JGI JOINT GENOME INSTITUTE
A DOE OFFICE OF SCIENCE USER FACILITY

2015 Progress Report

U.S. Department of Energy

Joint Genome Institute

On the cover: The East River catchment, in Colorado's Rocky Mountains, is the site of Berkeley Lab's Genomes to Watershed Scientific Focus Area. DNA collected from floodplain soils are being sequenced. The information generated will go toward developing a simulation framework for exploring interactions between microbial metabolic potential, hydrology and watershed biogeochemistry.

Table of Contents

DOE JGI **Mission**

To assist in the DOE's mission of on developing clean, renewable and sustainable alternative fuel sources from plant material, the DOE JGI sequences and analyzes the genomes of candidate feedstocks, as well as fungi and microbes that could play roles in breaking down and converting the plant mass into liquid transportation fuel.

(AOtaro Flickr, CC BY 2.0)

The mission of the U.S. Department of Energy Joint Genome Institute (DOE JGI) is to serve the diverse scientific community as a national 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**

Two thousand fifteen was yet another successful year for the Department of Energy Joint Genome Institute (DOE JGI). We exceeded our productivity goals, made an abundance of high-quality resources available to our user community, and sustained our tradition of contributing high-impact publications to the scientific literature.

In reviewing the progress that the DOE JGI made in 2015, we measure our contributions in several ways. First and foremost, as a DOE national user facility, the DOE JGI supported 958 active primary users and collaborators in 2015. Associated with our work with primary users in the past year, we produced 11 plant, 114 fungal, and 1,360 prokaryotic *de novo* genomes, as well as 1,205 metagenomic data sets. Data from these projects, in addition to supporting the science of our primary users, was all provided to an even greater number of scientists working on energy and environmental research problems via our various database platforms. These include the DOE JGI's Genome Portal and our comparative genomics platforms: the Integrated Microbial Genomes system for microbes and metagenomes, Phytozome for plants, and MycoCosm for fungi. In fiscal year 2015 alone, 700,000 unique visitors downloaded a total of 1.8 million data sets from these DOE JGI data portals, further amplifying the impact of the data produced by the DOE JGI in collaboration with its users.

In addition to generating and disseminating data for our primary users through our portals, we host an annual conference for DOE JGI users and prospective users, the Genomics of Energy & Environment Meeting. This provides an opportunity — through presentations, poster sessions and workshops — to share ways in which genome sciences can be applied to solving pressing energy and environmental problems. Last year, 430 participants came to Walnut Creek and visited the DOE JGI for our 10th Annual User Meeting, representing the best-attended conference yet. A summary of the meeting can be found on page 60.

The DOE JGI's impact is also propagated by the significant number of scientific publications originating from the DOE JGI and its users. Last year alone, the DOE JGI contributed to 181 peer-reviewed publications. Nearly a quarter of these were in higher-impact journals, including seven in the most selective, such as *Science* and *Nature*. This level of high-impact scientific productivity is what the research community has come to expect from the DOE JGI and contributes to our being recognized as a source of value-rich information. Eight of our scientists were among the "Top 1% Thomson Reuters Highly Cited Researchers" in 2015. When we look back on the last half-dozen years, there have been over 25,000 citations of DOE JGI publications. Below are summaries of a selection of the top 2015 DOE JGI publications.

- **Searching for New Life.** In a perspective piece, we described how advances in genomic technologies make it possible for researchers not just to learn more about explored and unexplored species, but to discover new organisms as well. Nature has been tinkering with life for at least three and a half billion years, and we now have a new set of ways to look for novel life that has so far eluded discovery. Analysis of massive-scale environmental DNA and RNA sequence data to identify outliers to previously defined life represents a powerful means to explore the unknown that is already bearing fruit in DOE JGI studies (*Science*, November 6, 2014).

Computational Tools for Microbial Species. In exploring uncultured microbes and searching for new life, researchers have been hindered by the lack of a high-throughput process to review assembled genome sequences. Two DOE JGI papers help overcome these challenges:

- A description of ProDeGe, representing the first computational protocol for quick and automated removal of contaminant sequences from draft genomes. The tool works on any type of genome sequence and is part of our commitment to developing software solutions to mitigate daunting computational challenges, thus enhancing the utility and impact of the DNA sequences we generate (*The ISME Journal*, June 9, 2015).

- A new method for classifying microbial species based on genome data that supplements traditional approaches relied on by microbiologists for decades. Such computational tool development, facilitated by the DOE JGI's partnership with the National Energy Research Scientific Computing Center (NERSC), is essential to characterizing complex biological and environmental systems in support of DOE's research missions (*Nucleic Acids Research*, July 6, 2015).

Functional Genomic Studies of Algae. Understanding the regulatory mechanisms that switch genes on and off (epigenomics) is critical to a better understanding of the meaning of genomic data. Some algae produce energy-dense oils or lipids that can be converted into fuels, but it can be challenging to produce them in adequate quantities for biofuel production. An in-depth epigenomic study of the green alga *Chlamydomonas reinhardtii* revealed key regulators in algal lipid production, opening the door to more effective use of algae as a biofuel feedstock. This study further solidified our position as a leader in algal genomics; DOE JGI has published over 75 percent of all publicly available algal genomes, including the *Chlamydomonas* reference genome that enabled this study (*Nature Plants,* July 27, 2015).

Understanding Permafrost Carbon Cycling. Vast areas of the planet are covered by permafrost soils, but they are highly vulnerable to thawing as global temperatures continue to rise. Studies suggest such thawing could result in the largest contribution of carbon transferred to the atmosphere by a single terrestrial process, but the magnitude of this carbon output depends on the activities of microbes residing in the soil. With colleagues from several national laboratories, we applied multiple molecular technologies, collectively referred to as "omics," to three types of Alaskan soils ranging from completely thawed to completely frozen, deepening our understanding of the role of microbial communities in degrading permafrost organic carbon and ultimately producing CO_2 and methane (*Nature*, March 4, 2015).

Novel Approaches to Metagenome Sequencing. Studying a microbial community through metagenomics allows researchers to bypass the limits of cultivation, but it also presents a challenge in assembling meaningful sequence information out of the genomic fragments for multiple individuals. We helped evaluate the utility of "synthetic long reads" produced by a unique library preparation technique to bypass the need for large-scale *de novo* assembly, and found such data could enable better characterization of rare species than conventional short-read sequencing (*Genome Research*, April 2015).

- **Defining a Candidate Bacterial Phylum.** Surveying uncultivated bacteria and archaea not only fills in the still-unexplored branches of the tree of life, but also provides insights into potentially useful microbial capabilities that could help DOE advance its energy and environmental missions. By illuminating the so-called "microbial dark matter," we have been able to conduct a comprehensive look at the microbial diversity of two groups of uncultivated bacteria and found compelling evidence that they actually belong to a single candidate phylum (*The ISME Journal*, June 19, 2015).

- **Finding Friends in the Root Microbiome.** Understanding how microbial communities interact at the nexus of plant roots and soil is a key focus area for the DOE JGI, as cultivation of beneficial microbiomes will be critical to sustainable bioenergy crop production in the future. With colleagues, we used a combination of plant genetics, biochemistry and sequence-based community profiling to investigate the roles of three phytohormones in controlling the composition of the root microbiome in the model plant *Arabidopsis thaliana* (*Science*, August 21, 2015).

- **Mutualism for Beneficial Environmental Adaptations.** Drawing on 49 fungal genomes, many sequenced at DOE JGI, we helped conduct the first broad, comparative phylogenomic analysis of mycorrhizal fungi. The results help explain how plants and fungi developed symbiotic relationships, and how this mutualistic association provides host plants with beneficial traits for environmental adaptation (*Nature Genetics*, February 23, 2015).

- **Harnessing Destructive Tendencies for Biofuel Applications.** The shipworm, a worm-like, wood-eating marine bivalve, has long been considered a foe of wooden seagoing vessels, but its destructive capabilities could be useful for the industrial production of advanced biofuels from woody plant mass. Harnessing our metagenomics know-how, we helped colleagues describe the novel strategy by which the shipworm breaks down and digests wood (*PNAS*, November 10, 2014).

These represent just a few of the publications that would not have been possible without the technical and intellectual contributions of DOE JGI staff.

Few fields of science have seen the degree of disruptive technological advances that genomics has over the past decade. To ensure that the DOE JGI remains relevant in this changing environment, we are engaged in ongoing strategic planning. In this process begun three years ago, we recognized that we needed to focus on responding to the fundamental unsolved problem in genomics, the gap between DNA sequence data and the ability to assign biological function to these data sets. A strategic planning retreat held this past year reinforced this view, and identified both areas of progress in implementing our response to this challenge and areas in which we need to continue to grow. Some of the technologies established at the DOE JGI to help bridge the sequence to function gap include epigenomics, the exploration of "second code" modifications to DNA bases; high-throughput analysis of mutant populations; population genomics to survey whole populations of organisms rather than isolated individuals; and metabolomics, the study of the unique chemical fingerprints of organisms and communities. In each of these areas, we have coupled a capability with the appropriate analysis tools to enable our users to derive insights from these studies. To further empower our users in linking sequence to function, we have expanded our DNA Synthesis Science Program. This program's emphasis is now on working with users to study the role of secondary metabolites in microbe-microbe and plant-microbe interactions. This involves guiding users to generate hypotheses based on mining DOE JGI databases, helping them design and build targeted refactored pathways, and studying pathway expression in appropriate hosts.

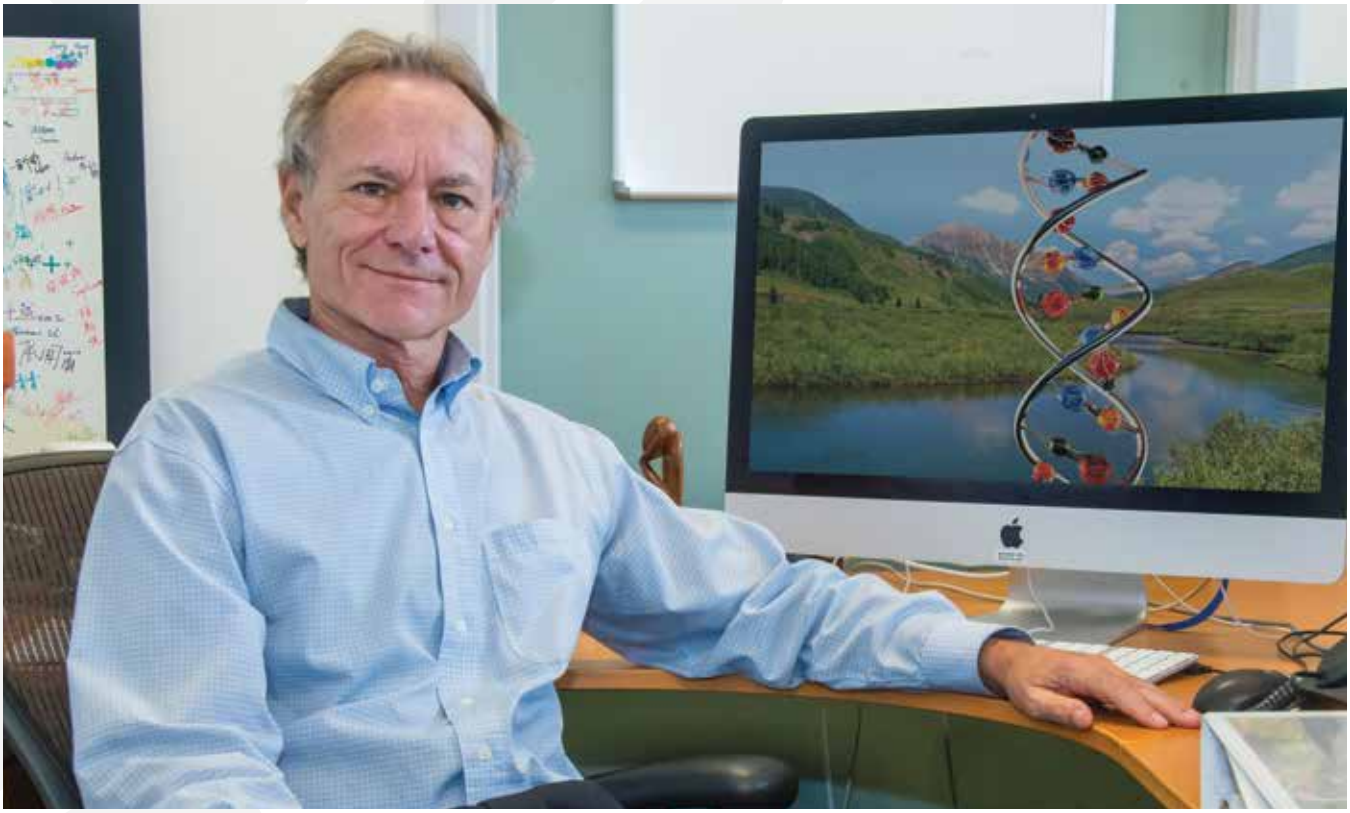
In addition to making our existing sequence to function capabilities available to users, we are committed to adding an evolving set of new capabilities. The Emerging Technologies Opportunity Program (ETOP) is targeted at providing our users with cutting-edge technologies being developed by leading external laboratories. During the last year, working closely with our external partners, we have supported activities in microfluidics, Raman spectroscopy and single-cell function-driven genomics, advanced DNA synthesis technologies, metagenomic assembly, and DNA acquisition for plant and fungal genomics.

To supplement the genomic capabilities the DOE JGI offers users, we are increasingly looking to link with other DOE National Laboratory user facilities to catalyze synergies for user science. As one of two National User Facilities under the umbrella of the DOE Office of Biological and Environmental Research, it was natural for us to join forces in a collaborative program with the Environmental Molecular Sciences Laboratory (EMSL) at Pacific Northwest National Laboratory (PNNL). This program, now called FICUS (Facilities Integrating Collaborations for User Science), represents "one-stop shopping" for our users seeking access to not only the DOE JGI's resources, but also the collection of mass spectrometry and imaging tools available through EMSL to characterize target organisms and communities structurally and biochemically. This program leverages each user facility's strengths to support exciting user science projects that are now beginning to bear fruit, three years after the first round of supported projects. Encouraged by the program's early success, we issued another call for proposals this last year that has resulted in the selected projects summarized on page 55.

The DOE JGI does not exist as an island. Rather, we are part of Lawrence Berkeley National Laboratory (Berkeley Lab), and while many of our users may not be aware of it, the DOE JGI's success is dependent on foundational resources provided to us by two other national user facilities: the National Energy Research Scientific Computing Center (NERSC) and the Energy Sciences Network (ESnet). NERSC is where all of the data we generate are stored, managed and analyzed. This information is then delivered full circle back to the community of researchers and scientists through ESnet, at a rate that's 15,000 times faster than the average home network.

In 2015, our relationship with NERSC was further strengthened with the opening of Shyh Wang Hall on the hillside of Berkeley Lab. NERSC has begun transitioning the DOE JGI's computational assets from its previous home in downtown Oakland to this new building, where NERSC and ESNet have developed and deployed the first-ever 400 billions of bits per second production network. Along with the information from many other projects, this resource is now conveying the DOE JGI's massive datasets.

One of the most exciting developments about the DOE JGI's future is that the facility expects to be moving to a new building on the Berkeley Lab main campus in Berkeley, California. We are far along in the planning, design and implementation of the DOE JGI's new home, the Integrative Genomics Building (IGB), which is slated to open in late 2018. The IGB will bring together the DOE JGI, currently located in Walnut Creek, California, and KBase, the DOE Systems Biology Knowledgebase, currently located near Berkeley in Emeryville. This planned collocation is anticipated to advance new predictive capabilities in support of DOE's mission in fundamental energy and environmental research. In addition, we expect to see significant scientific benefits accrued by the DOE JGI due to the close proximity to the richness of Berkeley Lab and University of California, Berkeley scientific activities. All of us at the DOE JGI are enthusiastic about the move, and I am confident that moving to the Berkeley Lab main campus will ultimately lead to productive collaborations and scientific discoveries that would have never occurred with the DOE JGI remaining in Walnut Creek isolated from the larger scientific community.



One additional development from 2015 that I'd like to share is that we launched a new diversity and inclusion (D&I) initiative, an activity about which I am particularly passionate. Over the trajectory of my career, I have witnessed how melting pots of perspectives and ideas have contributed to the vibrancy and success of the largest, most complex scientific endeavors. At the DOE JGI, we are dedicated to putting in place both top-down and bottom-up approaches to encourage the development of a D&I culture in our workplace. We are also supporting outreach programs to encourage the development of the next generation of scientists, including a vibrant one with the University of California, Merced, a campus with a large minority enrollment. I strongly believe that activities of the D&I initiative will position us to make the greatest scientific contributions as well as hire and retain the very best staff.

It has been a good year for the DOE JGI. I hope you will agree that based on the accomplishments of the past year the future of the DOE JGI remains bright. I encourage you to stay tuned as the DOE JGI continues to evolve as a state-of-the-art genomic science user facility.

