cyanobacterial populations at this level, MIT marine microbiologist Sallie Chisholm, a longtime DOE JGI collaborator, led a team that applied single-cell genomics to these cyanobacteria collected from the same environment at three separate times of the year.

As reported in the April 25, 2014, issue of *Science*, the team sequenced and assembled *Prochlorococcus* genomes from single cells in the northwestern Sargasso Sea between November 2008 and April 2009. They found cell clusters within known ecotypes that indicated that the relative abundance of cyanobacterial subpopulations shifted along with the seasons, likely in response to environmental changes.

The team found that the *Prochlorococcus* subpopulations within the clusters were genomically distinct, and that each of these so-called "genomic backbones" is comprised of highly conserved core gene alleles, or alternative forms, and a smaller distinct set of "optional" genes. DOE JGI's Rex Malmstrom helped validate the vast diversity of these abundant *Prochlorococcus* subpopulations. The team noted that if each of the cyanobacterial backbone subpopulations were counted as distinct species, then *Prochlorococcus* consists of thousands of species, each helping to maintain the cyanobacteria's stability in the global oceans.

"It was heartening to work with JGI on part of this project because that is where the *Prochlorococcus* genomics story began about 15 years ago," said Chisholm. "They sequenced the genomes of two of our strains that had been isolated from different oceans, giving us our first peek at the diversity within this group. Likewise, JGI's current collection of a number of *Prochlorococcus* genomes has been an invaluable resource in our more recent work. These 'reference genomes' are key to unlocking the mysteries of the patterns we see in the genomes of wild populations."

Tagging Strategy for Marine Virus Population Surveys

As many as 100 million cells of cyanobacteria can be found in a single liter of water — but up to 100 million viruses can be found in 1mL of seawater. Despite their abundance, researchers know very little about the viruses in the water, beyond their ability to drastically decrease cyanobacterial populations, affecting the global regulation of biogeochemical cycles.

To help resolve this conspicuous lack of knowledge and learn more about viral diversity, a team led by DOE JGI collaborator Matt Sullivan at the University of Arizona conducted a population-scale survey using a game-changing new technique. Their results published online on July 13, 2014, in *Nature* suggest that there is an ecology of viruses and it can be studied by harnessing more traditional approaches that have been applied to larger organisms.

Sullivan's team focused on the cyanophages isolated from a single sample of water collected in Monterey Bay, Calif. To resolve the challenge of figuring out "who infects whom" among marine viruses and cyanobacteria, they used a technique known as viral tagging, in which viruses and so-called "host bait" are stained with a fluorescent dye in order to find out with which hosts the phages associate. Sullivan credits the original idea for the project to study co-author Phil Hugenholtz, formerly a DOE JGI researcher and now Director of the Australian Centre for Ecogenomics at the University of Queensland.

After screening for cyanophages that were tagged as being associated with a single *Synechococcus* strain (SynWH803), samples of the viral community metagenomes were sent to the DOE JGI for sequencing as part of a Community Science Program project proposed by Sullivan. The results indicated that while there were "at least" 26 viral populations associated with the cyanobacterial strain and most had been cultured, viral tagging suggested there were another 42 new, uncultured viruses specific to the cyanobacterial strain.

"The thinking before was that the viral genome sequence space would be one big blur, but this suggests there are units that we can count and study," Sullivan said. "That represents a whole new ballgame and opens up viral ecology to utilize decades of theory and practice from the study of more traditional study of larger organisms. Additionally, our method of viral tagging should be generalizable to many other virus-host studies so it should transform the way viruses in nature are studied moving forward."

Watch Matt Sullivan's marine viruses presentation from the 2012 DOE JGI Genomics of Energy & Environment Meeting at http://bit.ly/JGI7Sullivan.

Decoding Virus-Host Interactions in the Oxygen-Starved Ocean

In the past 50 years, oxygen minimum zones (OMZs) have expanded due to climate change and increased waste runoff from farms and cities. There are currently more than 500 OMZs worldwide, encompassing roughly eight percent of ocean volume that is considered oxygen-starved. Knowing how microbial interactions change in response to OMZ expansion is crucial to understanding the organizing principles underlying coupled nutrient and biogeochemical cycling in the ocean and the balance of greenhouse gases in the atmosphere.

DOE JGI collaborator Steven Hallam of the University of British Columbia has been studying a microbial community dominated by SUP05, a currently uncultivated group of microorganisms, related to gill symbionts of deep-sea clams and mussels that thrive in the most oxygen-starved regions of the water column. In a study published August 29, 2014, in the journal *eLife*, Hallam teamed with DOE JGI collaborator Matt Sullivan (featured above) to investigate marine viruses infecting SUP05 to better understand how viral infection influences SUP05 ecology and metabolic potential.

In the study, the team collected several thousand individual bacterial cells from three depths spanning the Saanich Inlet (off of Vancouver Island, British Columbia) oxygen gradient. Nearly 130 SUP05 single amplified genomes (SAGs) were recovered and sequenced at other institutions before being assembled, quality-checked

and annotated at the DOE JGI. Forty-two of the sequenced SUP05 SAGs were found to contain a total of 69 viruses representing five new genera. These viral sequences provided new reference genomes and enabled further inquiries on topics such as including lineage-specific viral infection and mortality estimates, biogeography and accessory metabolic gene potential. For example, using the new viral genomes as "hooks," the team queried 186 viral and microbial metagenomic datasets, many of which were generated through DOE JGI Community Science Program, to demonstrate that while SUP05 viruses were locally restricted to OMZs, they persisted in the environment over several years, with some viruses evolving over that time interval.

"This study represents the first of its kind, exploiting the unique strength of single-cell genomics to explore virus-host dynamics, including viral co-infections, in a completely cultivation-independent manner," noted Tanja Woyke, head of the DOE JGI Microbial Program and co-author of this study. "The resulting data provide a very robust foundation for future experimental work," she added. Woyke is also excited about the expansion of this work into dark matter branches of the microbial tree of life.

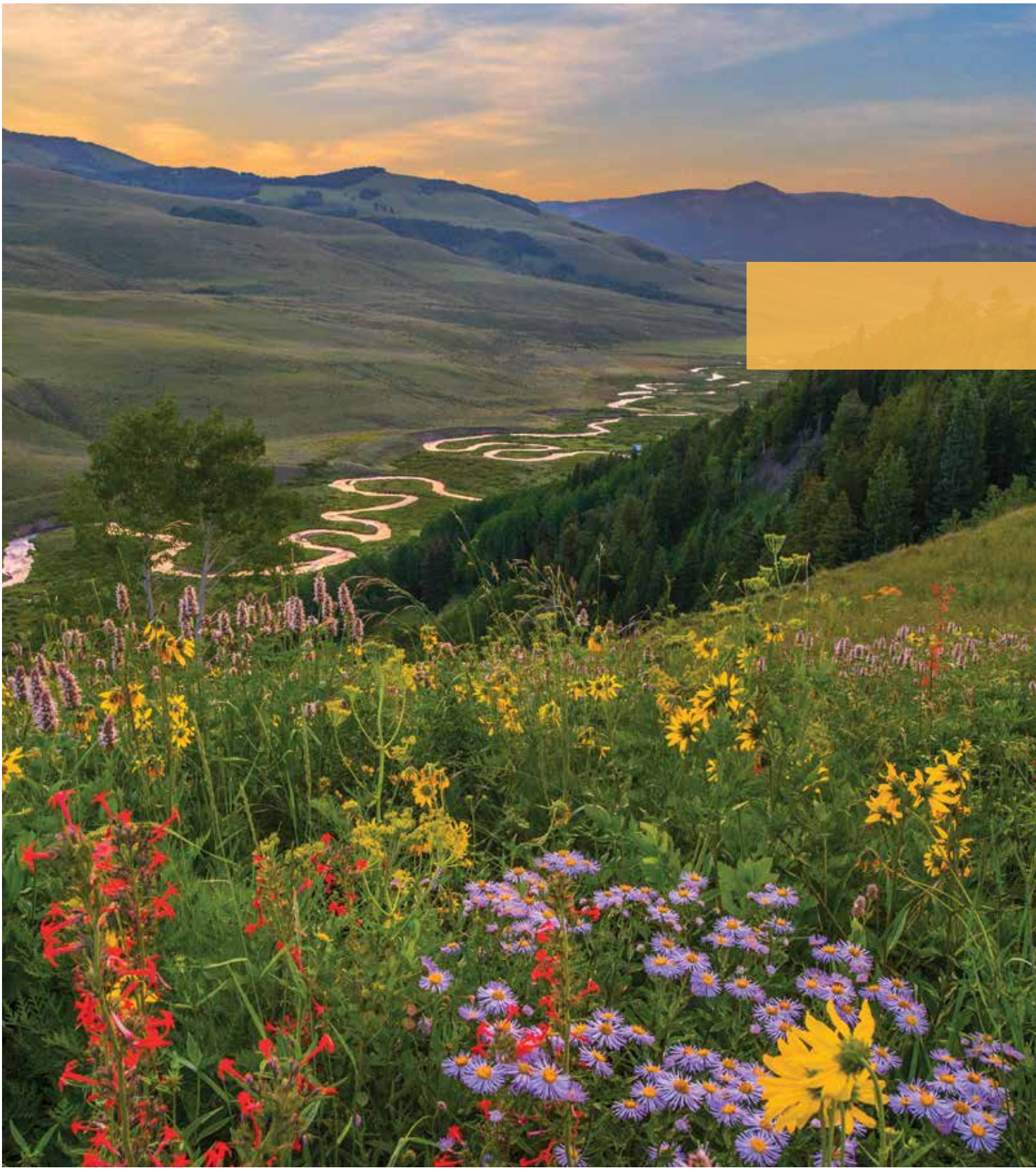
Genomic Gluttony of a Shrubbery

On a single island in the remote South Pacific grows the shrub *Amborella trichopoda*, which is the only plant in its family and genus. In a study published December 20, 2013, in *Science*, researchers from Indiana University, DOE JGI, Penn State University, and the Institute of Research for Development in New Caledonia described the genome sequence of *Amborella*'s mitochondria, the organelle that produces the plant's energy. In the process, they also detailed *Amborella*'s ability to acquire and retain whole mitochondrial genomes from green algae, mosses and other flowering plants.