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

Achieving the **DOE Mission**

Capable of breaking down wood and leaf litter, white rot fungi are being studied for their potential applications in optimizing the production of advanced biofuels from sustainable plant sources and for cleaning up contaminated sites (see page 39). Commonly known as the angel wing, *Pleurocybella porrigens* is a white rot fungus. *(Francis Martin, INRA)*

The Department of Energy Joint Genome Institute (DOE JGI) is a national user facility funded by the DOE's Office of Biological and Environmental Research (BER) that conducts high-throughput DNA sequencing, synthesis and analysis aligned with BER's bioenergy and environmental missions. These missions mirror DOE and national priorities to:

- Develop renewable and sustainable sources of biofuels from plant biomass
- Understand the biological processes controlling greenhouse gas accumulation in the atmosphere (especially carbon dioxide and methane, key factors in global climate change)
- Gain insights into biogeochemical processes controlling the cycling of key nutrients in environments or the mobility of heavy metals and radionuclides at contaminated sites for which DOE has stewardship responsibilities

Bioenergy

The United States is the world's largest consumer of petroleum, and most of this energy is used for transportation and industry. This drives the DOE's focus on developing clean, renewable and sustainable alternative fuel sources from lignocellulosic biomass. Such fuels would ideally offer energy content on par with gasoline while being compatible with existing fuel distribution infrastructure. Sequencing projects at the DOE JGI that contribute to meeting this goal focus on one of three categories: terrestrial plants that can be used as feedstocks for biofuel production and their associated microbial communities (microbiomes); fungi, microbes and microbial communities that can break down the lignin and cellulose in plant walls; and organisms that can convert lignocellulose-derived sugars or lignols into biofuels or other bioproducts currently produced from petroleum, such as plastics.

Carbon Cycle

The global carbon cycle directly influences levels of atmospheric carbon dioxide and methane, which in turn affect the Earth's climate. As microbes constitute the largest component of the Earth's biodiversity and biomass, 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, particularly those found in less well-understood terrestrial, subsurface and terrestrial-aquatic interface ecosystems. Other projects aim to understand the contribution of biogenic emissions to atmospheric particle formation and evolution. This information is expected to contribute to better predictive models and strategies for mitigating the effects of increasing carbon dioxide and methane emissions on the global climate.

Biogeochemistry

Beyond the carbon cycle, many other DOE-relevant environmental processes are controlled by complex, interconnected biogeochemical reactions. The DOE JGI engages in projects that can couple a genome-enabled understanding of biological processes in the context of the physical, chemical and geochemical processes controlling the cycling and fate of key elements in environments impacting BER's energy and environmental missions. Microbes and microbial communities of interest to the DOE JGI as targets for sequencing include those involved in terrestrial nitrogen, phosphorus, sulfur and other macronutrient cycles that impact sustainable bioenergy crop growth or global carbon cycling, as well as those involved in the iron, sulfur and manganese cycles that mediate the transformation of DOE-relevant contaminants, such as heavy metals or radionuclides in soils, freshwater aquatic sediments and the subsurface.

Organizational Structure

Senior Management Team

Eddy Rubin
JGI Director

Dan Rokhsar
Chief Informatics Officer

Kjiersten Fagnan
NERSC User Support

Susannah Tringe
Deputy Director, User Programs

Tootie Tatum
Project Management Office

Len Pennacchio
Deputy Director, Genomic Technologies

Chia-Lin Wei
Sequencing Technologies

Atif Shahab
Institutional Informatics

Axel Visel
Deputy Director, Science Programs

Yasuo Yoshikuni
DNA Synthesis Science User Program

Trent Northen
Metabolomics Program

Dan Rokhsar
Eukaryote Super Program

Igor Grigoriev
Fungal Program

Jeremy Schmutz
Plant Program

Nikos Kyrpides
Prokaryote Super Program

Adam Rivers
Metagenome Program

Tanja Woyke
Microbial Program

Impact **2015**

Primary Users **Fiscal Year 2015**

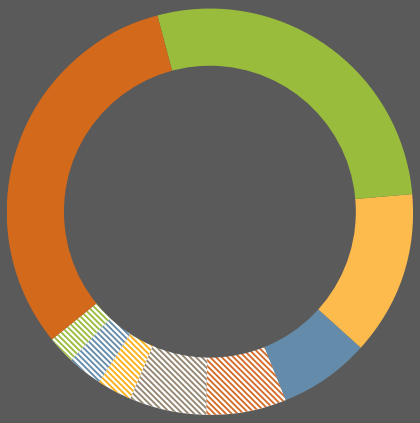
This category captures the primary users of the DOE JGI, which includes PIs and their collaborators on all user projects that were active during FY 2015. Each user is uniquely identifiable and is counted once per year regardless of the number of active projects in which he/she may be involved. This count does not include collaborators who are employed by the DOE JGI or funded through the DOE JGI's partner subcontracts.

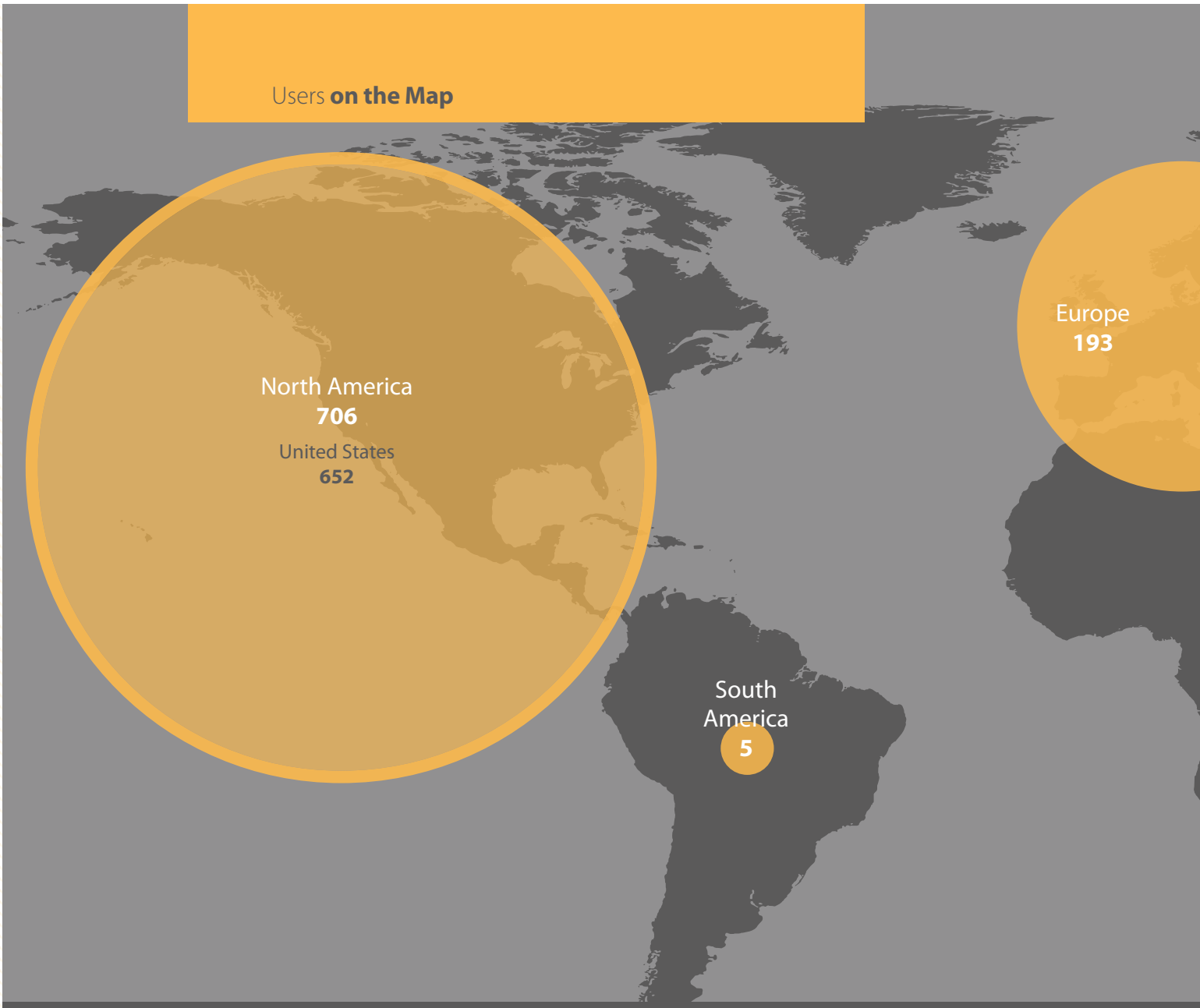
Users **by Institution Type**

Academic	676
Other	181
DOE National Laboratory	88
Company	13

JGI Expenses **FY2015**

- **32%** Science & Analysis
- **28%** Sequence/Data Generation
- **13%** R&D
- **7%** Operations
- **7%** Facilities
- **6%** National Energy Research Scientific Computing Center (NERSC)/IT
- **3%** Management
- 3% Emerging Technologoies Opportunity Program (ETOP)
- **2%** User Support





North America	706	South America	5	Europe	193		
United States	652	Brazil	2	Germany	38	Belgium	3
Canada	49	Columbia	2	France	28	Greece	3
Mexico	3	Peru	1	United Kingdom	28	Russian Federation	3
Costa Rica	1			Spain	18	Hungary	3
Panama	1			Netherlands	16	Portugal	3
				Austria	14	Turkey	2
				Switzerland	7	Denmark	2
				Finland	6	Iceland	1
				Italy	6	Estonia	1
				Sweden	5	Czech Republic	1
				Norway	5		

Africa	5	Asia	31	Australia & New Zealand	18
South Africa	3	China	7	Australia	12
Egypt	1	Japan	7	New Zealand	6
Senegal	1	India	6		
		Israel	4		
		Taiwan	3		
		Republic of Korea	2		
		Malaysia	1		
		Singapore	1		

Users of **JGI Tools & Data**

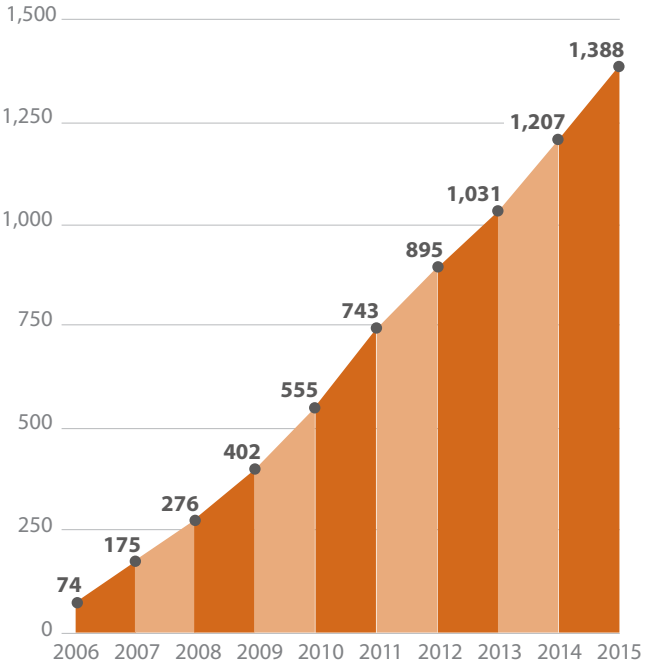
DOE JGI systems also support investigators who have utilized computational and/or data resources located at the DOE JGI, but are not included in the primary user count because their projects were not conducted as part of DOE JGI's user programs.

Workshops and Meetings	
Genomics of Energy & Environment	
10th Annual User Meeting Participants	**436**
Other Workshop Participants	**675**

Web Portal Visitors *(unique visits)*	
JGI Portal	**175,425**
IMG Systems	**267,970**
MycoCosm	**123,159**
GOLD	**125,436**
Phytozome	**117,771**

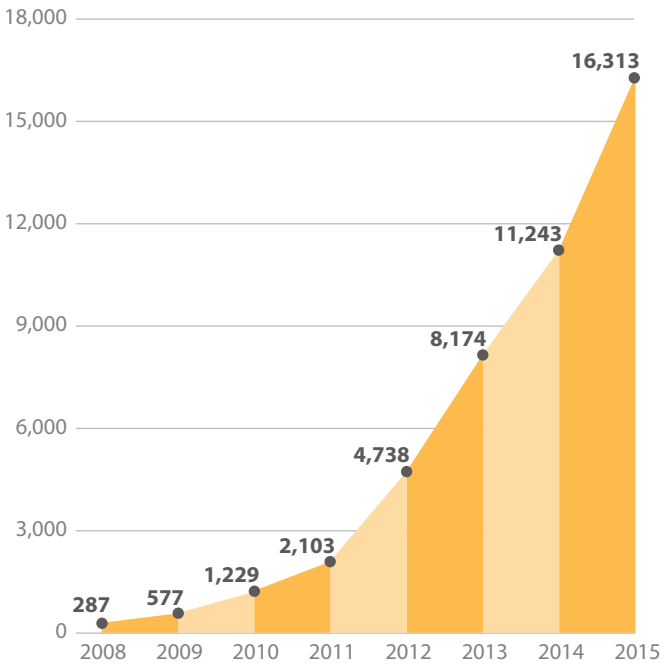
Sequence **Output**
(in billions of bases or GB)

Cumulative Number of **Scientific Publications**

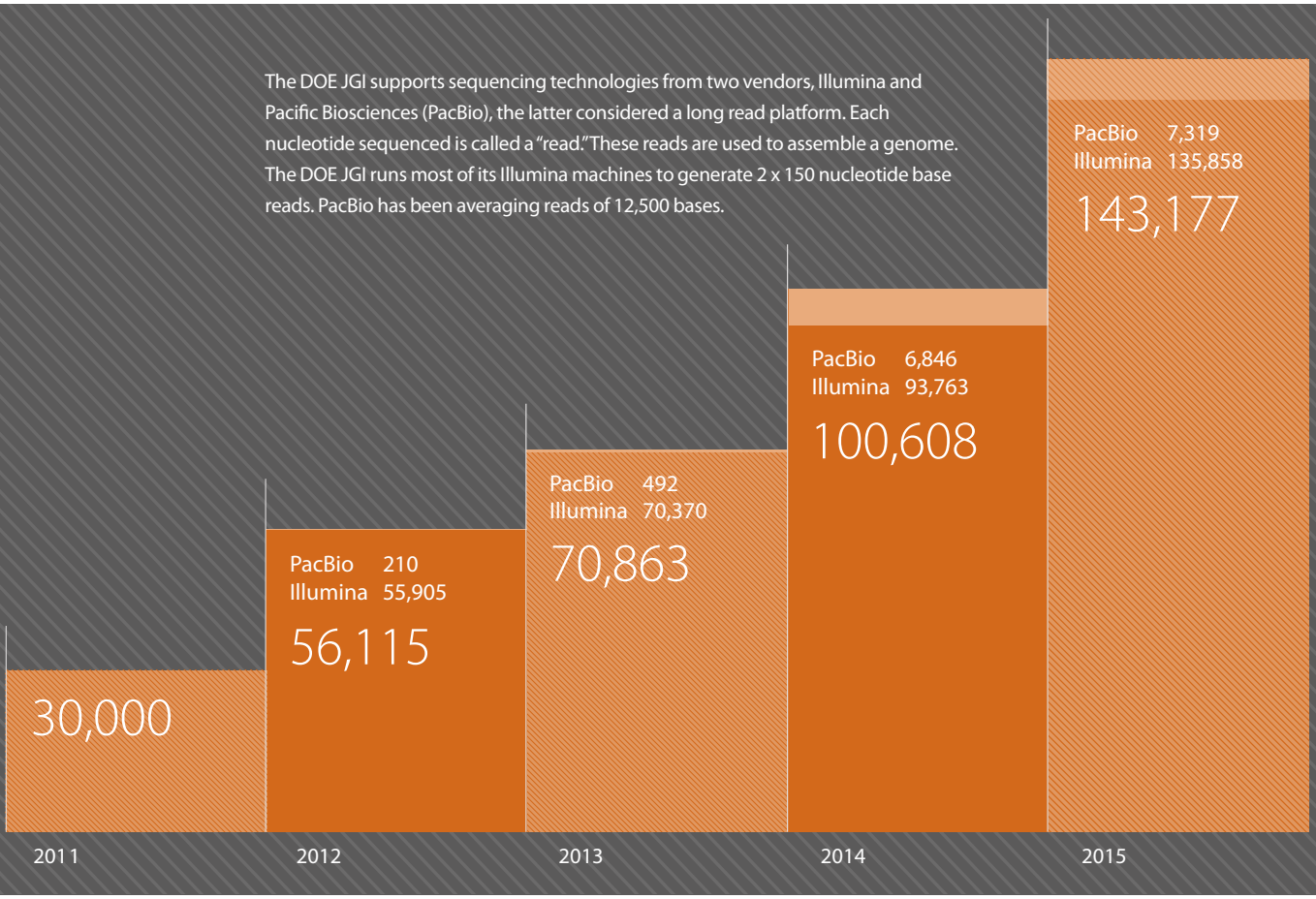


Papers published in 2011-2013 were cited nearly 5,000 times by researchers in 2014.

Cumulative Number of **Projects Completed**



The DOE JGI supports sequencing technologies from two vendors, Illumina and Pacific Biosciences (PacBio), the latter considered a long read platform. Each nucleotide sequenced is called a "read." These reads are used to assemble a genome. The DOE JGI runs most of its Illumina machines to generate 2 x 150 nucleotide base reads. PacBio has been averaging reads of 12,500 bases.



10 Years with IMG: **A Case Study**

The Integrated Microbial Genome (IMG) system marked its tenth anniversary with publication of the paper "Ten years of maintaining and expanding a microbial genome and metagenome analysis system" in the November 2015 issue of *Trends in Microbiology* **(http://bit.ly/JGI10IMGtim)**. IMG is developed and maintained by a team of software engineers and computer scientists from Berkeley Lab's Biosciences Computing Group led by Victor Markowitz, in partnership with a team of scientists from the DOE JGI's Prokaryotic Super Program led by Nikos Kyrpides.

When the first working prototype of IMG was launched in March 2005, it contained a total of 618 genomes — 318 bacterial, 25 archaeal, 15 eukaryotic and 260 bacterial phage — and had only a few users, mainly from the DOE JGI. "When we started developing the IMG system," recalls Kyrpides, "there were several data management and analysis systems for microbial genomics but IMG was unique in its comprehensive integration of genomic data across different data types and its powerful analysis tools provided through an intuitive user interface. There were also no systems for the emerging area of metagenomics, which IMG soon began to integrate as well."

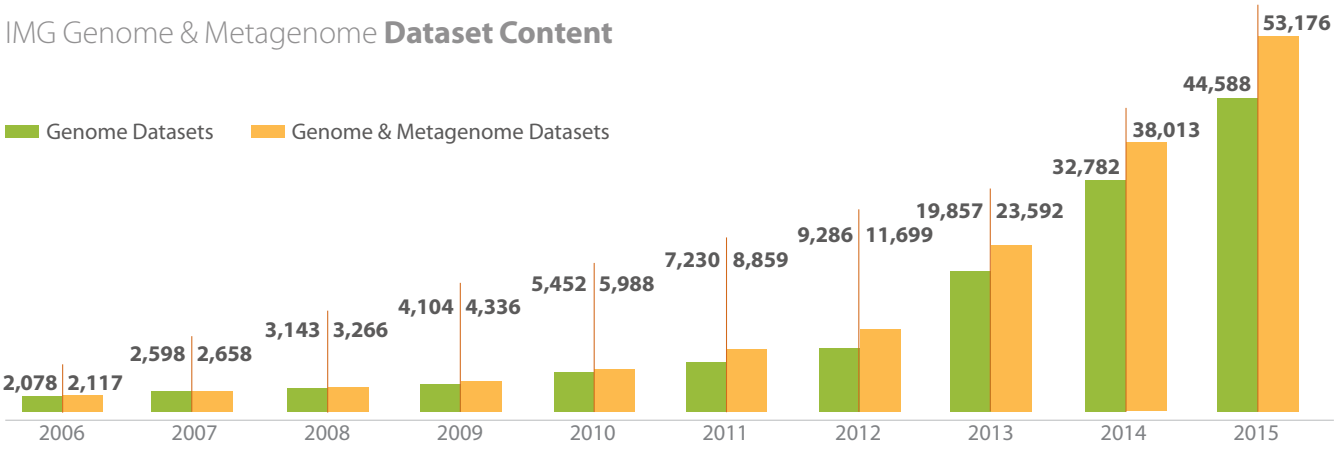
Today, IMG is one of the largest publicly-available data management and analysis systems for microbial genome and metagenome datasets, which also include entries such as plasmids, transcripts, and genome fragments. IMG contains about 44,500 genome datasets with 140 million genes and close to 8,600 metagenome datasets with over 38 billion genes, as shown in the figure on the next page. The system has more than 14,500 registered users from 95 countries across six continents, has contributed to thousands of published papers and has served as a tool for teaching genome and metagenome comparative analysis at tens of universities and colleges around the globe.

For the past 10 years, IMG has expanded rapidly with a massive increase in the number of datasets and the addition of new types of data and analysis tools supporting the integration of single cell genomes, "omics" data, as well as tools for analyzing metabolic pathways. Genome and metagenome data processing pipelines have been continuously extended in order to improve the consistency and completeness of annotations. As a DOE JGI production resource, IMG follows rigorous content update and tool maintenance procedures. New microbial genome and metagenome datasets are incorporated into IMG on a regular basis. As the number and size of genome and metagenome datasets processed by and integrated into IMG continues to grow, new genomic data organization paradigms and analysis methods will be developed. In particular, as thousands of microbial genomes per year are sequenced or extracted from metagenomes and single cells, strain-level genome datasets will be replaced by species-level pangenomes in order to support efficient comparative analysis in a rapidly expanding universe of datasets.

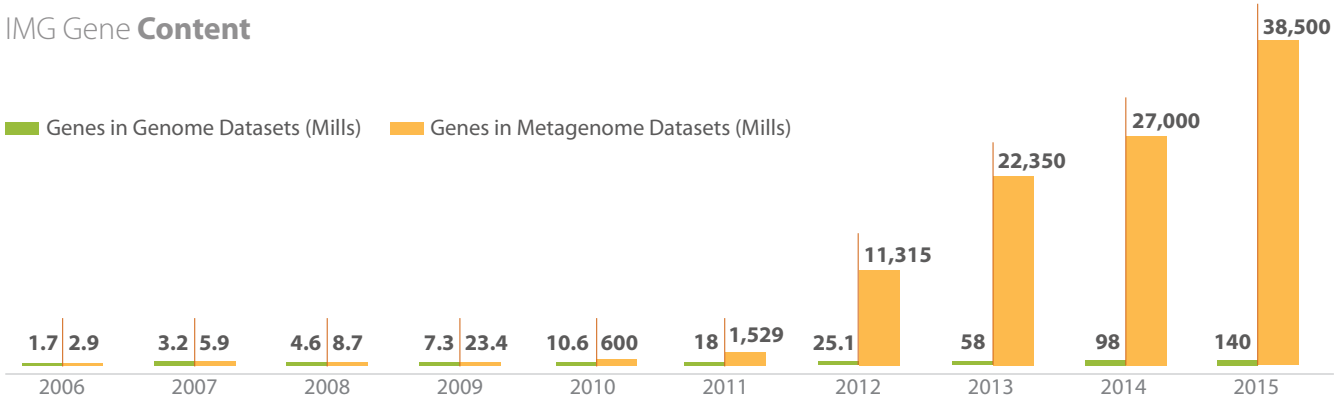
"In maintaining IMG as a leading genome and metagenome data management and analysis system," says Victor Markowitz, "we succeeded in overcoming the limitations of the data management tools and systems available to us. We are well-positioned to further improve the system once new database technology platforms become available at NERSC."

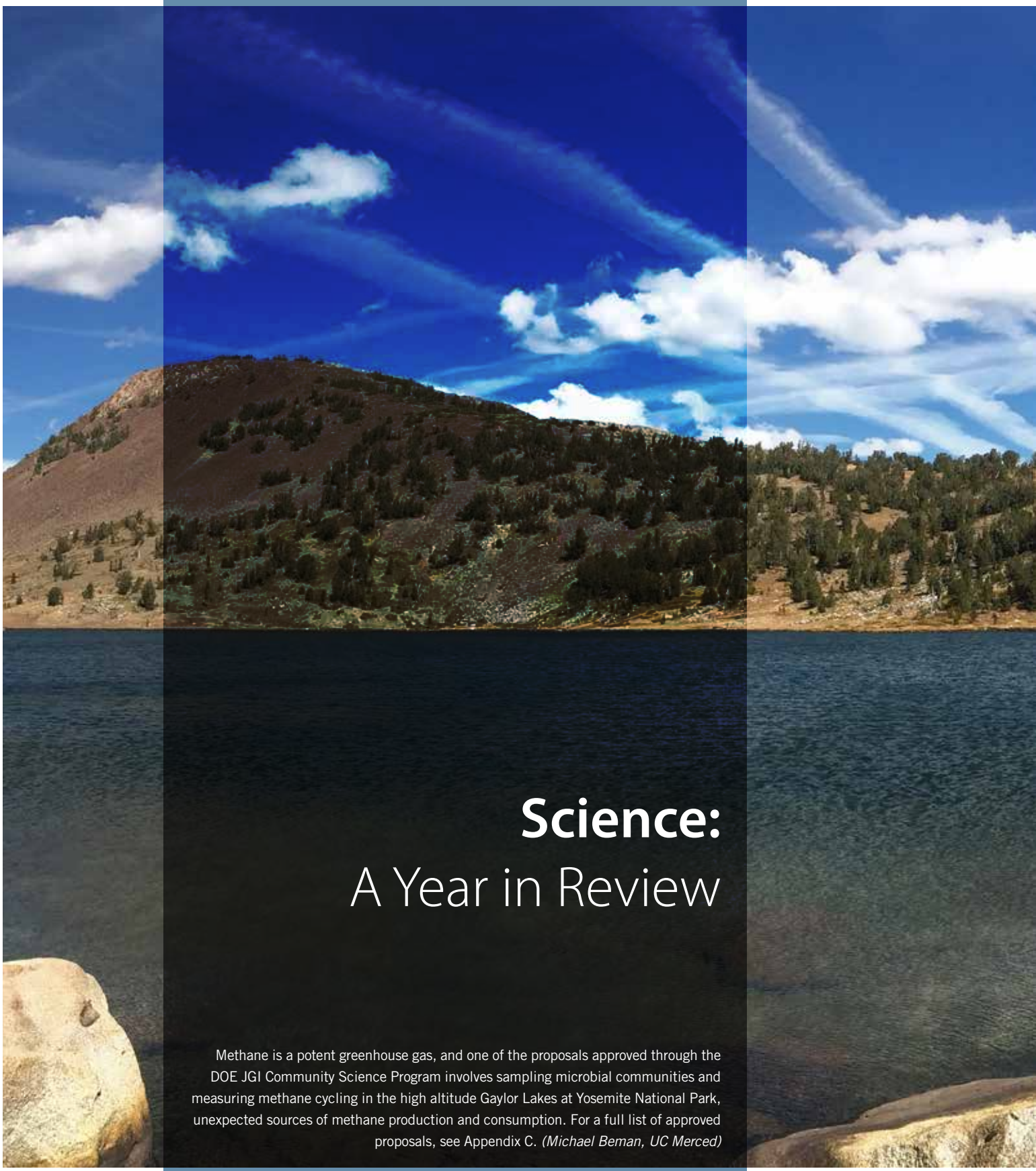
Victor Markowitz (left) and Nikos Krypides (right) lead the teams that developed and maintain the Integrated Microbial Genomes (IMG) system.

IMG Genome & Metagenome **Dataset Content**



IMG Gene **Content**





Science:
A Year in Review

Methane is a potent greenhouse gas, and one of the proposals approved through the DOE JGI Community Science Program involves sampling microbial communities and measuring methane cycling in the high altitude Gaylor Lakes at Yosemite National Park, unexpected sources of methane production and consumption. For a full list of approved proposals, see Appendix C. *(Michael Beman, UC Merced)*

The Department of Energy Joint Genome Institute (DOE JGI) is at the forefront of many scientific frontiers, including the search for undiscovered life. In a perspective piece published November 6, 2015, in the journal *Science*, DOE JGI Director Eddy Rubin and Microbial Program Head Tanja Woyke discussed why the time is right to apply genomic technologies to discover new life on Earth. They proposed the division of microbial life on Earth into three categories: explored, unexplored and undiscovered. This theme has been revisited in succeeding DOE JGI publications. For example, in the cover article of the April 2015 issue of *Genome Research*, a team including DOE JGI and Lawrence Berkeley National Laboratory (Berkeley Lab) researchers compared two different ways of using the same next-generation sequencing platform, one of which produced significantly longer reads than the other. The evaluation was done to help researchers determine how to best assemble meaningful sequence information, or even complete genomes, out of the genomic fragments from a population of organisms.

The DOE JGI's Emerging Technologies Opportunity Program (ETOP) continued to support targeted new technologies that the DOE JGI could incorporate into the portfolio of resources it offers to its users. Through this program, researchers at MIT and the University of Vienna, Austria, are developing technology to accelerate the functional characterization of organisms from metagenomic sequencing experiments. The early results of this ETOP project were described in the January 13, 2015, issue of the *Proceedings of the National Academy of Sciences*. Working toward a universally applicable technique that would allow researchers to evaluate microbial activities at the single-cell level, the team conducted proof-of-principle experiments to study the efficacy of a process that involved incorporating heavy water (deuterium) isotopes into molecules, and then using Raman spectroscopy to chemically identify the microbes. David Berry from the University of Vienna spoke about the ETOP project during the 2014 DOE JGI Genomics of Energy & Environment Meeting. Watch his presentation at **http://bit.ly/JGIUM9_Berry**.

To assist researchers in the quest for undiscovered life, DOE JGI teams have also developed computational tools enabling researchers to overcome the lack of a high-throughput process to review assembled genome sequences. A tool called ProDeGe, described in the June 9, 2015, issue of *The ISME Journal*, works on any type of genome sequence, removing contaminant sequences through a quick and automated process. Another tool, known as the MiSI method, described in the July 6, 2015, issue of *Nucleic Acids Research*, classifies organisms using genome sequence data using a combination of two metrics for determining how closely related two genomes are to complement conventional methods for species identification.

Poplar is a candidate bioenergy crop, and the first tree to have its genome sequenced. Since the DOE JGI published the poplar genome analysis in 2006 on the cover of the journal *Science*, researchers have developed plantations to study genetic variation within these populations and customize poplar for bioenergy production. Understanding how diseases impact these plantations is crucial to ensuring ample crop yields (see story on page 27.) These are rows of hybrid poplar at the Greenwood Resources tree farm in Boardman, Oregon. *(Image courtesy of Oregon Department of Forestry Flickr, CC BY 2.0)*

Bioenergy

Developing Eucalyptus Community Resources

Like cartographers who map landmarks, cities and countries to help people navigate, geneticists create genetic maps that link genes to particular traits or conditions. These maps help researchers figure out what chromosome the genes are on and where in the chromosomes these genes reside, and help suggest genes' functions. In a study that appeared online November 10, 2014, in *New Phytologist*, a team that included DOE JGI Plant Program Head Jeremy Schmutz reported the development of two genetic maps that improve the genome assembly of eucalyptus.

Grown on 40 million acres in 100 countries, the fast-growing and oil-rich eucalyptus is considered a candidate biomass energy crop, and its genome was sequenced under the DOE JGI Community Science Program (CSP). French researchers initiated whole-genome resequencing of two eucalyptus species, *Eucalyptus urophylla* and *Eucalyptus grandis*. Using the DNA sequence variations, or SNPs, found in the *E. grandis* genome sequence, published by the DOE JGI-led team in June 2014, the researchers then designed a high-quality SNP DNA-based genotyping array that allowed them to determine exactly which version of each gene is present in a given plant by analyzing crossbred offspring. These findings ultimately helped improve the *E. grandis* genome assembly available on the DOE JGI plant portal Phytozome, which, in turn, will improve the usefulness of the eucalyptus sequence for bioenergy studies.

Schmutz was also part of the team that first described the *E. grandis* genome of a Brazilian tree (BRASUZ1) in a *Nature* paper published in June 2014. A video featuring Jerry Tuskan, a longtime DOE JGI collaborator at Oak Ridge National Laboratory, and a member of the original team who worked on the eucalyptus genome, can be viewed at **http://bit.ly/eucalyptusTuskan**.

The June 2015 issue of *New Phytologist* was entirely dedicated to eucalyptus papers. The effort was spearheaded by DOE JGI collaborator Zander Myburg of the University of Pretoria, South Africa, one of the leads on the eucalyptus genome CSP project, and a co-author of several of the papers published in the special issue.

A stand of eucalyptus. *(Image courtesy of the University of Pretoria)*

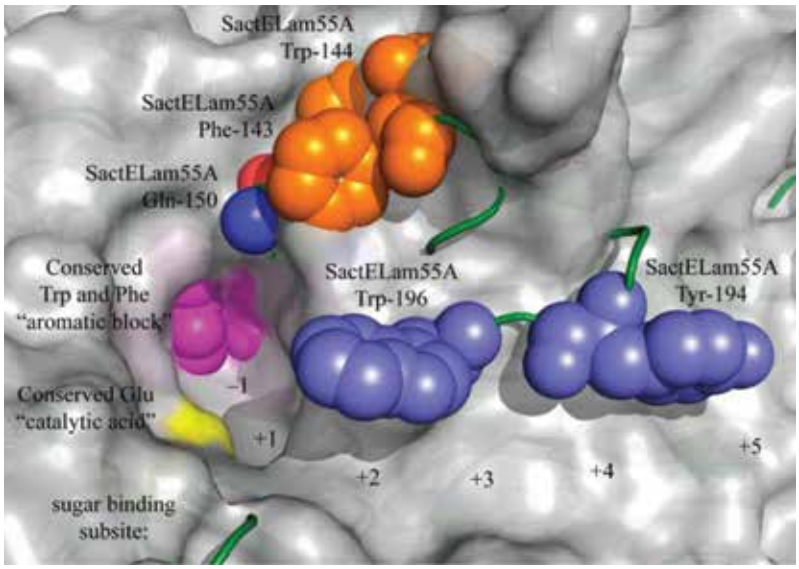
Producing a Complete Functional Map of an Entire Enzyme Family

Several of the microbial and metagenome projects conducted at the DOE JGI focus on biomass degrading microbial communities, such as those in the guts of insects and animals, to investigate how they break down lignocellulosic material to create energy. For example, when DOE JGI researchers published the cow rumen metagenome in *Science* in 2011, they identified nearly 30,000 candidate carbohydrate-active enzyme genes.

The process of functionally annotating each one of these genes, however, can be time-consuming. In a collaboration involving two DOE user facilities and a DOE Bioenergy Research Center, a team including researchers from the DOE JGI and the Great Lakes Bioenergy Research Center characterized the structure and function of a glycoside hydrolase (GH) 55 protein: SacteLam55A. The study was published May 8, 2015, in the *Journal of Biological Chemistry*.

Much of the work previously done on enzymes in the GH55 family of cellulose-degrading genes has been carried out on fungi. For example, the structure of the protein PcLam55A was derived using the white-rot fungus *Phanerochaete chrysosporium*, which had been sequenced by the DOE JGI.