"It swallowed whole genomes from other plants and algae as well as retained them in remarkably whole forms for eons," said Indiana University's Jeffrey Palmer, the study's senior author. The DNA that the *Amborella* mitochondria absorbed and retained through horizontal gene transfer amounts to at least one million base pairs, bringing the size of its mitochondrial genome to 3.9 million base pairs, several times that of typical plant mitochondrial genomes that contain around 500,000 base pairs.

The research provides strong evidence for the hypothesis that plant mitochondria can take up new traits by fusing with the mitochondria of other species. In *Amborella*, the mitochondria have ample opportunities to come into contact with mitochondria of other plants, for example epiphytes, plants that grow on other plants. When injured, *Amborella* often generates rapid growth at these locations, a virtual petri dish for mitochondria of different species to come into direct contact with each other. This, coupled with a low rate of losing mitochondrial genes over time, has created *Amborella*'s huge mitochondrial DNA glut.

The plant's mitochondria have acquired six genome equivalents of foreign DNA, marking the first time that an organelle has captured entire "foreign" genomes. *(Joel McNeal via PennState Flickr CC BY-NC-ND 2.0)*



Biogeochemistry

Retracing Early Citrus Cultivation Steps

Originally domesticated in Southeast Asia thousands of years ago before spreading throughout Asia, Europe, and the Americas via trade, citrus, the world's most widely cultivated fruit crop, is now under attack from an insidious emerging infectious disease that is destroying entire orchards. To help defend against citrus greening and other threats, researchers worldwide are mobilizing to apply genomic tools and approaches to understand how citrus varieties arose and how they respond to disease and other stresses.

The DOE JGI contributed to the citrus pilot project, and was part of teams that generated the sweet orange and Clementine mandarin genomes. These early contributions helped lay the foundation for the study published in the June 2014 edition of *Nature Biotechnology*. An international consortium of researchers from the United States, France, Italy, Spain, and Brazil analyzed and compared the genome sequences of 10 diverse citrus varieties, including sweet and sour orange along with several important mandarin and pummelo cultivars. The consortium found that these diverse varieties are derived from two wild citrus species that diverged in Southeast Asia over five million years ago.

One of these wild species gave rise to cultivated pummelo, while mandarins, in contrast, were found to be genetic mixtures of a second species and pummelo. Sweet orange, the most widely grown citrus variety worldwide, turned out to be a complex genetic hybrid of mandarin and pummelo. Seville or sour orange, commonly used in marmalade, was found to be an unrelated interspecific hybrid.

By inferring the past hybridization events that gave rise to these common citrus varieties, the team hopes to enable strategies for improving citrus, including resistance to greening and other diseases. Revisit Fred Gmitter's citrus talk at the 2012 DOE JGI Genomics of Energy & Environment Meeting at http://bit.ly/JGI7Gmitter.

Common Bean Genome for Crop Improvements

All plants require nitrogen to thrive, and nitrogen fixation is the process by which atmospheric nitrogen is converted into ammonia. However, many agricultural lands are deficient in nitrogen, leading farmers to rely on fertilizers to supply the needed nutrient for their crops. The DOE Office of Science has targeted research into the common bean because of its importance in enhancing nitrogen use efficiency for sustainability of bioenergy crops, and for increasing plant resilience and productivity with fewer inputs, on marginal lands, and in the face of the changing climate and environment.

Legumes such as the common bean and soybean can form symbiotic relationships with nitrogen-fixing bacteria.

In a report published online on June 8, 2014, in the journal *Nature Genetics*, a team led by researchers at the University of Georgia, the DOE JGI, the HudsonAlpha Institute for Biotechnology, and North Dakota State University sequenced and analyzed the genome of the common bean, *Phaseolus vulgaris*. The project was supported by the DOE, the National Institute of Food and Agriculture, and the U.S. Department of Agriculture.

For the study, the team sequenced and assembled a 473-million-base-pair genome of the common bean. Though it is thought to have originated in Mexico, the common bean was domesticated separately at two different geographic locations in Mesoamerica and the Andes, diverging from a common ancestral wild population more than 100,000 years ago. The team then compared sequences from pooled populations representing these regions, finding only a small fraction of shared genes. This indicated that different events had been involved in the domestication process at each location.

The team then compared the high-quality common bean genome against the sequence of its most economically important relative, soybean, which had also been done at the DOE JGI. They found evidence of synteny, in which a gene in one species is present in another. They also noted that the common bean's genome had evolved more rapidly than soybean's since diverging from their last common ancestor nearly 20 million years ago. Understanding how the common bean responds to biotic and abiotic stresses will provide researchers with information about genomic regions that have been targeted either during domestication or early improvement. These findings in turn indicate areas of interest for future crop improvement efforts.

Phil McClean presented on the common bean genome project at the 2014 DOE JGI Genomics of Energy & Environment Meeting. The video is available at http://bit.ly/JGIUM9McClean.

Identifying Stop Codon Reassignments in the Wild

"All along, we presumed that the code or vocabulary used by organisms was universal, applying to all branches of the tree of life, with vanishingly few exceptions," said DOE JGI Director Eddy Rubin. "We have now confirmed that this just isn't so."

It has long been assumed that there is only one "canonical" genetic code that all organisms use to build proteins. DOE JGI researchers who discovered large numbers of exceptions from the canonical genetic code and published

the work in the May 23, 2014, edition of the journal *Science* have challenged this paradigm. This research was conducted under the DOE JGI's continuing effort to explore the biological frontier known as "microbial dark matter."

There are 64 codons, and all but three of these triplets encode actual amino acids — the building blocks of all proteins. The remaining three are "stop codons," which bring the molecular machinery to a halt, terminating the translation of RNA into protein. Each has a given name: Amber, Opal and Ochre. When an organism's machinery reads the instructions in the DNA, builds a protein composed of amino acids, and reaches Amber, Opal or Ochre, this triplet would signal that it has arrived at the end of a protein.

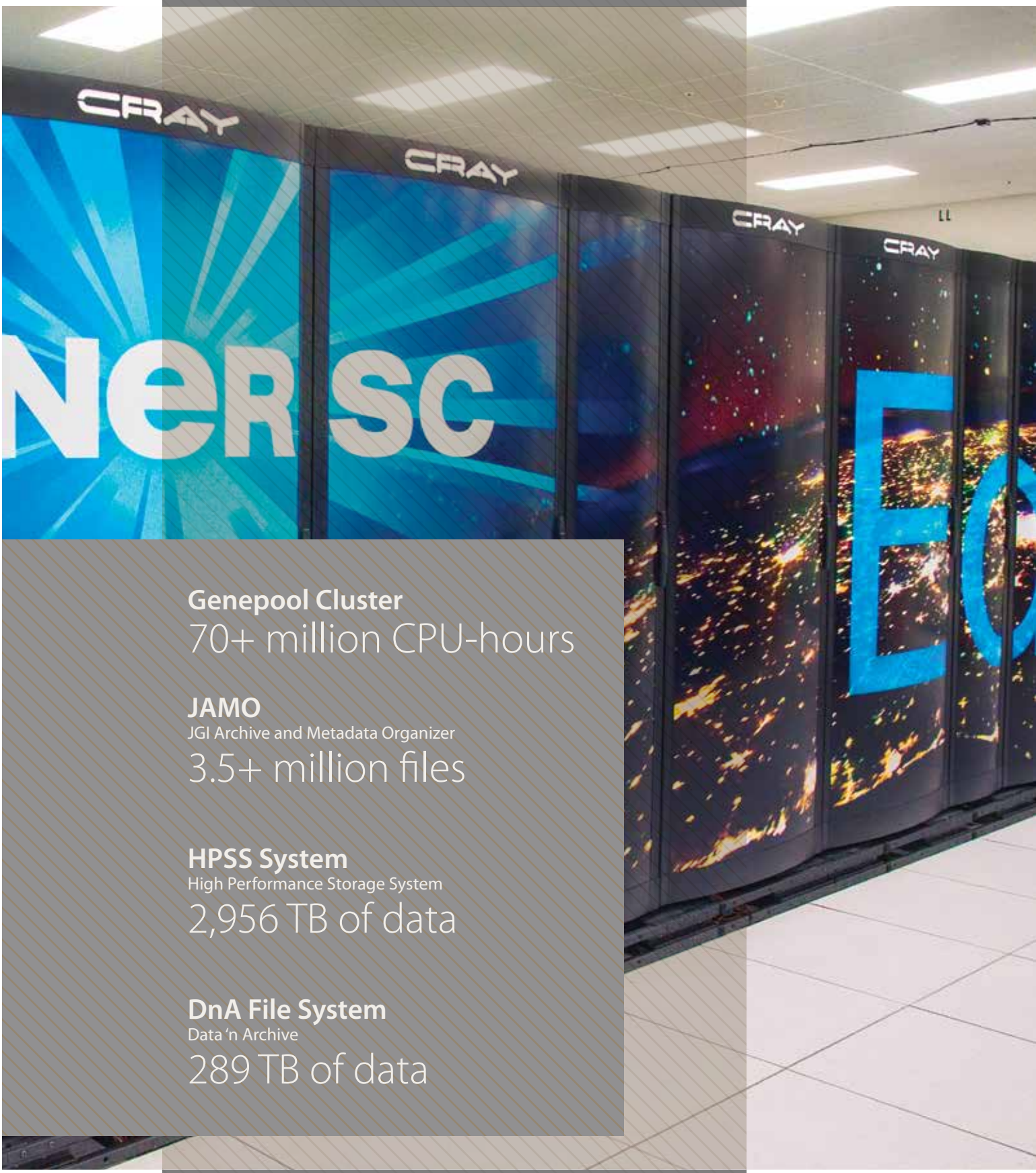
The particular observation that caught the team's interest in looking for breakdowns in the canonical genetic code was when the study's lead investigator came across an anomaly: bacteria with extraordinarily short genes of only 200 base pairs in length. Typically, genes from microbes are about 800-900 base pairs long. When one of the three stop codons was "reinterpreted" to add an amino acid, the odd proteins suddenly looked of normal length. This was quickly experimentally confirmed, indicating that some microbes "read" a particular stop codon to instruct the machinery to add an amino acid.

Following this finding they wanted to see how frequently this occurs in nature and looked for similar occurrences in enormous amounts of sequence data from uncultured microbes. Computationally they sifted through a massive "haystack" of sequence data, 5.6 trillion letters of genetic code (the equivalent of nearly 2,000 human genomes). These came from over 1,700 samples sourced from far-flung and esoteric locations that span the globe — marine, freshwater, and terrestrial environments — to those much closer to home and more prosaic — from the human mouth and gut.