The gene that encodes SacteLam55A was isolated from the microbe SirexAA-E in the gut of the pinewood-boring wasp *Sirex noctilio*. It was found when the microbe was grown on cellobiose, xylan and pretreated switchgrass samples, suggesting it has cellulolytic properties.



Combining the scientific resources of two DOE Office of Science user facilities, the Advanced Photon Source and the DOE JGI, researchers were able to examine the dynamic motion of several residues surrounding the active site of the substrate-bound structure of SacteLam55A, a glycoside hydrolase enzyme. *(Image from Bianchetti CM et al, (2015)* J Biol Chem.*)*

To determine the protein's structure, researchers relied on diffraction data collected at the Advanced Photon Source, a DOE user facility at Argonne National Laboratory, to develop high-resolution crystal structures. Through assays and techniques such as gene synthesis and cell-free protein translation, the team was also able to characterize the biochemistry and structure of the GH55 family. "The combination of gene synthesis, cell-free translation and assays using a diagnostic panel of substrates across the entire GH55 represents, to our knowledge, the most complete functional mapping of an entire GH family available to date," the team reported.

Mapping Water Management Traits in a Switchgrass Relative

The DOE considers switchgrass (*Panicum virgatum*) a potential bioenergy feedstock because it is a perennial grass that is also salt-tolerant and drought-tolerant, and thrives on marginal land. However, the plant is also polyploid, containing multiple copies of its genome, and this makes it difficult for researchers to map its genetic code. To develop a model system, the DOE JGI is sequencing a smaller, close relative of switchgrass with a shorter life cycle: Hall's panic grass (*Panicum hallii*).

In the January 2015 issue of *New Phytologist*, a team from the University of Texas-Austin, DOE JGI and other partner institutions utilized the genetic similarities between switchgrass and panic grass to identify traits related to the grasses' ability to thrive under various water availability conditions. Using two varieties of *P. hallii* — var. *hallii*, which thrives in very dry habitats, and var. *filipes*, which thrives in habitats with moderate water levels — the team used a statistical approach known as quantitative trait locus (QTL) analysis to identify genetic loci responsible for divergent traits that distinguish the two varieties.

"The same set of ecotype-differentiating traits found between xeric and mesic varieties of *P. hallii* are also involved in the divergence of upland and lowland ecotypes of the closely related bioenergy crop switchgrass," the team noted in their report. "Gaining insight into the molecular details of colocalizing QTL in these systems will be

Researchers looked at how two varieties of panic grass (*Panicum hallii*) handle water availability. *(Image courtesy of Tom Juenger)*

an important next step in understanding the factors constraining or facilitating adaptation and ecotypic differentiation to different habitats in nature."

The team members noted in their study that their findings suggest some traits involved in the formation of variants may evolve in tandem, while other traits evolve independently. This work will allow the association of genomic features in panic grasses with their responses to water availability. Given the similarity of panic grass to switchgrass, this will point researchers to those genomic regions that could serve as targets for further studies aimed at the improvement of switchgrass for bioenergy applications.

Stressful Searches for Algal Lipid Production

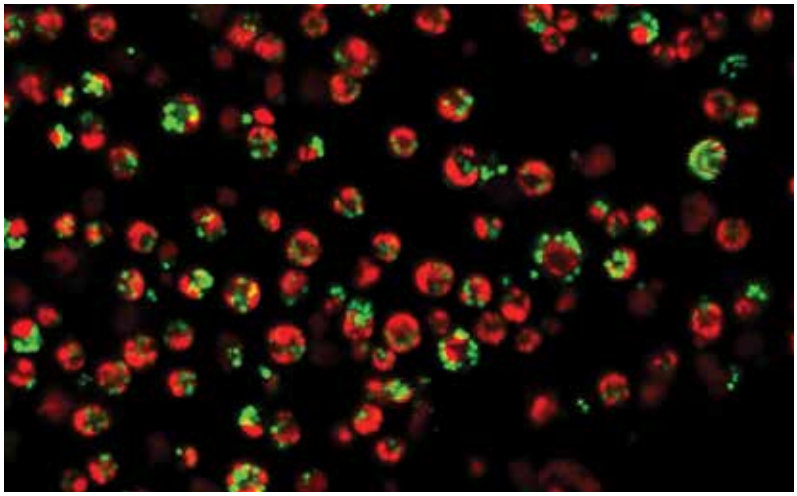
Some algae can produce energy-dense oils or lipids that can then be converted into fuels. However, these lipids are generally produced in response to stress, so it can be challenging to avoid "killing the goose that lays the golden eggs;" that is, to stress the algae just enough to produce lipids, but not enough to kill them. In the inaugural July 27, 2015, issue of *Nature Plants*, a DOE JGI team harnessed epigenomics and gene expression data to help algal bioenergy researchers develop more targeted approaches for producing lipids for fuel.

"We know how to stress the algae," said the study's first author, Chew Yee Ngan of the DOE JGI. "What we didn't know was how to keep the algae alive at the same time, until now."

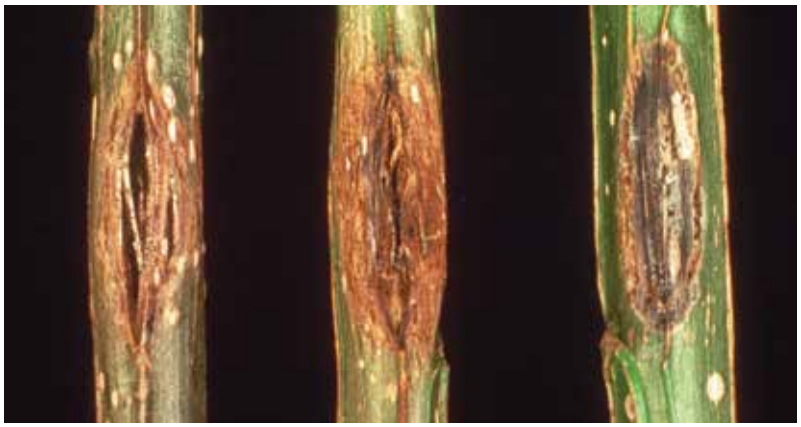
The DOE JGI has published over 75 percent of all publicly available algal genomes, including the *Chlamydomonas reinhardtii* — referred to as "Chlamy" by its research community — reference genome used for this study. Researchers analyzed the genes that are activated during algal lipid production, and in particular the molecular machinery that orchestrates these gene activities inside the cell when it produces lipids.

To find more protein factors that can regulate lipid production, the team cultured Chlamy cells and starved them of nitrogen or sulfur, and in response to the resulting stress conditions, Chlamy produces lipids. They then analyzed the complex of DNA and proteins known as chromatin, as well as the expression profiles or transcriptome. These analyses revealed what genes are being activated under stress conditions, which can then be compared with non-stressed Chlamy cells. Through careful analysis of genome-wide data sets, they narrowed down their search to identify two transcription factors that appeared to play a pivotal role in lipid accumulation, and then studied one of them, PSR1, in detail.

"The study also demonstrated how cells can be tricked into producing lots of lipid without dying of starvation by overexpression of PSR1, which is a strategy that could potentially be applied in other industrial algal species better suited for large-scale biofuel production," said study co-author Axel Visel, DOE JGI Deputy for Science Programs.



Algal cells of *Chlamydomonas reinhardtii* grown under nitrogen starvation conditions to produce lipids. The red signal is the autofluorescence from the chlorophyll of the cells, while the green indicates the lipid bodies following lipid staining with Lipidtox Green. *(Image prepared by Rita Kuo, DOE JGI)*



Cankers caused by the fungal tree pathogen *M. populorum* on poplar stems. *(T.H. Filer Jr., USDA, Bugwood.org CC BY-NC-3.0)*

Toward Developing Disease-Resistant Poplar Plantations

Trees such as poplar and eucalyptus are considered candidate bioenergy crops, but they are not yet grown in plantations with the consistency of agricultural crops. The transition from a forest full of a variety of tree species to a plantation growing only a single tree species has raised concerns about tree pathogens that would not have been so damaging to the diverse trees in a wild environment. One such fungal tree pathogen is Mycosphaerella populorum, which causes stem deformations known as cankers. In a plantation full of susceptible hybrid poplars, this pathogen can cause tree trunk breakage, defoliation, and is now considered "the most important factor limiting poplar plantations in eastern North America."

As reported in the March 17, 2015 *Proceedings of the National Academy of Sciences*, a team including DOE JGI researchers conducted a comparative analysis of two fungal tree pathogens, both previously sequenced at the DOE JGI, to find out how one of them has gained the capability to significantly damage hybrid poplar plantations. They compared the genomes of *M. populorum* and its close relative *M. populicola*, a pathogen that causes leaf damage to a much smaller range of species, to learn more about how fungal tree pathogens adapt to different hosts and ecosystems..

Along the team's findings is a cluster of secondary metabolites resulting from a horizontal gene transfer to the *M. populorum* genome that gives it the ability to infect woody tissues, a capability *M. populicola* does not possess. The team noted that their findings identified genes that, "can have major impacts on tree domestication because it transforms an innocuous coevolved pathogen into one that causes serve mortality and economic damage to plantation forestry." Characterizing these genes provides scientists targets for reducing or even eliminating the potential of this fungus to damage prospective bioenergy feedstocks.

Expanding Barley Genetics Resources

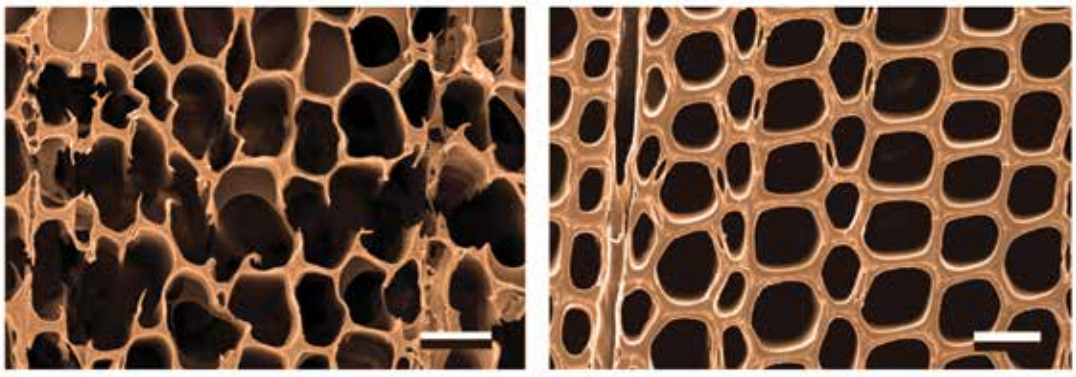
One of the reasons barley is a candidate bioenergy crop is that, as one of the most widely grown food crops, it has been bred to produce high yields. For commercial purposes, both the straw and the grain can be utilized to produce biofuels. However, producing a reference sequence for barley has been challenging because over 80 percent of the genome (which is already 67 percent larger than the human genome) is repetitive. To help with the international effort to produce a reference barley genome, the DOE JGI selected a proposal to develop a genetic map of the barley genome as a 2011 CSP project.

Building off of worldwide efforts, a team involving DOE JGI researchers recently reported that nearly two-thirds of barley's gene space has been mapped. In a study first published September 21, 2015, in *The Plant Journal*, the team identified and sequenced over 15,000 bacterial artificial chromosomes containing barley genes, comprising roughly 1.7 billion base pairs (Gbp) of sequence out of the estimated 5.1 billion base pairs (Gbp) that make up the barley genome.

Sequencing and assembling the barley genome is challenging not just because of its 5.1 billion base pairs size, but because over 80 percent of the sequence is repetitive. *(Craig Nagy, Flickr, CC BY-SA 2.0)*

"These ~1.7 Gbp of gene-rich genomic sequence expand our knowledge of the characteristic features of the gene-containing regions," the team reported. "Furthermore, this resource will improve the speed and precision of map-based cloning and marker development in barley and closely related species while supporting ongoing efforts in obtaining a complete reference sequence of barley."

The researchers made use of an earlier project in which a team, also involving DOE JGI researchers, evaluated a method for assembling complex plant genomes. Using the technique called POPSEQ, researchers rapidly and inexpensively assembled barley genome datasets as proof of principle. The knowledge of these particular genes will materially help the community of scientists interested in research on barley exploit them. Importantly, having a much higher-resolution sequence-based map of the barley genome will make it easier for scientists to search out and identify genes involved in traits of interest for a variety of uses, among them generation of biomass for energy.



Scanning electron microscopy (SEM) image of pine that was substantially eroded by the white rot Phlebiopsis gigantea compared with a SEM of sound wood (right). Bar = 40 µm. *(Image from Hori et al. (2014)* Plos Genet.*)*

How a White Rot Tackles Freshly-Cut Wood

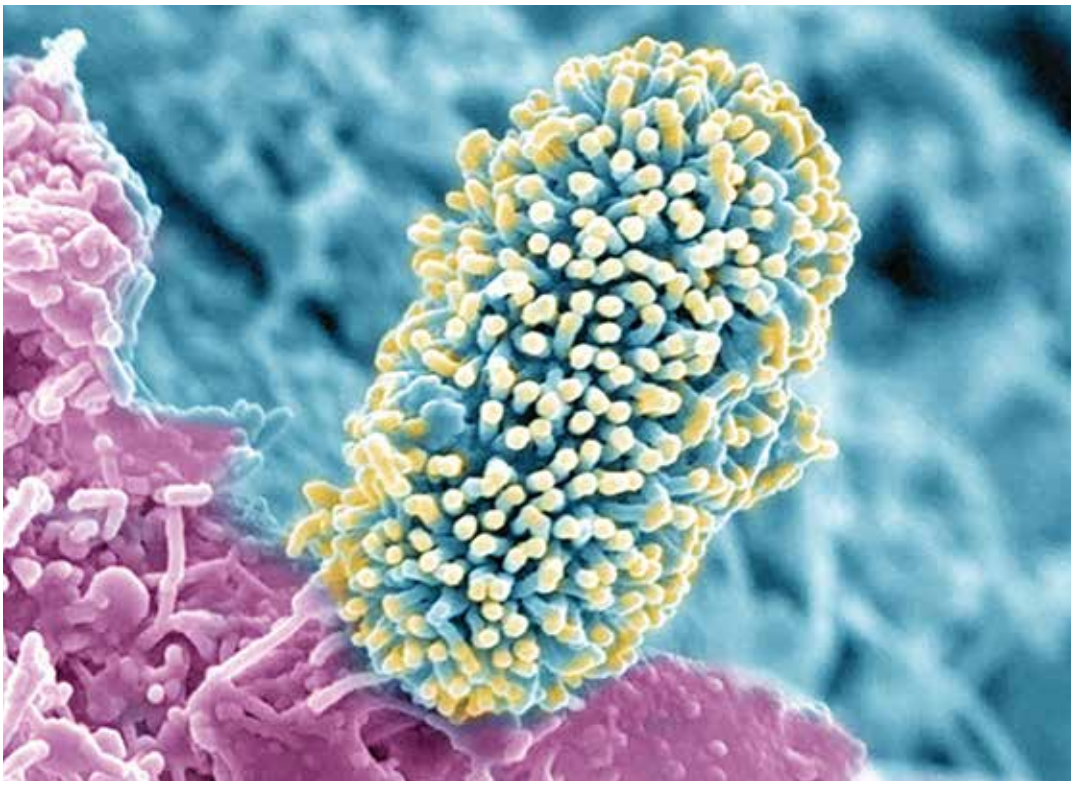
As plant biomass is a sustainable alternative energy source the DOE is interested in developing, researchers are probing the fungal processes involved in breaking down and converting plants into sugars, information that they hope to harness at industrial-scale settings.

Researchers at the DOE JGI have been sequencing and analyzing two primary groups of fungi: brown rots and white rots. While both groups can break down the cellulose and hemicellulose in plant cell walls, only white rots can break down lignin as well. Many of the white rot fungal genomes sequenced and analyzed at DOE JGI are

known to break down decaying wood. In a study published December 4, 2014, in *Plos Genetics*, however, a consortium including DOE JGI researchers reported on the mechanisms by which the white rot *Phlebiopsis gigantea* breaks down plant cell walls and resinous components in the freshly-cut sapwood of conifers.

In one aspect of the study, researchers monitored the progress of the fungus on wafers of sapwood from aspen, pine and spruce over three months and compared these results against un-inoculated wafers. In another experiment, they measured the fungus' levels of gene expression in a scenario where glucose had been extracted from freshly-harvested pine wood using acetone, and a scenario where the glucose had not already been extracted from the freshly-harvested wood.

"The results advance understanding of the early and exclusive colonization of coniferous wood by *P. gigantea* and also provide a framework for developing effective wood protection strategies, improving biocontrol agents and identifying useful enzymes," the team concluded.



An intricately structured soil bacterium, less than a micron in size, makes its home on the root surface of an *Arabidopsis* plant. *(Courtesy of Pacific Northwest National Laboratory, Flickr, CC BY-NC-SA 2.0. Image was captured with the Helios Nanolab dual-beam focused ion beam/scanning electron microscope at the Environmental Molecular Sciences Laboratory (EMSL) and was created by Alice Dohnalkova.)*

Finding Friends in the Root Microbiome

Microbial communities thriving at the nexus where roots and soil intersect play important roles in plant health and growth. A plant's immune system can distinguish between friends and foes among these microbes, and upon detecting pathogens, can produce regulatory chemicals called phytohormones to activate a defensive response.

In the August 21, 2015, issue of *Science*, a team led by collaborators at the University of North Carolina looked at roles of three phytohormones in controlling the composition of the root microbiome in the model plant *Arabidopsis thaliana*. The work was initiated as part of a DOE JGI Grand Challenge project studying the root-associated microbiomes of *Arabidopsis*, then grew into both a CSP project and a National Science Foundation-funded microbial systems biology grant investigating multiple plant species in the greenhouse and the field. "This is a key focus area for the DOE JGI, as we believe cultivation of beneficial microbiomes will be critical to sustainable bioenergy crop production in the future," said DOE JGI Metagenome Program Lead Susannah Tringe.

To better understand the impact of phytohormones on root microbiomes, the team compared wild type root microbiomes with those of mutants lacking the ability to produce at least one of the following phytohormones: salicylic acid, jasmonic acid and ethylene. The team examined root microbiome assembly using bacterial isolates, originally cultured from plants grown in wild soils, to colonize sterile roots in a sterile soil-like substrate. The resulting plant-associated communities were then tracked for eight weeks following inoculation using a sequence-based readout, revealing very different patterns of colonization in the phytohormone mutants.

A Wood-Digesting Strategy Unlike Any Other

For millennia, mariners were concerned by the shipworm, a worm-like, wood-eating marine clam also known as the "termite of the sea." However, what vexed pre-18th century sailors may prove useful for generating biofuels. In the November 25, 2014, issue of the *Proceedings of the National Academy of Sciences*, DOE JGI researchers and colleagues at the Ocean Genome Legacy Center of New England Biolabs at Northeastern University described the novel strategy by which the shipworm breaks down and digests wood.

"Most animals, including people, have beneficial bacteria in their digestive system to help them digest food and would quickly become sick and malnourished without them," said Daniel Distel, Director of the Ocean Genome Legacy Center, who originally proposed the project under the DOE JGI Community Science Program. "But shipworms have no bacteria in the part of the gut where their food is digested. Instead, they house symbiotic bacteria inside specialized cells in their gills, a location far removed from the gut."

The team is still pondering the reasoning behind having the wood-degrading enzymes produced away from where digestion actually takes place, but they suggested that this strategy allows the shipworm to serve as a simple model system to work out the minimal enzyme requirements for efficiently breaking down cellulose.

"Because only selected wood-degrading enzymes are transported, the shipworm system naturally identifies those endosymbiont enzymes most relevant to lignocellulose deconstruction without interference from other microbial proteins," they wrote in the paper. "Thus, this work expands the known biological repertoire of bacterial

The "termite of the sea" may prove useful for generating biofuels. *(Dan Distel, Ocean Genome Legacy Center of New England Biolabs)*

Biosciences researchers Aindrila Mukhopadhyay and Heather Jansen at JBEI work on a related project that involves engineering *E. coli* to produce bio-gasoline.

endosymbionts to include digestion of food and identifies new enzymes and enzyme combinations of potential value to biomass-based industries, such as cellulosic biofuel production."

Thus organisms that used to be dreaded at sea may supply both insights and tangible tools for the generation of biofuels from woods on land. To learn more, watch Distel's presentation, "How to Eat a Wooden Ship: A Genomic View of Wood-Eating Bacterial Endosymbiosis in the Shipworm *Bankia setacea*," from the DOE JGI's 2011 Annual Genomics of Energy & Environment Meeting at **http://bit.ly/JGIUM6Distel**.

Enhancing Microbial Pathways for Biofuel Production

Terpenes are hydrocarbons in plants, and bioenergy researchers consider them high-energy metabolites that could be used for producing biofuels from plant feedstocks. For example, terpene production in eucalyptus is of interest to members of an international consortium — including DOE JGI scientists — that described the eucalyptus genome in *Nature* (see page 23).

In the January 2015 issue of *Applied and Environmental Microbiology*, DOE JGI researchers collaborated with a DOE Bioenergy Research Center, the Joint BioEnergy Institute (JBEI), to find ways of enhancing terpene yield in bacteria. In previous studies, JBEI researchers had reported that bisabolane, a biofuel resulting from the precursor terpene bisabolene, could serve as an alternative to diesel fuel. They wanted to find a way to improve terpene production in *Escherichia coli* using one metabolic pathway, called DXP, which they expected to be more efficient in terms of final yield than the alternative mevalonate pathway. To develop a novel route that would take C5 sugars (such as the xylose formed when hemicellulose is broken down) to terpenes, they used a directed-evolution strategy and deleted specific genes involved at a key point in the pathway.

The results led to their discovery of two novel routes — one that arose through spontaneous mutations and one found through overexpression of a selected candidate gene — for producing the terpene and candidate biofuel bisabolene. The team also noted that applying the engineering process to the DXP pathways in plants and algae "could provide a more direct link from carbon fixation (Ru5P in the Calvin cycle) to the terpene pathway."

Permafrost-associated soils cover nearly a quarter of the terrestrial land surface, contain more than a third of the world's soil carbon, and are thawing due to climate change. For more information about the permafrost studies DOE JGI researchers are working on, go to page 33.

Carbon Cycle

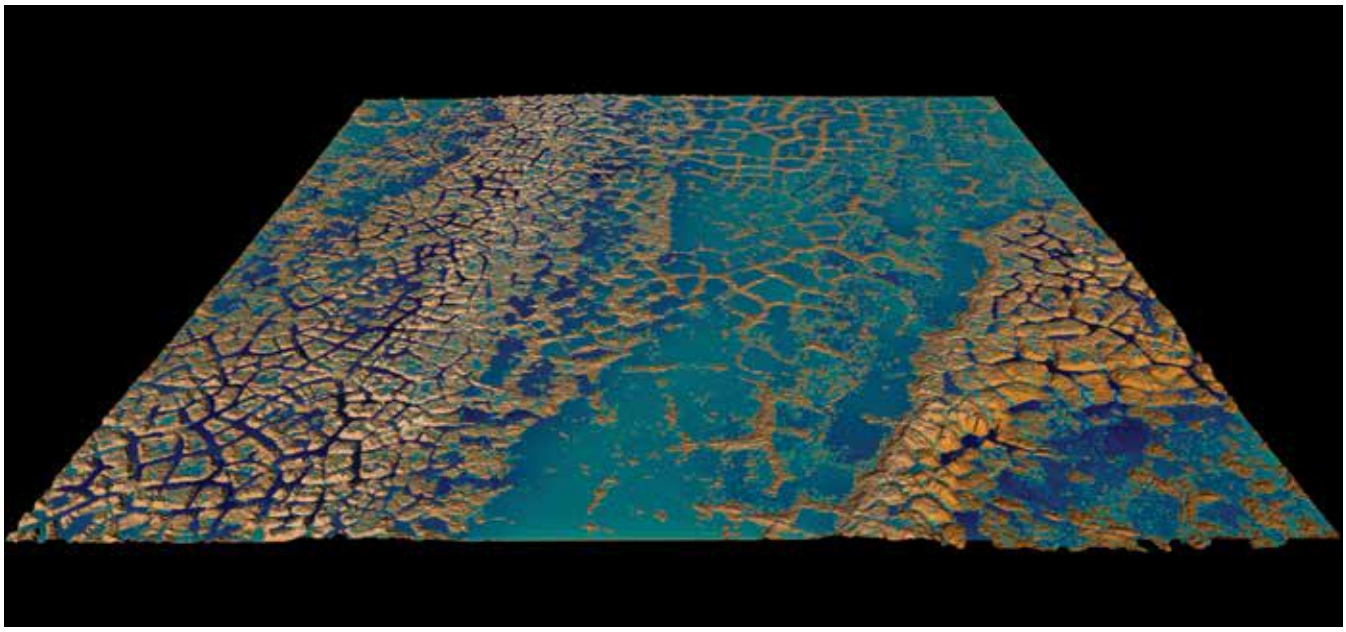
Omics Applied Toward Permafrost Studies

One of the abiding concerns regarding permafrost is that as global temperatures rise, as is projected to happen over the coming centuries, soils may thaw completely, resulting in the largest contribution of carbon transferred to the atmosphere by a single terrestrial process.

To help understand the processes that control the conversion of organic matter to carbon dioxide and methane, DOE JGI scientists were part of a team that reported on the application of multiple molecular technologies, collectively referred to as "omics," to better characterize microbial activities in a paper published online March 4, 2015, in the journal *Nature*.

Microbial ecologist Janet Jansson from Pacific Northwest National Laboratory led the team, which used metagenomics, or environmental genomics, to identify the phylogeny — i.e., the history of organismal lineages — of the communities' microbial members and the functional gene composition. Metatranscriptomics allowed the team to determine which genes were being expressed. Finally, metaproteomics provided insights on which proteins were actually produced.

Comparing the findings from using three omics types on the three soils provided insight into the linkages between omics data and elemental cycling pathways. In the thawed soils of the thermokarst bog, for example, they found the highest rates of methane production and identified several microbes involved in this pathway. Additionally, several genes involved in methanogenesis were detected in both the metagenomic and metatranscriptomic data sets and corresponding proteins in the metaproteomic data sets. Three draft methanogen genomes were identified, and comparisons with sequenced methane producers suggest these are previously undescribed microbes.



Using computer simulations such as this snowmelt draining from polygonal ground near Barrow, Alaska, scientists are working to understand the fate of carbon stored in Arctic soils. *(LANL, Flickr CC BY-NC-ND 2.0)*

Study first author Jenni Hultman prepping permafrost samples. The team investigated three types of Alaskan soils, ranging from completely thawed to completely frozen. *(Janet Jansson, Pacific Northwest National Laboratory)*

The work builds on findings from a previous collaboration between the DOE JGI and Jansson. The work done by Jansson and her colleagues is just one of the ecosystem studies being conducted by the DOE in Alaska. Through the Next-Generation Ecosystem Experiments (NGEE Arctic) project in Barrow, Alaska, a consortium of academic institutions and national laboratories is developing a process-driven ecosystem model that will allow researchers to better predict the evolution of Arctic ecosystems in a changing climate. Jansson is currently leading a CSP project at the DOE JGI for sequencing of samples collected for the NGEE Arctic project.

Mutualism for Beneficial Environmental Adaptations

Below ground, plants interact with many species of fungi and other microbes in the surrounding environment. These exchanges can impact the plant's health and tolerance to stressors such as drought or disease, as well as the global carbon cycle. Mycorrhizal fungi live in the roots of host plants, where they exchange sugars that plants produce by photosynthesis for mineral nutrients that fungi absorb from the soil. They are of interest to bioenergy researchers because they play roles in maintaining the health of candidate feedstock crop trees. Recent studies

indicate that mycorrhizal fungi also play a significant role in belowground carbon sequestration, which may mitigate the effects of man-made carbon dioxide emissions.

To understand the basis for fungal symbiotic relationships with plants, a team of DOE JGI researchers and longtime collaborators at the French National Institute for Agricultural Research (INRA) and Clark University conducted the first broad, comparative phylogenomic analysis of mycorrhizal fungi, which appeared in the April 2015 issue of *Nature Genetics*. Drawing on 49 fungal genomes, 18 of which were sequenced for this study, these researchers describe how the comparative analyses of these genomes allowed them to track the evolution of mycorrhizal fungi. The results help researchers understand how plants and fungi developed symbiotic relationships, and how the mutualistic association provides host plants with beneficial traits for environmental adaptation.

To understand the genetic shifts underlying the repeated origins of mycorrhizal lifestyles, the researchers focused on enzymes that degrade plant cell walls from 16 gene families associated with plant cell wall degradation.

The analyses of the fungal genomes and fossils suggested that in comparison with brown-rot fungi and white rot fungi that evolved over 300 million years ago, ectomycorrhizal fungi emerged fairly recently from several species and then spread out across lineages less than 200 million years ago. The team also found that up to 40 percent of the symbiosis-induced genes were restricted to a single mycorrhizal species.

"Together these studies tell a story about how mushroom-forming fungi evolved a complex mechanism for breakdown of plant cell walls in 'white rot' and then cast it aside following the evolution of mycorrhizal associations, as well as the alternative decay mechanism of 'brown rot,'" said David Hibbett of Clark University, one of the study's senior authors. "The other major part of the story is that in mycorrhizal lineages there is a huge turnover in genes that are upregulated in the symbiosis — many of these have no homologs in even closely related species, suggesting that the evolution of the symbiosis is associated with massive genetic innovation."

Mycorrhizal fungi include some of the most conspicuous forest mushrooms, such as fly agaric (*Amanita muscaria*). *(Francis Martin, INRA)*

A transmission electron micrograph of one of the smallest known eukaryotic algae, *Micromonas*. *(Transmission electron microscopy image by A.Z. Worden, T. Deerinck, M. Terada, J. Obiyashi and M. Ellisman, the Monterey Bay Aquarium Research Institute and NCMIR)*

Tracking Down the Origins of Algal Light Ancestors

Phytochromes are protein-based sensors found in plants, fungi and bacteria that regulate events in an organism's life cycle based on changes in light wavelengths received from the sun. In plants, phytochromes play a key role in development as well as perceiving and avoiding shade. However, their origins are poorly understood. To learn more, a team including the Monterey Bay Aquarium Research Institute (MBARI) and DOE JGI researchers sequenced and analyzed multiple sets of all RNA molecules in a cell (or gene transcripts, i.e., transcriptomes) of several kinds of phytoplankton. In the study published November 4, 2014, in *Proceedings of the National Academy of Sciences*, they found that phytochromes were present in prasinophytes and glaucophytes, two phylogenetic lineages of photosynthetic algae.

Among the transcriptomes studied was one from the tiny marine phytoplankton *Micromonas pusilla*, which was the focus of a previous collaboration between MBARI researcher Alexandra Worden and the DOE JGI published in *Science* in 2009.

Since phytochromes are master regulators in land plants, researchers think they likely play critical roles in algal growth and biomass production. Additional studies of phytochrome diversity and regulatory roles could reveal ways to exploit more of the light spectrum for the generation of bioenergy compounds.

Short-Term Results of Wetlands Restoration

Wetlands cover less than 10 percent of the Earth's land surface, but they can trap as much as 30 percent of global soil carbon, as well as emit nearly 40 percent of natural global methane emissions. In the late 19th and early 20th centuries, several freshwater tidal marshes in California's Sacramento-San Joaquin Delta were drained and converted for agricultural purposes. A century later, the U.S. Geological Survey and the California Department of Water Resources launched a pilot project focused on Twitchell Island to restore these wetlands for a number of reasons, including the mitigation of levee failure, land surface subsidence and erosion.

Before agencies can consider long-term wetland restoration projects, they want to learn more about whether or not restored wetlands can serve as carbon sinks or act as greenhouse gas emitters. In a study published May 19,

After a day collecting samples from Twitchell Island, located in the Sacramento-San Joaquin Delta of California, U.S. Geological Survey scientist and study co-author Lisamarie Windham-Myers (left), study senior author Susannah Tringe (center), and study first author Shaomei He (right) go over their preliminary findings. *(Image by David Gilbert, DOE JGI)*

2015, in the journal *mBio*, a team led by DOE JGI User Programs Deputy Susannah Tringe reported on microbial community composition and carbon emissions patterns from data collected at different times from Twitchell Island sites near the inlet and leading toward the interior of the wetland.

The team collected an abundance of data relating to microbial community composition and greenhouse gas emissions. DNA collected from the soil samples was sequenced and analyzed at the DOE JGI. One of the team's findings was that the microbial communities present in the restored wetlands were more diverse than the composition of the microbial communities in an adjacent cornfield. Additionally, the river water became slightly more acidic as it moved from the river inlet toward the wetland interior, likely due to the decomposing plant mass found "inland." A similar pattern was observed regarding the methane emissions, with lower emissions detected at the inlet compared with the interior sites.

"The differences in microbial community composition and functional profiles reveal complex interactions among functional guilds and among wetland plants, microorganisms and their environment," the team reported. "Based on these findings, more specific studies can be carried out to evaluate the impact of wetland water chemistry and hydrology on CH_4 emission and carbon sequestration before large-scale restoration projects are implemented."

In 2011, Susannah Tringe, Metagenome Program Lead, received a DOE Office of Science Early Career Research Program grant to study microbial communities in restored wetlands. Watch a video of Tringe introducing the project and the work at Twitchell Island at **http://bit.ly/JGIwetlandsvid**. In recent years, Tringe has expanded her focus to included restored wetlands around the San Francisco Bay Area; watch a recent video regarding the project at **http://bit.ly/JGI15CCMarshes**.

Grand Prismatic Spring at Yellowstone National Park. Yellowstone's geothermal ecosystem contains an extensive array of unique high-temperature environments that host understudied microbes that play undetermined roles in global biogeochemical cycles. By learning more about the microbial diversity in these environments, researchers hope to harness the information to overcome energy and environmental challenges.

Biogeochemistry

White Rot Fungi for Contaminant Cleanup

Fungi play many roles that address DOE mission areas, particularly in bioenergy and environment applications. Their capabilities in breaking down wood and leaf litter are being studied by researchers at several institutions including the DOE JGI for optimizing the production of advanced biofuels from sustainable plant sources.

One area in which metabolically versatile fungi have received less attention is their ability to clean up contaminated sites. The button mushroom (*Agaricus bisporus*), for example, is known for its taste and its ability to decompose leaves and wood in forests, while related species have been shown to help remove heavy metals from contaminated areas. This capability is what researchers in Massachusetts want to harness to remove the dense, viscous Bunker C fuel oil from the waters of Blackstone Canal, which were contaminated during a 1999 fire at the Fisherville Mill.