"We were surprised to find that an unprecedented number of bacteria in the wild possess these codon reassignments, from 'stop' to amino-acid encoding 'sense,' up to 10 percent of the time in some environments," said Rubin.

This unexpected discovery may have implications for a range of DOE interests, among them virus microbe interactions, genomic alteration technologies, and biocontainment.

With BER support for Subsurface Systems Scientific Focus Area 2.0 (SFA 2.0), Berkeley Lab Earth Sciences Division and DOE JGI scientists are conducting long-term subsurface research on contaminant mobility near Rifle, Colorado.



Genepool Cluster
70+ million CPU-hours

JAMO
JGI Archive and Metadata Organizer
3.5+ million files

HPSS System
High Performance Storage System
2,956 TB of data

DnA File System
Data 'n Archive
289 TB of data

Computational Infrastructure

The DOE JGI's capacity as a next-generation genomics user facility has generated petabytes of data and analysis. In 2014, our genome sequence data alone consisted of over 100 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 its 8,000+ core computing cluster Genepool, 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 2014, QA/QC group member Michael Barton developed and deployed the site nucleotid.es for the consistent evaluation of genome assemblers. Barton took advantage of the cutting-edge Docker container software to build containers for individual assemblers that can run anywhere Docker is supported. The nucleotid.es site is currently using compute resources in the Amazon cloud, but NERSC is collaborating with the DOE JGI to support this software on the Genepool and Cray compute systems. This collaboration has the potential to change the way scientists share software, evaluate computational tools and make scientific computing more reproducible.

Also in 2014, the DOE JGI used more than 10 million CPU hours on NERSC's petascale supercomputers, Hopper and Edison. These calculations could not have been completed on the Genepool cluster. In 2014, the DOE JGI broadened efforts to explore high-performance computing for bioinformatics. A notable outcome from these initiatives was the development of UPC Meraculous, an assembler for plants developed under the guidance of Dan Rokhsar (DOE JGI Chief Informatics Officer) and Katherine Yelick (Berkeley Lab Associate Lab Director for Computing Sciences). The new version of Meraculous allows for the assembly of the human genome in minutes across thousands of processors on Edison, NERSC's 2 petaflop supercomputer.

In 2015, NERSC will be moving to a new, state-of-the-art facility on the main Berkeley Lab site, and the computational infrastructure of the DOE JGI will also make the move. The Computational Research and Theory (CRT) Facility will be on the forefront of high-performance supercomputing research and will be DOE's most efficient facility of its kind. Designed to take advantage of the cool Berkeley climate, the CRT is anticipated to set a new standard in energy efficiency for high-performance computing.

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.

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)
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
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
GenePRIMP	Gene PRediction IMprovement Pipeline
GLBRC	Great Lakes Bioenergy Research Center
GOLD	Genomes OnLine Database
HPC	High Performance Computing
HPSS	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
MGM	Microbial Genomics & Metagenomics
NERSC	National Energy Research Scientific Computing Center
NREL	National Renewable Energy Laboratory
ORNL	Oak Ridge National Laboratory
PMO	Project Management Office
PNNL	Pacific Northwest National Laboratory
SAC	Scientific Advisory Committee
WIP	Work Initiation Process

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.

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.

ChIP-Seq: A method of analyzing protein interactions with DNA.

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 fluidic 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.

Mass spectrometry: An analytical technique that can identify unknown compounds through their molecular weight. It can also be used to determine a molecule's structure and chemical properties.

Metabolomics: A comparison of biological samples based on their metabolite profiles.

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.

Nanopore sequencing: Nucleotide sequencing method that utilizes nanopores, which are small nanometer holes in a membrane. An electric potential is applied to the membrane and an electric current will drive ions through the nanopore, and as molecules pass through they will cause characteristic disruptions to the current and can be used to identify nucleotides in a DNA strand as it passes through the nanopore.

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.

Proteomics: The large-scale study of proteins, as well as their structures and functions.

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: Proprietary sequencing technique used by Pacific Biosciences (PacBio) for 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
Baliga, Nitin	Institute of Systems Biology	Genome-wide analysis of chromatin accessibility and miRNA-mediated transcriptional regulation of lipid accumulation in *Chlamydomonas reinhardtii*
Banfield, Jill	University of California, Berkeley	Genome-resolved metagenomic analysis of microbial function in the subsurface
Banfield, Jill	University of California, Berkeley	Tree-driven diel microbial carbon dynamics in the vadose zone
Barrasa, Jose Maria	University of Alcala, Spain	Study of the lignocellulolytic machinery in saprobic wood and leaf litter degrading Agaricales
Busby, Posy	University of Washington	Mechanisms of interaction in the foliar fungal microbiome of *Populus trichocarpa*
Cadillo-Quiroz, Hinsby	Arizona State University	Microbial composition and metagenomic functional potential across tropical peatlands: Comparative evaluation and modeling of C decomposition to greenhouse gases
Campbell, Barbara	Clemson University	Biogeochemical cycling links between terrestrial and marine systems
Cavicchioli, Rick	University of New South Wales, Australia	Seasonal variation in Antarctic microbial communities: ecology, stability, and susceptibility to ecosystem change
Chistoserdova, Ludmila	University of Washington	Methane oxidation as a community function: defining partnerships and strategies through sequencing metagenomes and metatranscriptomes of laboratory manipulated microcosms
DiFazio, Steven	West Virginia University	Sex determination in the Salicaceae
Frank, Carolin	University of California, Merced	Diazotrophic aboveground endophytes in native pines, poplar, and willow

PROPOSER	AFFILIATION	PROJECT DESCRIPTION
Fredrickson, Jim	PNNL	Spatio-temporal functional profiling in model microbial communities
Hallam, Steven	University of British Columbia, Canada	Microbial engines driving organic matter transformations in the dark ocean: An integrated biological and chemical perspective
James, Timothy	University of Michigan	Revealing the ecological function of uncultured fungal dark matter in freshwater ecosystems using single cell genomics
Klassen, Jonathan	University of Connecticut	Metagenomic mining of natural product diversity and understanding its contribution to ecosystem function in the cellulolytic fungus-growing ant symbiosis
Klenk, Hans-Peter; Goker, Markus	University of Newcastle, UK; DSMZ, Germany	Exploiting the genomes of the Actinobacteria: plant growth promoters and producers of natural products and energy relevant enzymes united in a taxonomically unresolved phylum
Lamendella, Regina	Juniata College	Systems biology approach to fracking for environmental monitoring
Laudencia-Chingcuanco, Debbie	USDA-ARS	Creating a genome-wide sequence-indexed collection of grass mutants
Liu, Wen-Tso	University of Illinois	Shedding light on the anaerobic wastewater treatment "black-box" in anthropogenic carbon cycling: Exploring the uncharted ecological function of uncultured microbial taxa through next-generation sequencing technology
Loper, Joyce	USDA-ARS	Exploring the genomic diversity of the *Pseudomonas fluorescens* group
Medina-Munoz, Monica	Penn State University	How do coral hosts communicate with their associated microbial community?

Community Science Program Projects *(continued)*

PROPOSER	AFFILIATION	PROJECT DESCRIPTION
Peay, Kabir	Stanford University	Coprophilous fungi as a model system for understanding the metagenomics of carbon cycling in microbial eukaryote communities
Simon, Holly	Oregon Health & Science University	A systems approach to evaluate physical constraints and microbial controls on fluxes of nutrients and energy in a coastal ecosystem
Sorek, Rotem	Weizmann Institute of Science, Israel	Comparative genomics of a single organism: In search of acquired immunity in trees
Swaminathan, Kankshita	University of Illinois	Understanding variance in allele specific expression in the polyploid Saccharinae
Thon, Michael	University of Salamanca, Spain	Evolution and adaptation of carbohydrate utilization in the *Colletotrichum acutatum* species complex
Treseder, Kathleen	University of California, Irvine	Genomes and transcriptomes of decomposer fungi responding to warming in Alaskan boreal forest
Wildermuth, Mary	University of California, Berkeley	Comparative genomics of powdery mildews and associated host plants
Wosten, Han	Utrecht University	Functional genomics of lignocellulose degradation by Agaricomycete fungi
Wurzbacher, Christian	IGB, Germany	Whole genome sequencing of aquatic fungi responsible for the degradation of recalcitrant substrates in liquid environments
Young, Erica	University of Wisconsin	Metagenome and metatranscriptome of complex algal communities growing in wastewater: Bioremediation, nutrient transformation, carbon sequestration
Zhang, Baohong	East Carolina University	*Panicum virgatum* small RNA sequencing to identify gene expression changes related to biofuel traits

JGI/EMSL Collaborative Science Initiative Projects

The DOE JGI and the Environmental Molecular Sciences Laboratory (EMSL) accepted 12 projects submitted during the 2014 call for Collaborative Science Initiative proposals. The call represents a unique opportunity for researchers to combine the power of genomics and molecular characterization in one research project to help advance the missions of the Department of Energy's Office of Biological and Environmental Research. The accepted proposals began on October 1, 2014, providing the researchers with access to the capabilities of both user facilities and datasets beyond what could be generated by either facility alone.

PROPOSER	AFFILIATION	PROJECT DESCRIPTION
Colwell, Frederick	Oregon State University	Integrated biogeochemical modeling of microbial consortia mediating anaerobic oxidation of methane in dynamic methane hydrate-bearing sediments
Crump, Byron	Oregon State University	Decoding DOM degradation: How does carbon source and sunlight exposure alter microbial metabolism and expression of genome-encoded metabolic degradation of permafrost organic matter?
Cumming, Jonathan	West Virginia University	Mapping the metabolism of nutrient and carbon exchange in the plant-microbe symbiosis
de Vries, Ronald	CBS-KNAW Fungal Biodiversity Centre	Dissecting intraspecies diversity in fungal wood decay
DiChristina, Thomas	Georgia Institute of Technology	Sensing external metals by outer membrane beta-barrel proteins
Lebeis, Sarah	University of Tennessee	Uncovering the composition and function of the aquatic microbiome for duckweeds
Luthey-Schulten, Zaida	University of Illinois at Urbana-Champaign	Quantifying differential expression and identifying bottlenecks in methanogenic pathways
Magnuson, Jon	Pacific Northwest National Laboratory	Elucidating the influences of engineered n-glycosylation motifs in bacterial biomass hydrolyzing enzymes upon heterologous and native gene expression, secretion and degradation in *Aspergillus niger*

JGI/EMSL Collaborative Science Initiative Projects *(continued)*

PROPOSER	AFFILIATION	PROJECT DESCRIPTION
Rich, Virginia	University of Arizona	Systems-level insights into carbon transformations in thawing permafrost by parallel high-resolution organic matter and microbial community characterizations
Stegen, James	Pacific Northwest National Laboratory	Coupling microbial communities to carbon and contaminant biogeochemistry in the groundwater-surface water interaction zone
Vilgalys, Rytas	Duke University	Integrated genomic/transcriptomic/secretomic study of plant-fungal interactions between pines and their symbiotic ectomycorrhizal fungi in the mushroom genus *Suillus*
Wrighton, Kelly	The Ohio State University	Microbial controls on biogeochemical cycling in deep subsurface shale carbon reservoirs

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, *Broad Institute (Chair)*

Steve Briggs, *University of California, San Diego*

Jeff Dangl, *University of North Carolina*

David Dooling, *Monsanto*

Joe Ecker, *Salk Institute*

Claire M. Fraser, *University of Maryland*

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

Glenn Kubiak, *Lawrence Berkeley National Laboratory*

Mary Ann Moran, *University of Georgia*

Rick Myers, *HudsonAlpha Institute for Biotechnology*

Julian Parkhill, *The Sanger Institute*

James Tiedje, *Michigan State University*

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

Informatics Advisory Committee (IAC)

Members

Adam Arkin, *Lawrence Berkeley National Laboratory*

David Dooling, *Monsanto*

Saul Kravitz, *The Howard Hughes Medical Institute (HHMI)*

Stan Letovsky, *SynapDx*

Jill Mesirov, *Broad Institute (IAC Chair)*

Granger Sutton, *J. Craig Venter Institute*

Kathy Yelick, *Lawrence Berkeley National Laboratory*

Fungal Program User Advisory Committee

Members

Scott Baker, *Pacific Northwest National Laboratory*

Randy Berka, *Novozymes*

Ronal de Vries, *CBS (Netherlands)*

Audrey Gasch, *Great Lakes Bioenergy Research Center*

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

Stephen Goodwin, *Purdue University*

David Hibbett, *Clark University*

Francis Martin, *INRA (France)*

Joseph Spatafora, *Oregon State University*

Adrian Tsang, *Concordia University (Canada)*

Plant Program User Advisory Committee

Members

Jeff Dangl, *University of North Carolina*

Joe Ecker, *The Salk Institute for Biological Studies*

Samuel Hazen, *University of Massachusetts, Amherst*

Eva Huala, *Carnegie Institute/TAIR*

Sabeeha Merchant, *University of California, Los Angeles*

Thomas Mitchell-Olds, *Duke University*

Stephen Moose, *University of Illinois*

Gary Stacey, *University of Missouri*

Prokaryotic Super Program Advisory Committee

Members

Jill Banfield, *University of California, Berkeley*

Cameron Currie, *University of Wisconsin*

Ed DeLong, *University of Hawaii*

Jonathan Eisen, *University of California, Davis*

Jed Fuhrman, *University of Southern California*

George Garrity, *Michigan State University (Chair)*

Steve Hallam, *University of British Columbia*

Trina McMahon, *University of Wisconsin-Madison (Vice Chair)*

Folker Meyer, *Argonne National Laboratory*

Mary Ann Moran, *University of Georgia*

Nancy Moran, *University of Texas at Austin*

Karen Nelson, *J. Craig Venter Institute*

Rich Roberts, *NEB*

Doug Rusch, *Indiana University Bloomington*

Ramunas Stepanauskas, *Bigelow Laboratory for Ocean Sciences*

Niels van der Lelie, *FMC Corporation*

Phil Hugenholtz, *University of Queensland (Australia)*

Synthetic Biology Advisory Committee

Members

Sunil Chandran, *Amyris, Inc.*

Michael Fischbach, *University of California, San Francisco*

Jay Keasling, *Lawrence Berkeley National Laboratory*

Megan Palmer, *University of California, Berkeley*

Jeffrey Sampson, *Agilent Technologies, Inc.*

Jay Shendure, *University of Washington*

Christina Smolke, *Stanford University*

Appendix E
2014 Genomics of Energy and Environment Meeting

Keynote Speakers

Stephen Quake (left), shown here with DOE JGI Director Eddy Rubin (right), a bioengineering professor at Stanford University, delivered the opening keynote address on single-cell genomics. Renowned as an inventive scientist and creator of new businesses, Quake pioneered automated single-molecule gene sequencing, large-scale integrated microfluidic circuits, and co-founded half a dozen companies, including leading microfluidic chipmaker Fluidigm. Now mentor to a new generation of innovators, his talk focused on the confluence of three technologies that has made it possible for single-cell sequencing to be accessible enough to be named *Nature Methods*' 2013 Method of the Year. Quake's talk is available online at http://bit.ly/JGIUM9_Quake.

Annalee Newitz, editor of the highly acclaimed San Francisco-based blog *io9*, shared her thoughts on how genome science might play a role in surviving the ongoing sixth mass extinction in her closing keynote. Drawing from her latest book on how humans will survive a mass extinction, she listed alternative fuels, biofuels, solar fuels, and artificial photosynthesis as adaptations that could make cities more livable and more carbon neutral. She also speculated on biological tools that could be harnessed to grow future buildings from seeds. Newitz's talk is available online at http://bit.ly/JGIUM9_Newitz.

Other Featured Speakers

(in order of appearance):

Martin Ackermann, *ETH Zurich (Switzerland)*

Mary Berbee, *University of British Columbia (Canada)*

Rytas Vilgalys, *Duke University*

Masaru Nobu, *University of Illinois at Urbana-Champaign*

Eddy Rubin, *DOE Joint Genome Institute*

Nicole Dubilier, *Max Planck Institute (Germany)*

Erin Nuccio, *Lawrence Livermore National Laboratory*

Michael Fischbach, *University of California, San Francisco*

Anne Osbourn, *John Innes Centre (United Kingdom)*

Maria Mercedes-Roca, *Zamorano Pan-American Agriculture School (Honduras)*

Scott Hickey, *University of California, Berkeley*

Kelly Matzen, *Oxitec, Ltd. (United Kingdom)*

Phil McClean, *North Dakota State University*

Cameron Coates, *University of California, San Diego*

June Medford, *Colorado State University*

Steve Rounsley, *University of Arizona*

Kankshita Swaminathan, *Energy Biosciences Institute*

Joshua Coomey, *University of Massachusetts, Amherst*

Gerald Tuskan, *Oak Ridge National Laboratory*

Pamela Ronald, *University of California, Davis*

Len Pennacchio, *DOE Joint Genome Institute*

Adam Arkin, *Lawrence Berkeley National Laboratory*

Itai Sharon, *University of California, Berkeley*

David Berry, *University of Vienna (Austria)*

Dan Voytas, *University of Minnesota*

Learn more about the meeting talks at
http://usermeeting.jgi.doe.gov/past-meetings/2014-agenda/

Videos of the talks are available on DOE JGI's YouTube channel at
http://bit.ly/JGIUM9videos
and on our SciVee channel at
http://www.scivee.tv/node/62351

Appendix F
2014 Publications

October 2013 – September 2014

Anderson I et al. Genome sequence of *Frateuria aurantia* type strain (Kondô 67(T)), a xanthomonade isolated from *Lilium auratium* Lindl. *Stand Genomic Sci.* 2013 Oct 2;9(1):83-92. doi: 10.4056/sigs.4338002. eCollection 2013 Oct 16.

Barling A et al. A detailed gene expression study of the *Miscanthus* genus reveals changes in the transcriptome associated with the rejuvenation of spring rhizomes. *BMC Genomics.* 2013 Dec 9;14:864. doi: 10.1186/1471-2164-14-864.

Barriuso J et al. Fungal genomes mining to discover novel sterol esterases and lipases as catalysts. *BMC Genomics.* 2013 Oct 18;14:712. doi: 10.1186/1471-2164-14-712.

Beck DA et al. The expanded diversity of methylophilaceae from Lake Washington through cultivation and genomic sequencing of novel ecotypes. *PLoS One.* 2014 Jul 24;9(7):e102458. doi: 10.1371/journal.pone.0102458. eCollection 2014.

Beissinger TM et al. A genome-wide scan for evidence of selection in a maize population under long-term artificial selection for ear number. *Genetics.* 2014 Mar;196(3):829-40. doi: 10.1534/genetics.113.160655.

Biddle AS et al. The complete genome sequence of *Clostridium indolis* DSM 755T. *Stand. Genomic Sci.* 2014 Mar 18;9(3):1089-104. doi: 10.4056/sigs.5281010. eCollection 2014.

Blaby IK et al. The *Chlamydomonas* genome project: A decade on. *Trends Plant Sci.* 2014 Jun 17. pii: S1360-1385(14)00143-5. doi: 10.1016/j.tplants.2014.05.008. [Epub ahead of print] Review.

Boscaro V et al. *Polynucleobacter necessarius*, a model for genome reduction in both free-living and symbiotic bacteria. *Proc Natl Acad Sci U S A.* 2013 Nov 12;110(46):18590-5. doi: 10.1073/pnas.1316687110.

Breider S et al. Genome sequence of *Leisingera nanhaiensis* strain DSM 24252T isolated from marine sediment. *Stand. Genomic Sci.* 2014 Jan 25;9(3):687-703. doi:10.4056/sigs.3828824. eCollection 2014.

Buddruhs N et al. Complete genome sequence of the marine methyl-halide oxidizing *Leisingera methylohalidivorans* type strain (DSM 14336(T)), a representative of the Roseobacter clade. *Stand Genomic Sci.* 2013 Oct 4;9(1):128-41. doi: 10.4056/sigs.4297965. eCollection 2013 Oct 16.

Busquets A et al. Draft genome sequence of *Pseudomonas azotifigens* strain DSM 17556T (6H33bT), a nitrogen fixer strain isolated from a compost pile. *Genome Announc.* 2013 Oct 31;1(6). pii: e00893-13. doi: 10.1128/genomeA.00893-13.

Byrne RT et al. Evolution of extreme resistance to ionizing radiation via genetic adaptation of DNA repair. *Elife.* 2014 Jan 1;3:e01322. doi: 10.7554/eLife.01322.