In a report published June 25, 2015, in *Plos One*, a team including DOE JGI scientists assessed the hydrocarbon-degrading capabilities of six white-rot fungi — known for their ability to break down both lignin and cellulose in plant matter — for potential use in a mycofiltration system that would cleanse the contaminated canal water. Their findings suggest that beyond the capability of these fungi to degrade plant cell wall components and improve the production of fuels from biomass, they also are capable of breaking down organic pollutants such as polycyclic aromatic hydrocarbons, polychlorinated biphenyls and organochlorine pesticides

All told, the team evaluated six white-rot fungi: *Phlebia radiata*, *Trametes versicolor*, and *Pleurotus ostreatus*, all of which were sequenced and analyzed at the DOE JGI through various CSP projects; and *Irpex lacteus*, *Trichaptum biforme* and *Punctularia strigosozonata*. The last of these was sequenced and analyzed at the DOE JGI to see how the oil affected its gene expression, and was studied separately.

The fungal strains were inoculated onto white pine sawdust, the primary wood being considered for use in the mycofiltration unit. Researchers then added the cultures to petri dishes along with some Bunker C oil and incubated for six months to see how well the fungi broke down the hydrocarbons.

The team found that after six months, most of the 10-carbon and 14-carbon alkane chains in the oil samples had indeed been broken down by five strains of white-rot fungi. Additionally, in a 20-day test, the sixth strain, *P. strigosozonata*, degraded 99 percent of the initial C10 alkane. "The degradation of these compounds demonstrates the promise that white-rot fungi hold as agents of bioremediation for similar oil compounds," the team wrote. Though the processes by which the white-rot fungi can break down hydrocarbons have not yet been determined, the team recommends that the tests be repeated using the actual bioremediation system being developed to see how these conditions would affect the results seen in the laboratory.

The white rot *Trametes versicolor*, colloquially known as the turkey tail, was one of six fungi tested for potential bioremediation use in Massachusetts. *(Image cropped from photo by Luc De Leeuw, Flickr, CC-BY-NC-SA-2.0)*

Understanding Nitrogen-Fixing Cyanobacteria

In the global oceans, cyanobacteria such as *Prochloroccus* and *Synechococcus* are critical because despite their small size, their sheer abundance involves them in the fixation of nearly half of total atmospheric carbon captured annually and its conversion into organic material through photosynthesis. Published April 7, 2015, in the *Proceedings of the National Academy of Sciences,* a consortium including DOE JGI researchers sequenced and analyzed the genome of a strain of the cyanobacteria *Trichodesmium erythraeum* isolated off the coast of North Carolina. Unlike *Synechococcus* and *Prochlorococcus*, this cyanobacterium plays a key role in fixing atmospheric nitrogen. The absence of this nutrient can be a limiting growth factor.

DOE JGI researchers have previously collaborated on studies that indicated that some cyanobacteria selectively reduce their genomes to thrive in nutrient-poor environments. In the case of the *Trichodesmium* genome however, the team found that just under two-thirds of the total DNA was predicted to encode proteins and of the remaining non-coding DNA, nearly 90 percent was expressed as RNAs. Additionally, this non-coding DNA contained numerous DNA repeat sequences. The team compared the genome of this *T. erythaeum* strain to a draft genome of another *T. erythaeum* strain isolated from the tropical Atlantic, and a partial draft of a different *Trichodesmium* species isolated in Hawaii. They found that all three *Trichodesmium* strains had similarly low gene densities along with large genomes. One suggestion offered by the team was that the noncoding regions might contain regulatory elements required for the proper expression of needed genes for growth, both in the lab and in the wild.

The golden-brown surface mats that indicate a *Trichodesmium erythaeum* bloom have led to them being called "sea sawdust." *(Image by FWC Fish and Wildlife Research Institute, Flickr, CC BY-NC-ND 2.0)*

Another theory the team proposed was, "that the intergenic regions experience gradual inflation during certain evolutionary intervals characterized by [algal] bloom-driven selective sweeps." Unlike other cyanobacteria, *Trichodesmium* can thrive both in large clusters and with other microbes. In the latter situation, their shared co-existence could lead to multiple *Trichodesmium* subpopulations and contribute to the non-streamlined genomes observed. The relationship of "extra" DNA in Trichodesmia and this important capability, central to understanding microbial ecology, remains to be worked out.

Seeking "Gold Standard" Wastewater Treatments

The DOE JGI launched ETOP with the aim of bringing new technologies developed at other institutions in-house and making them available to its users for energy and environment applications. Among the projects approved in the inaugural ETOP call was one from metagenomics pioneer Jill Banfield at the University of California, Berkeley. Banfield proposed building and characterizing a pipeline that would allow researchers to isolate and study both near-complete and complete microbial genomes from environmental samples. One of the results from this project was published July 28, 2015, in *Environmental Microbiology*.

Natural ecosystems and the diverse communities of microbes (and other organisms) within them are extremely difficult to study because they are not closed systems, where all inputs and outputs can be accessed. To make the compositions and activities of microbial communities more tractable for analyses with genomic methods, this study focused on microbial communities in laboratory bioreactors that were being studied as a potential method of treating wastewater contaminated by gold ore processing. From these bioreactors, the team reconstructed draft and curated microbial genomes using high-throughput metagenomic sequencing of biofilm and supernatant samples. One bioreactor contained a mixture of cyanide (CN-) and thiocyanate (SCN-) being degraded, while in the other bioreactor, only thiocyanate was being degraded. Cyanide is used for processing gold ore while thiocyanate is a byproduct of the process.

"This is the first application of genome-resolved metagenomics to characterize SCN- and CN- bioreactors," the team noted, "revealing a complex community containing novel organisms and genes." The analyses allowed the team to outline the structures of the microbial communities and diagram potential nutrient flow paths. For example, they found evidence indicating that the microbes were not relying on the molasses included in the media as an energy source, which could help reduce bioreactor operating costs on the commercial scale. They also recovered several genome sequences allowing them to determine the composition of the communities in the bioreactors. They found some functions are shared, such as the ability to adapt to temperature and oxygen fluctuations. Others are less so; they found a complete denitrification pathway in one microbial species in the CN-SCN processing bioreactor.

Ultimately, this study will provide new approaches for the scientific community towards characterizing microbial communities involved in activities of major interest to DOE, including support of bioenergy feedstock plants, terrestrial carbon cycling and waste cleanup.

Liquid gold being poured into a cast at Gold Reef City in Johannesburg, South Africa. Processing gold ore involves the use of cyanide, resulting in contaminated wastewater that needs to be remediated. In this study, researchers conducted metagenomic analyses to identify microbes that could help with the bioremediation process. *(Credit: Dan Brown, Flickr, CC BY 2.0)*

Genepool Cluster
70+ million CPU-hours

Edison
1.6 million CPU-hours

JAMO
3.9 million files ecords

DnA File System
460 TB of data

Computational Infrastructure

The Department of Energy Joint Genome Institute's (DOE JGI) capacity as a next-generation genomics user facility has generated petabytes of data and analysis. In 2015 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.

The DOE JGI used several million central processing unit hours on the National Energy Research Scientific Computing Center's (NERSC) petascale supercomputers, Hopper and Edison. These calculations could not have been completed on the Genepool cluster. In 2015 NERSC deployed Cori Phase 1, a $21-million investment in data-intensive computing infrastructure. One of the core new features of Cori is new hardware called the Burst Buffer, an array of non-volatile random-access memory (NVRAM) nodes that will enable input/output-intensive workloads to run at scale on the new system. NERSC also deployed new software, called Shifter, to allow Docker containers to run on the Cray supercomputers. This work grew out of collaboration with the DOE JGI staff and has reduced the hurdles required to move bioinformatics, light source and astronomy workloads to move to Cori. The DOE JGI and NERSC will co-organize a workshop at the DOE JGI User Meeting in 2016. In this workshop, users can learn about the Docker technology and get hands-on experience running containers on the Cray systems as well as the cloud. In 2016 the DOE JGI user community will be able to apply for computational time with its Community Science Program applications. DOE JGI users with a compelling case for running analyses at NERSC with DOE JGI data or high-performance computing software will have access to high-performance computing resources.

The collaboration between Dan Rokhsar's (DOE JGI Chief Informatics Officer) and Katherine Yelick's (Berkeley Lab Associate Lab Director for Computing Sciences) research groups continued. The new version of Meraculous, called HipMer (High-performance Meraculous) allows for the assembly of the human genome in minutes across thousands of processors on Edison, NERSC's 2 petaflop supercomputer—capable of 2 quadrillion (thousand trillion) floating point operations per second. HipMer has the potential to speed up the assembly of metagenomes, making analysis of these datasets more tractable. The group made several algorithmic improvements to the software and is preparing to release an open-source version. This software was awarded the HPCwire Reader's Choice Award.

In 2015 NERSC moved to a new, state-of-the-art facility on the main Berkeley Lab site, and the computational infrastructure of the DOE JGI will make the move in 2016. The high-speed link between the two facilities enabled the transfer of the 3+ petabytes of DOE JGI data from Oakland to Berkeley without any disruption to the DOE JGI staff. 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 facility is anticipated to set a new standard in energy efficiency for high-performance computing.

Appendices

Appendix A
Acronyms at a Glance

AF	Alignment fraction
ANI	Average Nucleotide Identity
BER	Office of Biological and Environmental Research (at DOE)
BERAC	Biological and Environmental Research Advisory Committee
BESC	BioEnergy Science Center (at ORNL)
BLAST	Basic Local Alignment Search Tool
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
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
MetaBAT	Metagenome Binning with Abundance and Tetra-nucleotide frequencies
MGM	Microbial Genomics & Metagenomics
MiSI	Microbial Species Identifier
NERSC	National Energy Research Scientific Computing Center
NGEE	Next-Generation Ecosystem Experiments
NREL	National Renewable Energy Laboratory
ORNL	Oak Ridge National Laboratory
PMO	Project Management Office
PNNL	Pacific Northwest National Laboratory
ProDeGe	Protocol for Decontamination of Genomes
SAC	Scientific Advisory Committee
SFA	Scientific Focus Area
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 unwantedsubstances in waste and other substances.

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

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

Carbon cycle: The biogeochemical process by which carbon is exchanged among 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 contiguous sequence resulting from the assembly of smaller sequence fragments.

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

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

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

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

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

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

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

Flow cell: Resembles a microscopic slide, only with eight channels, on which DNA samples are loaded for analysis on 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 and is made of 5-carbon sugars. Unlike cellulose, which is crystalline, strong, and resistant to being broken down, hemicellulose is much more fragile, and has a random structure.

Informatics: The study of the science of information. At the DOE JGI, it is the science of managing and interpreting information with computational tools.

Library: A collection of DNA fragments that together represent all the DNA present in a particular 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 used to detect, identify and quantify molecules in simple and complex mixtures based on their mass and charge (m/z). It can be used to answer a wide range of biological questions.

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—derived from a microbial community—that is transcribed into RNA molecules and provides information on gene expression and gene function.

Microbe: Another name for a microorganism.

Microbiome: A community of microorganisms, or their collective genetic material.

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): Non-PCR based method of amplifying DNA, allowing amplification of genomes in single cells.

Nitrogen cycle: The biogeochemical process by which nitrogen is exchanged among 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 membrane-bound nucleus, mitochondria, and any other membrane-bound organelles.

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: Sequencing technique that, as used by Illumina systems, supports massively parallel sequencing through a proprietary method that detects single bases as they are incorporated into growing DNA strands.

Single-cell genomics: Method for the genomic analysis, including sequencing, of a single cell.

Single-molecule real-time (SMRT) sequencing: A parallelized single-molecule DNA sequencing method, that, as used by Pacific Biosciences systems, tracks real-time nucleotide incorporation events on a nanoscale space through a proprietary method.

Subcloning: The process of transferring a DNA fragment from one vector to another to study its functionality.

Sulfur cycle: The biogeochemical process by which sulfur is exchanged among 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 organism that serves as a snapshot of global gene expression.

Appendix C
2015 User Programs Supported Proposals

Community Science Program Proposals

INVESTIGATOR	AFFILIATION	DESCRIPTION
Catcheside, David	Flinders University (Australia)	Acquisition of the sequestrate (truffle like) habit by basidiomycete macrofungi
Cooper, Elizabeth	Clemson University	Comparative transcriptomics of sweet and grain sorghum to understand the mechanism and timing of sugar accumulation in an important bioenergy crop
de Vries, Ronald	CBS-KNAW Fungal Biodiversity Centre (Netherlands)	Dissecting the different approaches of ascomycete fungi to degrade plant biomass
DeLong, Ed	University of Hawaii at Manoa	Going long and going deep: Comprehensive open ocean community single cell genome sequencing at the model open ocean time series study site, station ALOHA
Dollhofer, Veronika	Bavarian State Research Center for Agriculture (Germany)	Anaerobic fungi and assessment of their potential for biogas production
Duplessis, Sebastien	INRA (France)	Sequencing a reference genome for *Phakopsora pachyrhizi*, the fungal pathogen responsible for the Asian Soybean Rust
Francis, Christopher	Stanford University	Metagenomic characterization of nitrogen-cycling microbial communities impacting uranium release in the Upper Colorado River Basin
Hamelin, Richard	University of British Columbia (Canada)	Pathobiome of bioenergy trees
Hibbett, David	Clark University	Comparative and functional genomics of shiitake mushrooms: an international collaboration to resolve evolutionary relationships, substrate specificity, growth profiles, and routes to domestication in the amphi-Pacific genus *Lentinula*
Juenger, Tom	University of Texas at Austin	Exploring natural genetic diversity in switchgrass (*Panicum virgatum*) and its microbiome

Community Science Program Proposals *(continued)*

INVESTIGATOR	AFFILIATION	DESCRIPTION
Kalyuzhnaya, Marina	University of Washington	Systems level insights into methane cycling in arid and semi-arid ecosystems via community metagenomics and metatranscriptomics
Lorito, Matteo	University of Naples (Italy)	Supporting the development of microbial probiotics for grasses useful in sustainable bioenergy production
Martin, Francis	INRA (France)	1KFG: Deep sequencing of ecologically-relevant Dikarya
Mayali, Xavier	Lawrence Livermore National Laboratory	Influence of phycosphere-associated bacteria on microalgal biofuel production
McMahon, Katherine	University of Wisconsin-Madison	Diel cycles of gene expression in oligotrophic, dystrophic, and eutrophic lakes to identify new gene functions and dissect carbon cycling metabolisms
Merchant, Sabeeha	University of California, Los Angeles	Comparative genomics and expression profiling of snow algae *Chlamydomonas cribrum* and *Chloromonas nivalis*
Niyogi, Kris	University of California, Berkeley	Functional genomics of photosynthesis in *Chlamydomonas*, JGI's flagship alga
O'Malley, Michelle	University of California, Santa Barbara	Genomic basis for syntrophic interactions between anaerobic gut fungi and methanogenic archaea
Pires, J. Chris	University of Missouri-Columbia	Investigating the diversity of mycorrhizal fungi to understand the evolution and function of symbiosis with orchids
Plett, Jonathan	University of Western Sydney (Australia)	Exploring the genomic basis for the global diversification by the ectomycorrhizal genus *Pisolithus*
Poland, Jesse	Kansas State University	The Intermediate Wheatgrass Genome: A resource for understanding mechanisms of perenniality and accelerating the development of perennial crops

Community Science Program Proposals *(continued)*

INVESTIGATOR	AFFILIATION	DESCRIPTION
Rappe, Michael	University of Hawaii at Manoa	Metagenomics of viral and microbial communities inhabiting warm, anoxic fluids of the sediment-buried deep ocean crust
Schachtman, Daniel	University of Nebraska	Systems analysis of the physiological and molecular mechanisms of Sorghum nitrogen use efficiency, water use efficiency and interactions with the soil microbiome
Spatafora, Joey	Oregon State University	Genomics of the early diverging lineages of fungi and their transition to terrestrial, plant-based ecologies
Umen, James	Donald Danforth Plant Science Center	Single cell and population dynamics of chromatin across the diurnal cycle in the model alga *Chlamydomonas*
Walsh, David	Concordia University (Canada)	Microbial metagenomics of carbon cycling communities in northern aquatic ecosystems
Wrighton, Kelly	Ohio State University	Life in the extreme deep terrestrial subsurface: microbial metabolism before and after shale gas extraction

Small-Scale Proposals

INVESTIGATOR	AFFILIATION	DESCRIPTION
Beman, J Michael	University of California, Merced	Metagenomics of methane production and oxidation in high altitude lakes of Yosemite National Park
Chistoserdova, Ludmila	University of Washington	Understanding methane cycling through manipulation of synthetic methane-oxidizing communities
Cullings, Ken	NASA Ames	Extreme endosymbiosis: An investigation into a unique fungal microbiome found in geothermal ecosystems in Yellowstone National Park and New Zealand
DeAngelis, Kristen	University of Massachusetts, Amherst	Expanding genomic diversity of terrestrial bacteria: linking genes to metabolism in the slower-growing members of forest soil bacterial communities
Kostka, Joel	Georgia Institute of Technology	The role of the Sphagnum microbiome in carbon and nutrient cycling in peatlands