Cai F et al. Engineering bacterial microcompartment shells: chimeric shell proteins and chimeric carboxysome shells. *ACS Synth Biol.* 2014 Aug 27. doi: 10.1021/sb500226j [Epub ahead of print]

Cameron JC et al. Biogenesis of a bacterial organelle: The carboxysome assembly pathway. *Cell.* 2013 Nov 21;155(5):1131-40. doi: 10.1016/j.cell.2013.10.044.

Campbell AG et al. Diversity and genomic insights into the uncultured *Chloroflexi* from the human microbiota. *Environ Microbiol.* 2014 Sep;16(9):2635-43. doi: 10.1111/1462-2920.12461.

Chang PJ et al. Force measurements of a magnetic micro actuator proposed for a microvalve array. *J. Micromech. Microeng.* 2014 Feb 19;24,034005. doi:10.1088/0960-1317/24/3/034005.

Childs KL et al. Generation of transcript assemblies and identification of single nucleotide polymorphisms from seven lowland and upland cultivars of switchgrass. *Plant Gen.* 2014 Jun 2. doi:10.3835/plantgenome2013.12.0041.

Cho Y et al. Transcriptional responses of the Bdtf1-deletion mutant to the phytoalexin brassinin in the necrotrophic fungus *Alternaria brassicicola. Molecules.* 2014 Jul 24;19(8):10717-32. doi: 10.3390/molecules190810717.

Cimermancic P et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell.* 2014 Jul 17;158(2):412-21. doi: 10.1016/j.cell.2014.06.034.

Clingenpeel S et al. Effects of sample treatments on genome recovery via single-cell genomics. *ISME J.* 2014 Jun 13. doi: 10.1038/ismej.2014.92. [Epub ahead of print]

Coleman-Derr D et al. Building the crops of tomorrow: advantages of symbiont-based approaches to improving abiotic stress tolerance. *Front Microbiol.* 2014 Jun 6;5:283. doi: 10.3389/fmicb.2014.00283. eCollection 2014.

Collavino M et al. *nifH* pyrosequencing reveals the potential for location-specific soil chemistry to influence N_2-fixing community dynamics. *Environ Microbiol.* 2014 Mar 19. doi: 10.1111/1462-2920.12423

Deng K et al. Rapid kinetic characterization of glycosyl hydrolases based on oxime derivatization and nanostructure-initiator mass spectrometry (NIMS). *ACS Chem Biol.* 2014 Jul 18;9(7):1470-9. doi: 10.1021/cb5000289.

Dogs M 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 Oct 3;9(1):142-59. doi: 10.4056/sigs.4287962. eCollection 2013 Oct 16.

Dogs M et al. Genome sequence of *Phaeobacter inhibens* type strain (T5(T)), a secondary metabolite producing representative of the marine *Roseobacter* clade, and emendation of the species description of *Phaeobacter inhibens. Stand Genomic Sci.* 2013 Dec 15;9(2):334-50. doi: 10.4056/sigs.4448212. eCollection 2013 Dec 20.

Duanmu D et al. Marine algae and land plants share conserved phytochrome signaling systems. *Proc Natl Acad Sci U S A.* 2014 Sep 29. pii: 201416751. doi: 10.1073/pnas.1416751111. [Epub ahead of print]

Durham B et al. Draft genome sequence of marine alphaproteobacterial strain HIMB11, the first cultivated representative of a unique lineage within the *Roseobacter* clade possessing an unusually small genome. *Stand. Genomic Sci.* 2014 Mar 15;9(3): 632-645. doi:10.4056/sigs.4998989. eCollection 2014.

Erbilgin O et al. Characterization of a planctomycetal organelle: A novel bacterial microcompartment for the aerobic degradation of plant saccharides. *Appl EnvironMicrobiol.* 2014 Apr;80(7):2193-205. doi: 10.1128/AEM.03887-13. Epub 2014 Jan 31.

Evans LM et al. Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nat Genet.* 2014 Oct; 46 (10):1089-96. doi: 10.1038/ng.3075. Epub 2014 Aug 24.

Field D et al. Genomic standards consortium projects. *Stand. Genomic Sci.* 2014 Feb 15;9(3):599-601. doi: 10.4056/sigs.5559680.

Galardini M et al. Permanent draft genome sequences of the symbiotic nitrogen fixing *Ensifer meliloti* strains BO21CC and AK58. *Stand Genomic Sci.* 2013 Dec 15;9(2):325-doi: 10.4056/sigs.3797438. eCollection 2013 Dec 20.

Gallo A et al. Identification and characterization of the polyketide synthase involved in ochratoxin: A biosynthesis in *Aspergillus carbonarius. Int J Food Microbiol.* 2014 Jun 2;179:10-7. doi: 10.1016/j.ijfoodmicro.2014.03.013. Epub 2014 Mar 18.

Gan F et al. Extensive remodeling of a cyanobacterial photosynthetic apparatus in far-red light. *Science.* 2014 Sep 12;345(6202):1312-7. doi: 10.1126/science.1256963.

Gao D et al. Annotation and sequence diversity of transposable elements in common bean *(Phaseolus vulgaris). Front Plant Sci.* 2014 Jul 11;5:339. doi: 10.3389/fpls.2014.00339. eCollection 2014.

Ghylin TW et al. Comparative single-cell genomics reveals potential ecological niches for the freshwater acI Actinobacteria lineage. *ISME J.* 2014 Aug 5. doi: 10.1038/ismej.2014.135. [Epub ahead of print]

Göker M et al. Genome sequence of the *Thermotoga thermarum* type strain (LA3(T)) from an African solfataric spring. *Stand. Genomic Sci.* 2014 May 15;9(3):1105-17. doi:10.4056/sigs.3016383. eCollection 2014.

Göker M et al. Genome sequence of the mud-dwelling archaeon *Methanoplanus limicola* type strain (DSM 2279(T)), reclassification of *Methanoplanus petrolearius* as *Methanolacinia petrolearia* and emended descriptions of the genera *Methanoplanus* and *Methanolacinia. Stand. Genomic Sci.* 2014 Mar 15;9(3):1076-88. doi:10.4056/sigs.5138968. eCollection 2014.

Golberg A et al. Proposed design of distributed macroalgal biorefineries: Thermodynamics, bioconversion technology, and sustainability implications for developing economies. *Biofuels Bioproducts & Biorefining-Biofpr.* 2014 Jan;8(1):67-82. doi: 10.1002/bbb.1438.

Golberg A et al. Cloud-enabled microscopy and droplet microfluidic platform for specific detection of *Escherichia coli* in water. *PLoS One.* 2014 Jan 27;9(1):e86341. doi: 10.1371/journal.pone.0086341. eCollection 2014.

Gordon SP et al. Genome diversity in *Brachypodium distachyon*: Deep sequencing of highly diverse inbred lines. *Plant J.* 2014 Aug;79(3):361-74. doi: 10.1111/tpj.12569.

Gostinçar C et al. Genome sequencing of four *Aureobasidium pullulans* varieties: Biotechnological potential, stress tolerance, and description of new species. *BMC Genomics.* 2014 Jul 1;15(1):549. [Epub ahead of print]

Grigoriev IV et al. MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 2014 Jan;42(Database issue):D699-704. doi: 10.1093/nar/gkt1183.

Haider B et al. Omega: An overlap-graph de novo assembler for metagenomics. *Bioinformatics.* 2014 Oct;30(19):2717-22. doi: 10.1093/bioinformatics/btu395. Epub 2014 Jun 19.

Han JI et al. Genome of the root-associated plant growth-promoting bacterium *Variovorax paradoxus* strain EPS. *Genome Announc.* 2013 Oct 24;1(5). pii: e00843-13. doi: 10.1128/genomeA.00843-13.

Hawley ER et al. Metagenomic analysis of microbial consortium from natural crude oil that seeps into the marine ecosystem offshore Southern California. *Stand. Genomic Sci.* 2014 Jan 2;9(3):1259-74. doi:10.4056/sigs.5029016.

Hawley ER et al. Metagenomic sequencing of two salton sea microbiomes. *Genome Announc.* 2014 Jan 23;2(1). pii: e01208-13. doi: 10.1128 genomeA.01208-13.

Hawley ER et al. Metagenomes from two microbial consortia associated with Santa Barbara seep oil. *Mar Genomics.* 2014 Jun 20. pii: S1874-7787(14)00074-9. doi: 10.1016/j.margen.2014.06.003. [Epub ahead of print]

Hedlund BP et al. Impact of single-cell genomics and metagenomics on the emerging view of extremophile "microbial dark matter." *Extremophiles.* 2014 Sep;18(5):865-75. doi: 10.1007/s00792-014-0664-7.

Heins RA et al. Phylogenomic guided identification of industrially relevant GH1 β-glucosidases through DNA synthesis and nanostructure-initiator mass spectrometry. *ACS Chem Biol.* 2014 Sep 19;9(9):2082-91. doi: 10.1021/cb500244v.

Heller et al. ARBitrator: A software pipeline for on-demand retrieval of autocurated nifH sequences from GenBank. *Bioinformatics.* 2014 Oct 15; 30(20): 2883-90. doi: 10.1093/bioinformatics/btu417. Epub 2014 Jul 2.

Hellsten U et al. Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. *Proc Natl Acad Sci U S A.* 2013 Nov 26;110(48):19478-82. doi: 10.1073/pnas.1319032110.

Hess J et al. Transposable element dynamics among asymbiotic and ectomycorrhizal *Amanita* fungi. *Genome Biol Evol.* 2014 Jun 12;6(7):1564-78. doi: 10.1093/gbe/evu121.

Hillson NJ. j5 DNA assembly design automation. *Methods Mol Biol.* 2014;1116:245-69. doi: 10.1007/978-1-62703-764-8_17.

Hirsch CN et al. Insights into the effects of long-term artificial selection on seed size in maize. *Genetics.* 2014 Sep;198(1):409-21. doi: 10.1534/genetics.114.167155.

Hirsch CN et al. Insights into the maize pan-genome and pan-transcriptome. *Plant Cell.* 2014 Jan;26(1):121-35. doi: 10.1105/tpc.113.119982.

Howe AC et al. Tackling soil diversity with the assembly of large, complex metagenomes. *Proc Natl Acad Sci U S A.* 2014 Apr 1;111(13):4904-9. doi: 10.1073/pnas.1402564111.

International Wheat Genome Sequencing Consortium (IWGSC). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science.* 2014 Jul 18;345(6194):1251788. doi: 10.1126/science.1251788.