Small-Scale Proposals *(continued)*

INVESTIGATOR	AFFILIATION	DESCRIPTION
McMahon, Katherine	University of Wisconsin-Madison	Reference genomes for abundant freshwater taxa – *Actinobacteria* and *Verrucomicrobia* phase 2
Meredith, Laura	Stanford University	Microbial, chemical, and physical drivers of COS fluxes and 18O-CO2 exchange rates in soils
Miller, Christopher	University of Colorado, Denver	Developing a systems-level understanding of biotic and abiotic controls on microbial methane cycling in freshwater wetlands
Redmond, Molly	University of North Carolina, Charlotte	Metagenomic sequencing of methane-oxidizing mesocosms from the Gulf of Mexico and Hudson Canyon
Rich, Jeremy	Brown University	Metagenomic sequencing of an uncultivated bacterial phylum in marine sediments amended with organic carbon and nitrate
Saito, Mak	Woods Hole Oceanographic Institution	Characterizing *Synechococcus* dominated populations from the Costa Rica dome and surrounding waters
Shade, Ashley	Michigan State University	Response and recovery of surface soil microbial communities to an ongoing underground coalmine fire
Simister, Rachel	University of British Columbia	Linking microbial genomic capacity to geochemical process in the deep terrestrial biosphere
Slonczewski, Joan	Kenyon College	Cyanobacterial communities of Antarctic Lake Fryxell liftoff mats and glacier meltwater
Tolar, Bradley	Stanford University	Monitoring the transcriptional response of a representative low-salinity ammonia-oxidizing thaumarchaeote to shifts in environmental conditions
Walsh, David	Concordia University	Metagenomics of western Arctic Ocean microbial communities
Wilkins, Michael	Ohio State University	Seasonal sulfur cycling as a control on methane flux in carbon-rich prairie pothole sediment ecosystems
Wrighton, Kelly	Ohio State University	Identifying key genomes and metabolisms responsible for near-surface methane cycling in freshwater wetlands

Synthesis Proposals

INVESTIGATOR	AFFILIATION	DESCRIPTION
Chang, Jui-Jen	Biodiversity Research Center, Academia Sinica	Designer Operons-A biomimic approach to the regulation of an enzyme cocktail for an artificial enzyme complex
Chen, Brandon	Genomatica, Inc.	Engineering efficient methanol utilization for renewable chemical
McCourt, Peter	University of Toronto	Exploring the perception landscape of the strigolactone receptor
Prather, Kristala	Massachusetts Institute of Technology	Combinatorial assembly, screening and functional characterization of a recombinant glucaric acid pathway in *S. cerevisiae*
Shen, Ben	The Scripps Research institute	Construction of bacterial artificial chromosome (BAC) vector libraries for comparative genomics, elucidation of gene functions and heterologous expression of targeted genes
Smanski, Mike	University of Minnesota	Towards a mechanistic understanding of disease suppressive soils: Refactoring natural product gene clusters
Wang, Clay	University of Southern California	Identification of fungal secondary metabolites with novel structures
Zimmer, Jochen	University of Virginia	Synthesis of cellulose synthase genes implicated in primary and secondary plant cell wall formation for structural and functional analyses of plant cellulose biosynthesis

JGI-EMSL Collaborative Science Initiative Proposals

The DOE JGI and the Environmental Molecular Sciences Laboratory (EMSL) accepted 8 proposals submitted during the 2015 call for Collaborative Science Initiative proposals. The call, now known as FICUS (Facilities Integrating Collaborations for User Science), 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, 2015, providing the researchers with access to the capabilities of both user facilities and datasets beyond what could be generated by either facility alone.

INVESTIGATOR	AFFILIATION	DESCRIPTION
Bell-Pedersen, Deborah	Texas A&M University	Specialized Ribosomes: A New Frontier in Gene Regulation
Cardon, Zoe	Marine Biological Laboratory	3D Reality Check: Developing Structural Support for Predicting Microbial Function and Interpreting Microbial "Omics" Data
Doty, Sharon	University of Washington	Nitrogen fixation in *Populus*: Identification and localization of the key diazotrophs in planta
Duhaime, Melissa	University of Michigan	Building the phage-host-environment interaction data to scale from genes-to-ecosystems: Towards predictive modeling of wild microbial and viral community dynamics
Eastwood, Dan	Swansea University (UK)	Genomes to dynamic decay communities: Understanding fungal interactions during early decomposition events in natural lignocellulosic substrate
Neumann, Rebecca	University of Washington	A Rhizosphere-Scale Investigation of the Relationship Between Plant Productivity and Methane Emissions from Wetlands
Orphan, Victoria	California Institute of Technology	Fluorescence-based cell sorting and targeted proteomic analysis of active methane-oxidizing syntrophic consortia from environmental samples
Pan, Chongle	Oak Ridge National Laboratory	Integrated Omics Analyses of a *Populus* Pedigree for Crop Improvement

Emerging Technologies Opportunity Program (ETOP)

The objectives of the ETOP are to identify and fund new and existing DOE JGI partners to develop promising new technical capabilities that could be provided to users. Successful pilot-scale proposals 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.

INVESTIGATOR	AFFILIATION	DESCRIPTION
2013		
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
2014		
Blainey, Paul	Broad Institute (AGI)	Microfluidic platform for high-sensitivity DNA library construction from microbial cells and gDNA
2015		
Fuchs, Bernhard	University of Washington	Accurate gene synthesis with tag-directed retrieval of sequence-verified DNA molecules
Pett-Ridge, Jennifer	Lawrence Livermore National Laboratory	SIP-Omics: Development of a semi-automated stable isotope probing pipeline for targeted metagenomics & metatranscriptomics
van den Engh, Ger	Center for Marine Cytometry	Development of flow sorting for microbial genome discovery

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

Mark Adams, *J. Craig Venter Institute (Chair)*

Carol Bult, *The Jackson Laboratory*

Steve Briggs, *University of California, San Diego*

Jeff Dangl, *University of North Carolina*

Claire M. Fraser, *University of Maryland*

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

Glenn Kubiak, *Lawrence Berkeley National Laboratory*

Trina McMahon, *University of Wisconsin-Madison*

Deirdre Meldrum, *Arizona State University*

Juan Meza, *University of California, Merced*

Mary Ann Moran, *University of Georgia*

Julian Parkhill, *The Sanger Institute*

Informatics Advisory Committee (IAC)

Members

Adam Arkin, *Lawrence Berkeley National Laboratory; University of California, Berkeley*

Judith Blake, *The Jackson Laboratory*

David Dooling, *Monsanto*

Paul Flicek, *European Molecular Biology Laboratory (EMBL)-European Bioinformatics Institute (EBI)*

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

Jill Mesirov, *Broad Institute (Chair)*

Granger Sutton, *J. Craig Venter Institute*

Cathy Wu, *Georgetown University*

Kathy Yelick, *Lawrence Berkeley National Laboratory*

Fungal Program User Advisory Committee

Members

Scott Baker, *Pacific Northwest National Laboratory*

Randy Berka, *Novozymes*

Ronald de Vries, *CBS (Netherlands)*

Audrey Gasch, *University of Wisconsin-Madison; Great Lakes Bioenergy Research Center*

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

Stephen Goodwin, *Purdue University*

David Hibbett, *Clark University*

Francis Martin, *INRA (France)*

Michelle O'Malley, *University of California, Santa Barbara*

Joseph Spatafora, *Oregon State University*

Kathleen Treseder, *University of California, Irvine*

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, *UMass Amherst*

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 at Manoa*

Jonathan Eisen, *University of California, Davis*

George Garrity, *Michigan State University (Chair)*

Steve Hallam, *University of British Columbia*

Phil Hugenholtz, *University of Queensland (Australia)*

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

Mary Ann Moran, *University of Georgia*

Nancy Moran, *University of Texas at Austin*

Victoria Orphan, *California Institute of Technology*

Rich Roberts, *NEB*

Ramunas Stepanauskas, *Bigelow Laboratory for Ocean Sciences*

Matt Sullivan, *The Ohio State University*

Synthetic Biology Advisory Committee

Members

Richard Bailey, *Independent consultant*

Doug Cameron, *Firstgreen Partners*

Sunil Chandran, *Amyris, Inc.*

James Flatt, *Synthetic Genomics*

Jay Keasling, *Lawrence Berkeley National Laboratory*

Megan Palmer, *University of California, Berkeley*

Elizabeth Sattely, *Stanford University*

Elizabeth Shank, *University of North Carolina, Chapel Hill*

David Weller, *USDA-ARS*

Appendix E
2015 Genomics of Energy and Environment Meeting

Keynote Speakers

To mark the 10th Annual Genomics of Energy & Environment Meeting, instead of a single opening keynote address, the DOE JGI invited representatives from the three Bioenergy Research Centers to give a series of short talks that highlighted their collaborations with the Institute.

Blake Simmons from the Joint BioEnergy Institute (JBEI) delivered the first of the opening keynotes. He reminded the audience about the need to move to sustainable, renewable fuels in the transportation sector, where fossil fuels currently provide 97 percent of the fuels.

Watch his talk at **http://bit.ly/JGI15UMSimmons**.

Shawn Kaeppler (foreground, with DOE JGI Director Eddy Rubin on the left) from Great Lakes Bioenergy Research Center then talked about his team's work on maize diversity, and their focus on the plant for its relationship to several candidate bioenergy grasses, including sorghum, switchgrass and miscanthus.

Watch his talk at **http://bit.ly/JGI15UMKaeppler**.

Jerry Tuskan from Oak Ridge National Laboratory (ORNL) and the Bioenergy Science Center (BESC) delivered the final opening keynote talk. He focused on work being done to understand the mutualistic relationship between the poplar tree and the fungus Laccaria bicolor, both of which were sequenced by the DOE JGI.

Watch his talk at **http://bit.ly/JGI15UMTuskan**.

User feedback session
From left to right: Susannah Tringe, Metagenome Program Lead; Igor Grigoriev, Fungal Program Lead; Tanja Woyke, Microbial Program Lead; Nikos Kyrpides, Prokaryote Super Program Lead; Jeremy Schmutz, Plant Program Lead; Eddy Rubin, Director; Kjiersten Fagnan, NERSC/JGI Engagement Lead.

Ed DeLong from the University of Hawaii at Manoa delivered the closing keynote. His talk focused on the application of omics techniques to discern the similarities between the daily rhythms of marine microbial communities located oceans apart. DeLong has collaborated with the DOE JGI for more than a decade, and his projects have focused on various marine microbes, ranging from deep-sea plankton to methane-oxidizing archaeon and Antarctic bacterioplankton.

Watch his talk at **http://bit.ly/JGI15UMDeLong**.

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

Jack Gilbert, *Argonne National Laboratory*

Francis Martin, *Institut National de la Recherche Agronomique (INRA; France)*

Joan Bennett, *Rutgers University*

Susanna Theroux, *DOE Joint Genome Institute*

Antonis Rokas, *Vanderbilt University*

Rotem Sorek, *Weizmann Institute of Science (Israel)*

Phil Hugenholtz, *University of Queensland (Australia)*

Barbara Jasny, *AAAS*

Atul Butte, *Stanford University*

Mark Burk, *Genomatica*

William Alexander, *University of Wisconsin-Madison*

Michelle C. Chang, *University of California, Berkeley*

Emiley Eloe-Fadrosh, *DOE Joint Genome Institute*

Marton Palatinszky, *University of Vienna (Austria)*

Thomas Brutnell, *Danforth Center*

Stephen Wright, *University of Toronto (Canada)*

Sue Rhee, *Carnegie Institution for Science*

Laura Bartley, *University of Oklahoma*

Sophien Kamoun, *The Sainsbury Laboratory (UK)*

Steve Briggs, *University of California, San Diego*

Tania González, *University of California, Berkeley*

Susan Lynch, *University of California, San Francisco*

Rick Cavicchioli, *University of New South Wales (Australia)*

Read about the meeting at
http://bit.ly/JGI15SpringPrimer

Learn more about the meeting talks at
http://usermeeting.jgi.doe.gov/past-meetings/2015-agenda/

Videos of the talks are available on DOE JGI's YouTube channel at
http://bit.ly/JGIUM2015videos

Appendix F
2015 Publications

October 2014 – September 2015

Abdul Rahman N et al. A molecular survey of Australian and North American termite genera indicates that vertical inheritance is the primary force shaping termite gut microbiomes. *Microbiome.* 2015 Feb 25;3:5. doi: 10.1186/s40168-015-0067-8.

Alawi M et al. Genome Sequence of *Methanosarcina soligelidi* SMA-21, Isolated from Siberian Permafrost-Affected Soil. *Genome Announc.* 2015 Apr 23;3(2). pii: e00270-15. doi: 10.1128/genomeA.00270-15.

Albertin CB et al. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature.* 2015 Aug 13;524(7564):220-4. doi: 10.1038/nature14668.

Ardley J et al. Genome sequence of the dark pink pigmented *Listia bainesii* microsymbiont Methylobacterium sp. WSM2598. *Stand Genomic Sci.* 2014 Dec 8. doi: 10.1186/1944-3277-9-5.

Axen SD et al. A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput Biol.* 2014 Oct 23;10(10):e1003898. doi: 10.1371/journal.pcbi.1003898.

Aylward FO et al. Convergent bacterial microbiotas in the fungal agricultural systems of insects. *MBio.* 2014 Nov 18;5(6):e02077. doi: 10.1128/mBio.02077-14.

Baker SE et al. Draft Genome Sequence of *Neurospora crassa* Strain FGSC 73. *Genome Announc.* 2015 Apr 2;3(2). pii: e00074-15. doi: 10.1128/genomeA.00074-15.

Baran R et al. Exometabolite niche partitioning among sympatric soil bacteria. *Nat Commun.* 2015 Sep 22;6:8289. doi: 10.1038/ncomms9289.

Bartholome J et al. High-resolution genetic maps of Eucalyptus improve *Eucalyptus grandis* genome assembly. *New Phytol.* 2015 Jun;206(4):1283-96. doi: 10.1111/nph.13150.

Beam JP et al. Ecophysiology of an uncultivated lineage of Aigarchaeota from an oxic, hot spring filamentous 'streamer' community. *ISME J.* 2016 Jan;10(1):210-24. doi: 10.1038/ismej.2015.83. Epub 2015 Jul 3.

Beck DA et al. Multiphyletic origins of methylotrophy in Alphaproteobacteria, exemplified by comparative genomics of Lake Washington isolates. *Environ Microbiol.* 2015 Mar;17(3):547-54. doi: 10.1111/1462-2920.12736.

Becker EA et al. Phylogenetically driven sequencing of extremely halophilic archaea reveals strategies for static and dynamic osmo-response. *PLoS Genet.* 2014 Nov 13;10(11):e1004784. doi: 10.1371/journal.pgen.1004784.

Berry D et al. Tracking heavy water (D2O) incorporation for identifying and sorting active microbial cells. *Proc Natl Acad Sci U S A.* 2015 Jan 13;112(2):E194-203. doi: 10.1073/pnas.1420406112.

Bianchetti CM et al. Active site and laminarin binding in glycoside hydrolase family 55. *J Biol Chem.* 2015 May 8;290(19):11819-32. doi: 10.1074/jbc.M114.623579.

Billis K et al. Comparative Transcriptomics between Synechococcus PCC 7942 and Synechocystis PCC 6803 Provide Insights into Mechanisms of Stress Acclimation. *PLoS One.* 2014 Oct 23;9(10):e109738. doi: 10.1371/journal.pone.0109738.

Blanc-Mathieu R et al. An improved genome of the model marine alga *Ostreococcus tauri* unfolds by assessing Illumina *de novo* assemblies. *BMC Genomics.* 2014 Dec 13;15:1103. doi: 10.1186/1471-2164-15-1103.

Boundy-Mills K et al. The United States Culture Collection Network (USCCN): Enhancing Microbial Genomics Research through Living Microbe Culture Collections. *Appl Environ Microbiol.* 2015 Sep 1;81(17):5671-4. doi: 10.1128/AEM.01176-15.

Branco S et al. Genetic isolation between two recently diverged populations of a symbiotic fungus. *Mol Ecol.* 2015 Jun;24(11):2747-58. doi: 10.1111/mec.13132.

Bräuer SL et al. Genome of *Methanoregula boonei* 6A8 reveals adaptations to oligotrophic peatland environments. *Microbiology.* 2015 Aug;161(8):1572-81. doi: 10.1099/mic.0.000117.

Brown CT et al. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature.* 2015 Jul 9;523(7559):208-11. doi: 10.1038/nature14486.

Brutnell TP et al. *Brachypodium distachyon* and *Setaria viridis*: Model Genetic Systems for the Grasses. *Annu Rev Plant Biol.* 2015 Apr;66:465-85. doi: 10.1146/annurev-arplant-042811-105528.

Cantor M et al. Elviz - exploration of metagenome assemblies with an interactive visualization tool. *BMC Bioinformatics.* 2015 Apr 28;16:130. doi: 10.1186/s12859-015-0566-4.

Carr SA et al. Abundant Atribacteria in deep marine sediment from the Adélie Basin, Antarctica. *Front Microbiol.* 2015 Aug 26;6:872. doi: 10.3389/fmicb.2015.00872

Castelle CJ et al. Genomic Expansion of Domain Archaea Highlights Roles for Organisms from New Phyla in Anaerobic Carbon Cycling. *Curr Biol.* 2015 Mar 16;25(6):690-701. doi: 10.1016/j.cub.2015.01.014.

Chaib De Mares M et al. Horizontal transfer of carbohydrate metabolism genes into ectomycorrhizal *Amanita*. *New Phytol.* 2015 Mar;205(4):1552-64. doi: 10.1111/nph.13140.

Chang Y et al. Phylogenomic Analyses Indicate that Early Fungi Evolved Digesting Cell Walls of Algal Ancestors of Land Plants. *Genome Biol Evol.* 2015 May 14;7(6):1590-601. doi: 10.1093/gbe/evv090.