Ivanova NN et al. Stop codon reassignments in the wild. *Science.* 2014 May 23;344(6186):909-13. doi: 10.1126/science.1250691.

Jansson JK et al. The microbial ecology of permafrost. *Nat Rev Microbiol.* 2014 Jun;12(6):414-25. doi: 10.1038/nrmicro3262.

Jay ZJ et al. Predominant *Acidilobus*-like populations from geothermal environments in yellowstone national park exhibit similar metabolic potential in different hypoxic microbial communities. *Appl Environ Microbiol.* 2014 Jan;80(1):294-305. doi: 10.1128/AEM.02860-13.

Kamke J et al. Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *ISME J.* 2013 Dec;7(12):2287-300. doi: 10.1038/ismej.2013.111.

Kamke J et al. The candidate phylum *Poribacteria* by single-cell genomics: new insights into phylogeny, cell-compartmentation, eukaryote-like repeat proteins, and other genomic features. *PLoS One.* 2014 Jan 31;9(1):e87353. doi: 10.1371/journal.pone.0087353. eCollection 2014.

Kashtan N et al. Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus. Science.* 2014 Apr 25;344(6182):416-20. doi: 10.1126/science.1248575.

Kazakov AE et al. New family of tungstate-responsive transcriptional regulators in sulfate-reducing bacteria. *J Bacteriol.* 2013 Oct;195(19):4466-75. doi: 10.1128/JB.00679-13.

Khanna N et al. Complete genome sequence of *Enterobacter* sp. IIT-BT 08: A potential microbial strain for high rate hydrogen production. *Stand Genomic Sci.* 2013 Dec 15;9(2):359-69. doi: 10.4056/sigs.4348035. eCollection 2013 Dec 20.

Kis-Papo T et al. Genomic adaptations of the halophilic Dead Sea filamentous fungus *Eurotium rubrum. Nat Commun.* 2014 May 9;5:3745. doi: 10.1038/ncomms4745.

Koike H. Comparative genomics analysis of *Trichoderma reesei* strains. *Indust Biotechnol.* Dec 2013, 9(6): 352-367. doi:10.1089/ind.2013.0015.

Kouvelis VN et al. Finished genome of *Zymomonas mobilis* subsp. *mobilis* strain CP4, an applied ethanol producer. *Genome Announc.* 2014 Jan 9;2(1). pii: e00845-13. doi: 10.1128/genomeA.00845-13.

Kuever J et al. Genome analysis of *Desulfotomaculum gibsoniae* strain Groll(T) a highly versatile Gram-positive sulfate-reducing bacterium. *Stand. Genomic Sci.* 2014 Mar 20;9(3):821-39. doi: 10.4056/sigs.5209235. eCollection 2014 Jun 15.

Kyrpides NC et al. Genomic Encyclopedia of Type Strains, Phase I: The one thousand microbial genomes (KMG-I) project. *Stand. Genomic Sci.* 2013 Dec 17;9(3). doi:10.4056/sigs.5068949

Kyrpides NC et al. Genomic encyclopedia of bacteria and archaea: Sequencing a myriad of type strains. *PLoS Biol.* 2014 Aug 5;12(8):e1001920. doi: 10.1371/journal.pbio.1001920. eCollection 2014 Aug.

Lamendella R et al. Assessment of the Deepwater Horizon oil spill impact on Gulf coast microbial communities. *Front Microbiol.* 2014 Apr 3;5:130. doi: 10.3389/fmicb.2014.00130. eCollection 2014.

Lassila JK et al. Assembly of robust bacterial microcompartment shells using building blocks from an organelle of unknown function. *J Mol Biol.* 2014 May 29;426(11):2217-28. doi: 10.1016/j.jmb.2014.02.025.

Leverenz RL et al. Structural and functional modularity of the orange carotenoid protein: Distinct roles for the N- and C-terminal domains in cyanobacterial photoprotection. *Plant Cell.* 2014 Jan;26(1):426-37. doi: 10.1105/tpc.113.118588.

Lewis S et al. Fuzzy GIS-based multi-criteria evaluation for US Agave production as a bioenergy feedstock. *GCB Bioenergy.* 2014 Feb 28. doi: 10.1111/gcbb.12116.

Lindemann SR et al. The epsomitic phototrophic microbial mat of Hot Lake, Washington: Community structural responses to seasonal cycling. *Front Microbiol.* 2013 Nov 13;4:323. doi: 10.3389/fmicb.2013.00323. eCollection 2013.

Linshiz G et al. PR-PR: Cross-platform laboratory automation system. *ACS Synth Biol.* 2014 Aug 15;3(8):515-24. doi: 10.1021/sb4001728. Epub 2014 Jan 14.

Lowry DB et al. The genetics of divergence and reproductive isolation between ecotypes of *Panicum hallii. New Phytol.* 2014 Sep 23. doi: 10.1111/nph.13027. [Epub ahead of print]

Marcussen T et al. Ancient hybridizations among the ancestral genomes of bread wheat. *Science.* 2014 Jul 18;345(6194):1250092. doi: 10.1126/science.1250092.

Markowitz VM et al. IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* 2014 Jan;42(Database issue):D568-73. doi: 10.1093/nar/gkt919.

Markowitz VM et al. IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res.* 2014 Jan;42(Database issue):D560-7. doi: 10.1093/nar/gkt963.

Martin JA et al. A near complete snapshot of the *Zea mays* seedling transcriptome revealed from ultra-deep sequencing. *Sci Rep.* 2014 Mar 31;4:4519. doi: 10.1038/srep04519.

Mascher M et al. Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J.* 2013 Nov;76(4):718-27. doi: 10.1111/tpj.12319. Epub 2013 Oct 10.

Mason OU et al. Single-cell genomics reveals features of a *Colwellia* species that was dominant during the Deepwater Horizon oil spill. *Front Microbiol.* 2014 Jul 8;5:332. doi: 10.3389/fmicb.2014.00332. eCollection 2014.

Massoumi Alamouti S et al. Comparative genomics of the pine pathogens and beetle symbionts in the genus *Grosmannia*. *Mol Biol Evol.* 2014 Jun;31(6):1454-74. doi: 10.1093/molbev/msu102.

Matzke NJ et al. Bayesian analysis of congruence of core genes in *Prochlorococcus* and *Synechococcus* and implications on horizontal gene transfer. *PLoS One.* 2014 Jan 21;9(1):e85103. doi: 10.1371/journal.pone.0085103. eCollection 2014 Jan 21.

Meier-Kolthoff JP et al. Genome sequence of the chemoheterotrophic soil bacterium *Saccharomonospora cyanea* type strain (NA-134(T)). *Stand Genomic Sci.* 2013 Sep 30;9(1):28-41. doi: 10.4056/sigs.4207886. eCollection 2013 Oct 16.

Moulin L et al. Complete Genome sequence of *Burkholderia phymatum* STM815(T), a broad host range and efficient nitrogen-fixing symbiont of *Mimosa* species. *Stand. Genomic Sci.* 2014 Mar 25;9(3):763-74. doi:10.4056/sigs.4861021. eCollection 2014 Jun 15.

Mukherjee S et al. Population level analysis of evolved mutations underlying improvements in plant hemicellulose and cellulose fermentation by *Clostridium phytofermentans*. *PLoS One.* 2014 Jan 22;9(1):e86731. doi: 10.1371/journal.pone.0086731. eCollection 2014.

Myburg AA et al. The genome of *Eucalyptus grandis*. *Nature.* 2014 Jun 19;509(7505):356-62. doi: 10.1038/nature13308.

Nagy LG et al. Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. *Nat Commun.* 2014 Jul 18;5:4471. doi: 10.1038/ncomms5471.

Nandasena K et al. Complete genome sequence of *Mesorhizobium ciceri bv. biserrulae* type strain (WSM1271(T)). *Stand. Genomic Sci.* 2013 Dec 31; 9(3):462-72. doi:10.4056/sigs.4458283. eCollection 2014.

Nobu MK et al. The genome of *Syntrophorhabdus aromaticivorans* strain UI provides new insights for syntrophic aromatic compound metabolism and electron flow. *Environ Microbiol.* 2014 Mar 3. doi: 10.1111/1462-2920.12444. [Epub ahead of print]

Nobu MK et al. Draft genome sequence of *Syntrophorhabdus aromaticivorans* strain UI, a mesophilic aromatic compound-degrading syntroph. *Genome Announc.* 2014 Feb 6;2(1). pii: e01064-13. doi: 10.1128/genomeA.01064-13.

Nordberg H et al. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res.* 2014 Jan;42(Database issue):D26-31. doi: 10.1093/nar/gkt1069.

Novichkov PS et al. RegPrecise 3.0 — a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics.* 2013 Nov 1;14:745. doi: 10.1186/1471-2164-14-745.

Ntougias S et al. High quality draft genome sequence of *Olivibacter sitiensis* type strain (AW-6(T)), a diphenol degrader with genes involved in the catechol pathway. *Stand. Genomic Sci.* 2014 Mar 20;9(3):783-93. doi:10.4056/sigs.5088950. eCollection 2014.

Nurk S et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol.* 2013 Oct;20(10):714-37. doi: 10.1089/cmb.2013.0084.

Ohm RA et al. Genomics of wood-degrading fungi. *Fungal Genet Biol.* 2014 May 20. pii: S1087-1845(14)00079-6. doi: 10.1016/j.fgb.2014.05.001. [Epub ahead of print]

Palaniappan K et al. Genome sequence of the moderately thermophilic sulfurreducing bacterium *Thermanaerovibrio* velox type strain (Z-9701(T)) and emended description of the genus *Thermanaerovibrio*. *Stand Genomic Sci.* 2013 Oct 2;9(1):57-70. doi: 10.4056/sigs.4237901. eCollection 2013 Oct 16.

Pedersen K et al. Complete genome sequence of the subsurface, mesophilic sulfate-reducing bacterium *Desulfovibrio aespoeensis* Aspo-2. *Genome Announc.* 2014 May 29;2(3). pii: e00509-14. doi: 10.1128/genomeA.00509-14.

Pedras MS et al. The phytoalexin camalexin induces fundamental changes in the proteome of *Alternaria brassicicola* different from those caused by brassinin. *Fungal Biol*. 2014 Jan;118(1):83-93. doi: 10.1016/j.funbio.2013.11.005. Epub 2013 Dec 1.

Pendleton AL et al. Duplications and losses in gene families of rust pathogens highlight putative effectors. *Front Plant Sci.* 2014 Jun 26;5:299. doi: 10.3389/fpls.2014.00299. eCollection 2014.

Pfeifer M et al. Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science.* 2014 Jul 18;345(6194):1250091. doi: 10.1126/science.1250091.

Piao H et al. Identification of novel biomass-degrading enzymes from genomic dark matter: Populating genomic sequence space with functional annotation. *Biotechnol Bioeng.* 2014 Aug;111(8):1550-65. doi: 10.1002/bit.25250.

Podell S et al. Seasonal fluctuations in ionic concentrations drive microbial succession in a hypersaline lake community. *ISME J.* 2014 May;8(5):979-90. doi: 10.1038/ismej.2013.221.

Poliakov A et al. GenomeVISTA — an integrated software package for whole-genome alignment and visualization. *Bioinformatics.* 2014 Sep 15;30(18):2654-5. doi: 10.1093/bioinformatics/btu355.

Prestat E et al. FOAM (Functional Ontology Assignments for Metagenomes): A Hidden Markov Model (HMM) database with environmental focus. *Nucleic Acids Res.* 2014 Sep 26. pii: gku702. doi: 10.1093/nar/gku702 [Epub ahead of print]

Ramirez MS et al. Genome sequences of two carbapenemase-resistant *Klebsiella pneumoniae* ST258 isolates. *Genome Announc.* 2014 Jun 19;2(3). pii: e00558-14. doi: 10.1128/genomeA.00558-14.

Rawat SR et al. Complete genome sequence of *Granulicella tundricola* type strain MP5ACTX9(T), an *Acidobacteria* from tundra soil. *Stand. Genomic Sci.* 2013 Dec 5;9(2). doi:10.4056/sigs.4648353

Reeve WG et al. Genome sequence of the *Trifolium rueppellianum*-nodulating *Rhizobium leguminosarum* bv. *trifolii* strain WSM2012. *Stand Genomic Sci.* 2013 Dec 15;9(2):283-93. doi: 10.4056/sigs.4528262. eCollection 2013.

Reeve WG et al. Genome sequence of the clover-nodulating *Rhizobium leguminosarum* bv. *trifolii* strain SRDI943. *Stand Genomic Sci.* 2013 Dec 15;9(2):232-42. doi: 10.4056/sigs.4478252. eCollection 2013.

Reeve WG et al. Genome sequence of the *Lebeckia ambigua*-nodulating "*Burkholderia sprentiae*" strain WSM5005(T.). *Stand Genomic Sci.* 2013 Dec 15; 9(2):385-94. doi: 10.4056/sigs.4558268. eCollection 2013.

Reeve WG et al. Complete genome sequence of *Mesorhizobium australicum* type strain (WSM2073(T)). *Stand Genomic Sci.* 2013 Dec 15;9(2):410-9. doi: 10.4056/sigs.4568282. eCollection 2013 Dec 20.

Reeve WG et al. Genome sequence of the *Listia angolensis* microsymbiont *Microvirga lotononidis* strain WSM3557T. *Stand. Genomic Sci.* 2013 Dec 31;9(3). doi:10.4056/sigs.4548266

Reeve WG et al. Complete genome sequence of *Mesorhizobium opportunistum* type strain WSM2075(T). *Stand Genomic Sci.* 2013 Dec 15;9(2): 294-303. doi: 10.4056/sigs.4538264. eCollection 2013 Dec 20.

Reeve WG et al. Genome sequence of the lupin-nodulating *Bradyrhizobium sp.* strain WSM1417. *Stand Genomic Sci.* 2013 Dec 15;9(2):273-82. doi: 10.4056/sigs.4518260. eCollection 2013 Dec 20.

Reeve WG et al. Genome sequence of the South American clover-nodulating *Rhizobium leguminosarum* bv. *trifolii* strain WSM597. *Stand Genomic Sci.* 2013 Dec 15;9(2):264-72. doi: 10.4056/sigs.4508258. eCollection 2013 Dec 20.

Reeve WG et al. Genome sequence of the *Ornithopus/Lupinus-nodulating Bradyrhizobium sp.* strain WSM471. *Stand Genomic Sci.* 2013 Dec 15;9(2):254-63. doi: 10.4056/sigs.4498256. eCollection 2013 Dec 20.

Reeve WG et al. Genome sequence of the clover-nodulating *Rhizobium leguminosarum* bv. *trifolii* strain TA1. *Stand Genomic Sci.* 2013 Dec 15;9(2):243-53. doi: 10.4056/sigs.4488254. eCollection 2013 Dec 20.

Reeve WG et al. Genome sequence of the clover-nodulating *Rhizobium leguminosarum* bv. *trifolii* strain SRDI565. *Stand Genomic Sci.* 2013 Dec 15;9(2):220-31. doi: 10.4056/sigs.4468250. eCollection 2013 Dec 20.

Reeve WG et al. Genome sequence of *Ensifer arboris* strain LMG 14919(T); a microsymbiont of the legume *Prosopis chilensis* growing in Kosti, Sudan. *Stand. Genomic Sci.* 2013 Dec 31;9(3):473-83. doi:10.4056/sigs.4828625. eCollection 2014 Jun 15.

Reeve WG et al. Genome sequence of *Ensifer medicae* strain WSM1115; an acid-tolerant *Medicago*-nodulating microsymbiont from Samothraki, Greece. *Stand. Genomic Sci.* 2013 Dec 31;9(3):514-26. doi:10.4056/sigs.4938652. eCollection 2014 Jun 15.

Reeve WG et al. Genome sequence of the *Medicago*-nodulating *Ensifer meliloti* commercial inoculant strain RRI128. *Stand. Genomic Sci.* 2014 Jan 25;9(3):602-13. doi:10.4056/sigs.4929626. eCollection 2014 Jun 15

Reeve WG et al. Genome sequence of *Microvirga lupini* strain LUT6(T), a novel *Lupinus* alphaproteobacterial microsymbiont from Texas. *Stand Genomic Sci.* 2014 Mar 1;9(3):1159-67. doi: 10.4056/sigs.5249382. eCollection 2014 Jun 15.

Rice DW et al. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science.* 2013 Dec 20;342(6165):1468-73. doi: 10.1126/science.1246275.

Riedel T et al. Genome sequence of the *Litoreibacter arenae* type strain (DSM 19593(T)), a member of the *Roseobacter* clade isolated from sea sand. *Stand Genomic Sci.* 2013 Oct 1;9(1):117-27. doi: 10.4056/sigs.4258318. eCollection 2013 Oct 16.

Riedel T et al. Genome sequence of the *Wenxinia marina* type strain (DSM 24838(T)), a representative of the *Roseobacter* group isolated from oilfield sediments. *Stand. Genomic Sci.* 2014 Mar 28;9(3):855-65. doi: 10.4056/sigs.5601028. eCollection 2014.

Riedel T et al. Genome sequence of the exopolysaccharide-producing *Salipiger mucosus* type strain (DSM 16094(T)), a moderately halophilic member of the *Roseobacter* clade. *Stand Genomic Sci.* 2014 Mar 15;9(3):1333-45. doi: 10.4056/sigs.4909790. eCollection 2014.

Riley R et al. Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *Proc Natl Acad Sci U S A.* 2014 Jul 8;111(27):9923-8. doi: 10.1073/pnas.1400592111.

Rinke C et al. Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat Protoc.* 2014 May;9(5):1038-48. doi: 10.1038/nprot.2014.067.

Roux S et al. Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *Elife.* 2014 Aug 29;3. doi: 10.7554/eLife.03125.

Sato TK et al. Harnessing genetic diversity in *Saccharomyces cerevisiae* for fermentation of xylose in hydrolysates of alkaline hydrogen peroxide-pretreated biomass. *Appl Environ Microbiol.* 2014 Jan;80(2):540-54. doi: 10.1128/AEM.01885-13.

Schmutz J et al. A reference genome for common bean and genome-wide analysis of dual domestications. *Nat Genet.* 2014 Jul;46(7):707-13. doi: 10.1038/ng.3008.

Shi W et al. Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Res.* 2014 Sep;24(9):1517-25. doi: 10.1101/gr.168245.113.

Shilova IN et al. A microarray for assessing transcription from pelagic marine microbial taxa. *ISME J.* 2014 Jul;8(7):1476-91. doi: 10.1038/ismej.2014.1. Epub 2014 Jan 30.

Singh G et ak. Rearrangement of a large novel *Pseudomonas aeruginosa* gene island in strains isolated from a patient developing ventilator-associated pneumonia. *J Clin Microbiol.* 2014 Jul;52(7):2430-8. doi: 10.1128/JCM.01626-13. Epub 2014 Apr 30.

Smith H et al. Draft genome sequence and description of *Janthinobacterium* sp. strain CG3, a psychrotolerant Antarctic supraglacial stream bacterium. *Genome Announc.* 2013 Nov 21;1(6). pii: e00960-13. doi: 10.1128/genomeA.00960-13.

Sulakhe D et al. High-throughput translational medicine: Challenges and solutions. *Adv Exp Med Biol.* 2014;799:39-67. doi: 10.1007/978-1-4614-8778-4_3.

Swan BK et al. Genomic and metabolic diversity of Marine Group I Thaumarchaeota in the mesopelagic of two subtropical gyres. *PLoS One.* 2014 Apr 17;9(4):e95380. doi: 10.1371/journal.pone.0095380. eCollection 2014.

Syed K et al. Genomewide annotation and comparative genomics of cytochrome P450 monooxygenases (P450s) in the polypore species *Bjerkandera adusta, Ganoderma* sp. and *Phlebia brevispora. Mycologia.* 2013 Nov-Dec;105(6):1445-55. doi: 10.3852/13-002.

Tak N et al. Genome sequence of *Ensifer* sp. TW10; a *Tephrosia wallichii* (Biyani) microsymbiont native to the Indian Thar Desert. *Stand Genomic Sci.* 2013 Dec 15;9(2):304-14. doi: 10.4056/sigs.4598281. eCollection 2013 Dec 20.

Taş N et al. Impact of fire on active layer and permafrost microbial communities and metagenomes in an upland Alaskan boreal forest. *ISME J.* 2014 Sep;8(9):1904-19. doi: 10.1038/ismej.2014.36.

Terpolilli J et al. Genome sequence of *Ensifer medicae* strain WSM1369; an effective microsymbiont of the annual legume *Medicago sphaerocarpos. Stand Genomic Sci.* 2013 Dec 17;9(2):420-30. doi: 10.4056/sigs.4838624. eCollection 2013 Dec 20.

Terpolilli J et al. Genome sequence of *Rhizobium leguminosarum* bv *trifolii* strain WSM1689, the microsymbiont of the one flowered clover *Trifolium uniflorum. Stand. Genomic Sci.* 2013 Dec 31;9(3). doi:10.4056/sigs.4988693

Terpolilli J et al. Genome sequence of *Ensifer meliloti* strain WSM1022; a highly effective microsymbiont of the model legume *Medicago truncatula* A17. *Stand Genomic Sci.* 2013 Dec 15;9(2):315-24. doi: 10.4056/sigs.4608286. eCollection 2013 Dec 20.

Thiel T et al. Complete genome sequence of *Anabaena variabilis* ATCC 29413. *Stand. Genomic Sci.* 2014 Jan 1;9(3):562-73. doi:10.4056/sigs.3899418. eCollection 2014 Jun 15.

Thrash JC et al. Single-cell enabled comparative genomics of a deep ocean SAR11 bathytype. *ISME J.* 2014 Jul;8(7):1440-51.doi: 10.1038/ismej.2013.243. Epub 2014 Jan 23.

Tisserant E et al. Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proc Natl Acad Sci U S A.* 2013 Dec 10;110(50):20117-22. doi: 10.1073/pnas.1313452110.

Toome M et al. Genome sequencing provides insight into the reproductive biology, nutritional mode and ploidy of the fern pathogen *Mixia osmundae. New Phytol.* 2014 Apr;202(2):554-64. doi: 10.1111/nph.12653. Epub 2013 Dec 24.

Toome M et al. Draft genome sequence of a rare smut relative, *Tilletiaria anomala* UBC 951. *Genome Announc.* 2014 Jun 12;2(3). pii: e00539-14. doi: 10.1128/genomeA.00539-14.

Tsiamis G et al. Microbial diversity for biotechnology. *Biomed Res Int*. 2014;2014:845972. doi: 10.1155/2014/845972. Epub 2014 Jan 23.

Visser M et al. Genome analyses of the carboxydotrophic sulfate-reducers *Desulfotomaculum nigrificans* and *Desulfotomaculum carboxydivorans* and reclassification of *Desulfotomaculum carboxydivorans* as a later syrnonym of *D. nigrificans*. *Stand. Genomic Sci.* 2014 Mar 1;9(3):655-675. doi:10.4056/sigs.4718645. eCollection 2014 Jun 15.

Wagner MR et al. Natural soil microbes alter flowering phenology and the intensity of selection on flowering time in a wild *Arabidopsis* relative. *Ecol Lett*. 2014 Jun;17(6):717-26. doi: 10.1111/ele.12276. Epub 2014 Apr 4.

Walker R et al. Genome sequence of the acid-tolerant *Burkholderia sp.* strain WSM2230 from Karijini National Park, Australia. *Stand. Genomic Sci.* 2013 Dec 31; 9(3):551-61. doi: 10.4056/sigs.5008793. eCollection 2014.

Walker R et al. Genome sequence of the acid-tolerant *Burkholderia sp.* strain WSM2232 from Karijini National Park, Australia. *Stand. Genomic Sci.* 2013 Nov 15;9(3):1168-80. doi: 10.4056/sigs.5018795. eCollection 2014.

Wang W et al. The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. *Nat Commun.* 2014 Feb 19;5:3311. doi: 10.1038/ncomms4311.

Weston DJ et al. Sphagnum physiology in the context of changing climate: Emergent influences of genomics, modeling and host-microbiome interactions on understanding ecosystem function. *Plant Cell Environ.* 2014 Sep 30. doi: 10.1111/pce.12458. [Epub ahead of print]

Willems A et al. Genome sequence of *Burkholderia mimosarum* strain LMG 23256(T), a *Mimosa pigra* microsymbiont from Anso, Taiwan. *Stand Genomic Sci.* 2013 Dec 31;9(3):484-94. doi:10.4056/sigs.4848627. eCollection 2014.

Williams TJ et al. Microbial ecology of an Antarctic hypersaline lake: Genomic assessment of ecophysiology among dominant *haloarchaea. ISME J.* 2014 Aug;8(8):1645-58. doi: 10.1038/ismej.2014.18. Epub 2014 Feb 20.

Wilson JG et al. Draft genome sequence of *Enterobacter cloacae* strain JD6301. *Genome Announc.* 2014 May 29;2(3). pii: e00381-14. doi: 10.1128/genomeA.00381-14.

Wilson MC et al. An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature.* 2014 Feb 6;506(7486):58-62. doi: 10.1038/nature12959. Epub 2014 Jan 29.

Wohlbach DJ et al. Comparative genomics of *Saccharomyces cerevisiae* natural isolates for bioenergy production. *Genome Biol Evol.* 2014 Sep 5. pii: evu199. doi: 10.1093/gbe/evu199 [Epub ahead of print]

Wu D et al. Systematic identification of gene families for use as "markers" for phylogenetic and phylogeny-driven ecological studies of bacteria and archaea and their major subgroups. *PLoS One.* 2013 Oct 17;8(10):e77033. doi: 10.1371/journal.pone.0077033. eCollection 2013.

Wu GA et al. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat Biotechnol.* 2014 Jul;32(7):656-62. doi: 10.1038/nbt.2906.

Wu YW et al. MaxBin: An automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome.* 2014 Aug 1;2:26. doi: 10.1186/2049-2618-2-26. eCollection 2014.

Xie G et al. Recruiting human microbiome shotgun data to site-specific reference genomes. *PLoS One.* 2014 Jan 15;9(1):e84963. doi: 10.1371/journal.pone.0084963. eCollection 2014.

Yamamoto K et al. Complete genome sequence of *Methanolinea tarda* NOBI-1(T), a hydrogenotrophic methanogen isolated from methanogenic digester sludge. *Genome Announc.* 2014 Sep 4;2(5). pii: e00876-14. doi: 10.1128/genomeA.00876-14.

Yamamoto K et al. Complete genome sequence of *Methanoregula formicica* SMSP(T), a mesophilic hydrogenotrophic methanogen isolated from a methanogenic upflow anaerobic sludge blanket reactor. *Genome Announc.* 2014 Sep 4;2(5). pii: e00870-14. doi: 10.1128/genomeA.00870-14.

Youssef NH et al. Insights into the metabolism, lifestyle and putative evolutionary history of the novel archaeal phylum 'Diapherotrites.' *ISME J.* 2014 Aug 1. doi: 10.1038/ismej.2014.141. [Epub ahead of print]

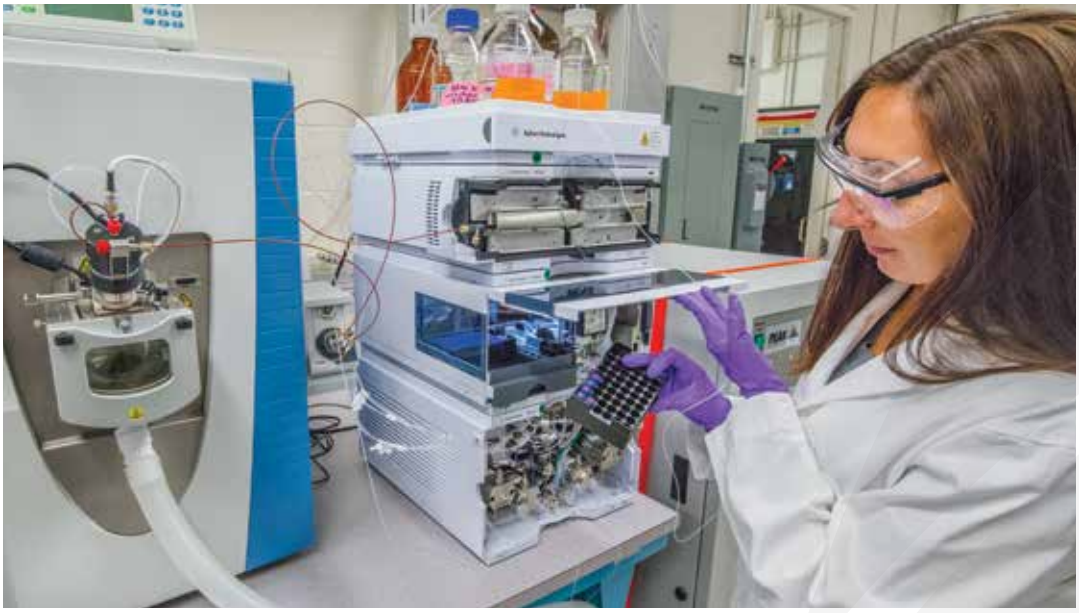
Zaremba-Niedzwiedzka K et al. Single-cell genomics reveal low recombination frequencies in freshwater bacteria of the SAR11 clade. *Genome Biol.* 2013 Nov 28;14(11):R130. doi: 10.1186/gb-2013-14-11-r130.

Zarzycki J et al. The crystal structures of the tri-functional *Chloroflexus aurantiacus* and bi-functional *Rhodobacter sphaeroide*s malyl-CoA lyases and comparison with CitE-like superfamily enzymes and malate synthases. *BMC Struct Biol.* 2013 Nov 9;13:28. doi: 10.1186/1472-6807-13-28.

Zhang Y et al. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature.* 2013 Dec 12;504(7479):306-10. doi: 10.1038/nature12716.

Zhou Y et al. High quality draft genome sequence of the slightly halophilic bacterium *Halomonas zhanjiangensis* type strain JSM 078169(T) (DSM 21076(T)) from a sea urchin in southern China. *Stand. Genomic Sci.* 2014 Apr 11;9(3):1020-30. doi:10.4056/sigs.5449586. eCollection 2014.

Ziemert N et al. Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora. Proc Natl Acad Sci U S A.* 2014 Mar 25;111(12):E1130-9. doi: 10.1073/pnas.1324161111.




JGI


JGI


MEETINGS
JGI

Comments?

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

Follow Us

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

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

15-JGI-1390

Headquartered in Walnut Creek, California, the U.S. Department of Energy Joint Genome Institute has over 250 staff devoted to advancing the frontiers of genomics in support of clean energy generation and environmental characterization and cleanup.

U.S. DEPARTMENT OF ENERGY
Office of Science

BERKELEY LAB

JOINT GENOME INSTITUTE

UNITED STATES DEPARTMENT OF ENERGY