Chapman JA et al. A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biol.* 2015 Jan 31;16(1):26.

Checcucci A et al. The integrated microbial genome resource of analysis. *Methods Mol Biol.* 2015;1231:289-95. doi: 10.1007/978-1-4939-1720-4_18.

Chochois V et al. Variation in Adult Plant Phenotypes and Partitioning among Seed and Stem-Borne Roots across *Brachypodium distachyon* Accessions to Exploit in Breeding Cereals for Well-Watered and Drought Environments. *Plant Physiol.* 2015 Jul;168(3):953-67. doi: 10.1104/pp.15.00095.

Christen M et al. Genome Calligrapher: A Web Tool for Refactoring Bacterial Genome Sequences for *de Novo* DNA Synthesis. *ACS Synth Biol.* 2015 Aug 21;4(8):927-34. doi: 10.1021/acssynbio.5b00087.

Christensen GA et al. Rex (encoded by DVU_0916) in *Desulfovibrio vulgaris* Hildenborough is a repressor of sulfate adenylyl transferase and is regulated by NADH. *J Bacteriol.* 2015 Jan 1;197(1):29-39. doi: 10.1128/JB.02083-14.

Chubukov V et al. Acute Limonene Toxicity in *Escherichia coli* Is Caused by Limonene Hydroperoxide and Alleviated by a Point Mutation in Alkyl Hydroperoxidase AhpC. *Appl Environ Microbiol.* 2015 Jul 15;81(14):4690-6. doi: 10.1128/AEM.01102-15.

Clingenpeel S et al. Reconstructing each cell's genome within complex microbial communities-dream or reality? *Front Microbiol.* 2015 Jan 8;5:771. doi: 10.3389/fmicb.2014.00771.

Cohen MF et al. Genome Sequence of the Alkaline-Tolerant Cellulomonas sp. Strain FA1. *Genome Announc.* 2015 Jun 18;3(3). pii: e00646-15. doi: 10.1128/genomeA.00646-15.

De Meyer SE et al. High-quality permanent draft genome sequence of the Lebeckia - nodulating *Burkholderia dilworthii* strain WSM3556(T). *Stand Genomic Sci.* 2015 Sep 19;10:64. doi: 10.1186/s40793-015-0048-3.

De Meyer SE et al. High-quality permanent draft genome sequence of the *Parapiptadenia rigida*-nodulating Burkholderia sp. strain UYPR1.413. *Stand Genomic Sci.* 2015 Jun 4;10:31. doi:10.1186/s40793-015-0018-9.

De Meyer SE et al. High-quality permanent draft genome sequence of the *Parapiptadenia rigida*-nodulating Cupriavidus sp. strain UYPR2.512. *Stand Genomic Sci.* 2015 Apr 14;10:13. doi:10.1186/1944-3277-10-13.

Dedysh SN et al. Draft Genome Sequence of *Methyloferula stellata* AR4, an Obligate Methanotroph Possessing Only a Soluble Methane Monooxygenase. *Genome Announc.* 2015 Mar 5;3(2). pii: e01555-14. doi: 10.1128/genomeA.01555-14.

Dhillon B et al. Horizontal gene transfer and gene dosage drives adaptation to wood colonization in a tree pathogen. *Proc Natl Acad Sci U S A.* 2015 Mar 17;112(11):3451-3456.

Dmytrenko O et al. The genome of the intracellular bacterium of the coastal bivalve, *Solemya velum*: a blueprint for thriving in and out of symbiosis. *BMC Genomics.* 2014 Oct 23;15:924. doi: 10.1186/1471-2164-15-924.

Doré J et al. Comparative genomics, proteomics and transcriptomics give new insight into the exoproteome of the basidiomycete *Hebeloma cylindrosporum* and its involvement in ectomycorrhizal symbiosis. *New Phytol.* 2015 Dec;208(4):1169-87. doi: 10.1111/nph.13546. Epub 2015 Jul 14.

Dubchak I et al. An integrative computational approach for prioritization of genomic variants. *PLoS One.* 2014 Dec 15;9(12):e114903. doi: 10.1371/journal.pone.0114903. eCollection 2014.

Earl D et al. Alignathon: a competitive assessment of whole-genome alignment methods. *Genome Res.* 2014 Dec;24(12):2077-89. doi: 10.1101/gr.174920.114.

Eichorst SA et al. Advancements in the application of NanoSIMS and Raman microspectroscopy to investigate the activity of microbial cells in soils. *FEMS Microbiol Ecol.* 2015 Oct;91(10). pii: fiv106. doi: 10.1093/femsec/fiv106.

Fiedler JD et al. High-Density Single Nucleotide Polymorphism Linkage Maps of Lowland Switchgrass using Genotyping-by-Sequencing. *Plant Gen.* 2015 Jul 10;8(2). doi:10.3835/plantgenome2014.10.0065

Field EK et al. Genomic insights into the uncultivated marine Zetaproteobacteria at Loihi Seamount. *ISME J.* 2015 Mar 17;9:857-70. doi: 10.1038/ismej.2014.183.

Firrincieli A et al. Genome sequence of the plant growth promoting endophytic yeast *Rhodotorula graminis* WP1. *Front Microbiol.* 2015 Sep 17;6:978. doi: 10.3389/fmicb.2015.00978.

Fitzgerald TL et al. Brachypodium as an emerging model for cereal-pathogen interactions. *Ann Bot.* 2015 Apr;115(5):717-31. doi: 10.1093/aob/mcv010.

Floudas D et al. Evolution of novel wood decay mechanisms in Agaricales revealed by the genome sequences of *Fistulina hepatica* and *Cylindrobasidium torrendii*. *Fungal Genet Biol.* 2015 Mar;76:78-92. doi: 10.1016/j.fgb.2015.02.002.

Gao XY et al. Draft genome sequence of *Halomonas lutea* strain YIM 91125(T) (DSM 23508(T)) isolated from the alkaline Lake Ebinur in Northwest China. *Stand Genomic Sci.* 2015 Jan 20;10:1. doi: 10.1186/1944-3277-10-1.

Garau G et al. Genome sequence of Ensifer medicae Di28; an effective N2-fixing microsymbiont of *Medicago murex* and *M. polymorpha*. *Stand Genomic Sci.* 2014 Dec 8. doi: 10.1186/1944-3277-9-4

Gaudana SB et al. Bioinformatic analysis of the distribution of inorganic carbon transporters and prospective targets for bioengineering to increase C uptake by cyanobacteria. *Photosynth Res.* 2014 Nov 16. doi: 10.1007/s11120-014-0059-8

Gilbert D et al. The Joint Genome Institute Offers Resources Beyond a Core Facility. *Microbe.* 2015 Jul;10(7):289-93. http://www.asmscience.org/content/journal/microbe/10.1128/microbe.10.289.1

Goordial J et al. Improved-high-quality draft genome sequence of Rhodococcus sp. JG-3, a eurypsychrophilic Actinobacteria from Antarctic Dry Valley permafrost. *Stand Genomic Sci.* 2015 Sep 3;10:61. doi: 10.1186/s40793-015-0043-8.

Gordon SP et al. Widespread Polycistronic Transcripts in Fungi Revealed by Single-Molecule mRNA Sequencing. *PLoS One.* 2015 Jul 15;10(7):e0132628. doi: 10.1371/journal.pone.0132628.

Gutierrez T et al. Genome Sequence of Halomonas sp. Strain MCTG39a, a Hydrocarbon-Degrading and Exopolymeric Substance-Producing Bacterium. *Genome Announc.* 2015 Jul 16;3(4). pii: e00793-15. doi: 10.1128/genomeA.00793-15.

Gutierrez T et al. Genome Sequence of *Polycyclovorans algicola* Strain TG408, an Obligate Polycyclic Aromatic Hydrocarbon-Degrading Bacterium Associated with Marine Eukaryotic Phytoplankton. *Genome Announc.* 2015 Mar 26;3(2). pii: e00207-15. doi: 10.1128/genomeA.00207-15.

Gutierrez T et al. Genome Sequence of *Porticoccus hydrocarbonoclasticus* Strain MCTG13d, an Obligate Polycyclic Aromatic Hydrocarbon-Degrading Bacterium Associated with Marine Eukaryotic Phytoplankton. *Genome Announc.* 2015 Jun 18;3(3). pii: e00672-15. doi: 10.1128/genomeA.00672-15.

Haase NJ et al. Shared Genomic Regions Between Derivatives of a Large Segregating Population of Maize Identified Using Bulked Segregant Analysis Sequencing and Traditional Linkage Analysis. *G3* (Bethesda). 2015 Jun 1. pii: g3.115.017665. doi: 10.1534/g3.115.017665.

Hadjithomas M et al. IMG-ABC: A Knowledge Base To Fuel Discovery of Biosynthetic Gene Clusters and Novel Secondary Metabolites. *MBio.* 2015 Jul 14;6(4). pii: e00932-15. doi: 10.1128/mBio.00932-15.

Hahnke RL et al. High quality draft genome sequence of *Flavobacterium rivuli* type strain WB 3.3-2(T) (DSM 21788(T)), a valuable source of polysaccharide decomposing enzymes. *Stand Genomic Sci.* 2015 Jul 30;10:46. doi: 10.1186/s40793-015-0032-y. eCollection 2015.

Hamilton R et al. Draft genomes of gammaproteobacterial methanotrophs isolated from terrestrial ecosystems. *Genome Announc.* 2015 Jun 4;3(3). pii: e00515-15. doi: 10.1128/genomeA.00515-15.

Haushalter RW et al. Development of an orthogonal fatty acid biosynthesis system in *E. coli* for oleochemical production. *Metab Eng.* 2015 Jul;30:1-6. doi: 10.1016/j.ymben.2015.04.003.

He S et al. Patterns in wetland microbial community composition and functional gene repertoire associated with methane emissions. *MBio.* 2015 May 19;6(3):e00066-15. doi: 10.1128/mBio.00066-15.

Hedlund BP et al. Uncultivated thermophiles: current status and spotlight on 'Aigarchaeota'. *Curr Opin Microbiol.* 2015 Jun;25:136-45. doi: 10.1016/j.mib.2015.06.008.

Hedrick PW et al. Examining the cause of high inbreeding depression: analysis of whole-genome sequence data in 28 selfed progeny of *Eucalyptus grandis*. *New Phytol.* 2015 Sep 10. doi: 10.1111/nph.13639. [Epub ahead of print]

Hori C et al. Analysis of the *Phlebiopsis gigantea* genome, transcriptome and secretome provides insight into its pioneer colonization strategies of wood. *PLoS Genet.* 2014 Dec 4;10(12):e1004759. doi: 10.1371/journal.pgen.1004759.

Hudson CM et al. Lignin-Modifying Processes in the Rhizosphere of Aridland Grasses. *Environ Microbiol.* 2015 Aug 17. doi: 10.1111/1462-2920.13020. [Epub ahead of print]

Hug LA et al. Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. *Environ Microbiol.* 2015 Jun 1. doi: 10.1111/1462-2920.12930. [Epub ahead of print]

Hug LA et al. Aquifer environment selects for microbial species cohorts in sediment and groundwater. *ISME J.* 2015 Aug;9(8):1846-56. doi: 10.1038/ismej.2015.2.

Hultman J et al. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature.* 2015 May 14;521(7551):208-12. doi: 10.1038/nature14238.

Hurley JM et al. Analysis of clock-regulated genes in Neurospora reveals widespread posttranscriptional control of metabolic potential. *Proc Natl Acad Sci* USA. 2014 Dec 2;111(48):16995-7002. doi: 10.1073/pnas.1418963111.

Hwang C et al. Complete Genome Sequence of Anaeromyxobacter sp. Fw109-5, an Anaerobic, Metal-Reducing Bacterium Isolated from a Contaminated Subsurface Environment. *Genome Announc.* 2015 Jan 22;3(1). pii: e01449-14. doi: 10.1128/genomeA.01449-14.

International Cassava Genetic Map Consortium (ICGMC). High-resolution linkage map and chromosome-scale genome assembly for cassava (*Manihot esculenta* Crantz) from 10 populations. *G3* (Bethesda). 2015 Jan 1;5(1):133-44. doi: 10.1534/g3.114.015008.

Kalyuzhnaya MG et al. Draft genome sequences of gammaproteobacterial methanotrophs isolated from lake washington sediment. *Genome Announc.* 2015 Mar 12;3(2). pii: e00103-15. doi: 10.1128/genomeA.00103-15.

Kang DD et al. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ.* 2015 Aug 27;3:e1165. doi:10.7717/peerj.1165

Kantor RS et al. Bioreactor microbial ecosystems for thiocyanate and cyanide degradation unravelled with genome-resolved metagenomics. *Environ Microbiol.* 2015 Jun 1. doi: 10.1111/1462-2920.12936. [Epub ahead of print]

Kelly S et al. Genome sequence of the Lotus spp. microsymbiont *Mesorhizobium loti* strain NZP2037. *Stand Genomic Sci.* 2014 Dec 8. doi: 10.1186/1944-3277-9-7

Kelly S et al. Genome sequence of the Lotus spp. microsymbiont *Mesorhizobium loti* strain R7A. *Stand Genomic Sci.* 2014 Dec 8. doi: 10.1186/1944-3277-9-6

Kirby J et al. Enhancing Terpene yield from sugars via novel routes to 1-deoxy-d-xylulose 5-phosphate. *Appl Environ Microbiol.* 2015 Jan;81(1):130-8. doi: 10.1128/AEM.02920-14.

Kohler A et al. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat Genet.* 2015 Apr;47(4):410-5. doi: 10.1038/ng.3223.

Kotak M et al. Complete Genome Sequence of the Opitutaceae Bacterium Strain TAV5, a Potential Facultative Methylotroph of the Wood-Feeding Termite R*eticulitermes flavipes*. *Genome Announc.* 2015 Mar 5;3(2). pii: e00060-15. doi: 10.1128/genomeA.00060-15.

Kourist R et al. Genomics and Transcriptomics Analyses of the Oil-Accumulating Basidiomycete Yeast *Trichosporon oleaginosus*: Insights into Substrate Utilization and Alternative Evolutionary Trajectories of Fungal Mating Systems. *MBio.* 2015 Jul 21;6(4). pii: e00918-15. doi: 10.1128/mBio.00918-15.

Kruse T et al. Genomic, Proteomic, and Biochemical Analysis of the Organohalide Respiratory Pathway in *Desulfitobacterium dehalogenans*. *J Bacteriol.* 2015 Mar 1;197(5):893-904. doi: 10.1128/JB.02370-14.

Kuo A et al. Expanding genomics of mycorrhizal symbiosis. *Front Microbiol.* 2014 Nov 4;5:582. doi: 10.3389/fmicb.2014.00582.

Labonté JM et al. Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *ISME J.* 2015 Nov;9(11):2386-99. doi: 10.1038/ismej.2015.48. Epub 2015 Apr 7.

Lam KK et al. FinisherSC: a repeat-aware tool for upgrading *de novo* assembly using long reads. *Bioinformatics.* 2015 Oct 1;31(19):3207-9. doi: 10.1093/bioinformatics/btv280. Epub 2015 Jun 3.

Lao J et al. Proteome profile of the endomembrane of developing coleoptiles from switchgrass (*Panicum virgatum*). *Proteomics.* 2015 Jul;15(13):2286-90. doi: 10.1002/pmic.201400487.

Lau SC et al. Genome sequence of the pink-pigmented marine bacterium *Loktanella hongkongensis* type strain (UST950701-009P(T)), a representative of the Roseobacter group. *Stand Genomic Sci.* 2015 Aug 11;10:51. doi: 10.1186/s40793-015-0050-9.

Laviad S et al. High quality draft genome sequence of *Leucobacter chironomi* strain MM2LBT (DSM 19883T) isolated from a *Chironomus* sp. egg mass. *Stand Genomic Sci.* 2015 May 8;10:21. doi:10.1186/s40793-015-0003-3.

Laviad S et al. High quality draft genome sequence of *Brachymonas chironomi* AIMA4T (DSM 19884T) isolated from a *Chironomus* sp. egg mass. *Stand Genomic Sci.* 2015 May 27;10:29. doi:10.1186/s40793-015-0010-4.

Lebeis SL et al. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science.* 2015 Aug 21;349(6250):860-4. doi: 10.1126/science.aaa8764.

Lee LL et al. Complete Genome Sequences of Caldicellulosiruptor sp. Strain Rt8.B8, Caldicellulosiruptor sp. Strain Wai35.B1, and "Thermoanaerobacter cellulolyticus". *Genome Announc.* 2015 May 14;3(3). pii: e00440-15. doi: 10.1128/genomeA.00440-15.

Lichius A et al. Genome sequencing of the *Trichoderma reesei* QM9136 mutant identifies a truncation of the transcriptional regulator XYR1 as the cause for its cellulase-negative phenotype. *BMC Genomics.* 2015 Apr 20;16:326. doi: 10.1186/s12864-015-1526-0.

Lim YW et al. Purifying the impure: sequencing metagenomes and metatranscriptomes from complex animal-associated samples. *J Vis Exp.* 2014 Dec 22;(94). doi: 10.3791/52117.

Lowe CJ et al. The deuterostome context of chordate origins. *Nature.* 2015 Apr 23;520(7548):456-65. doi: 10.1038/nature14434.

Lozano R et al. Identification and distribution of the NBS-LRR gene family in the Cassava genome. *BMC Genomics.* 2015 May 7;16:360. doi: 10.1186/s12864-015-1554-9.

Luef B et al. Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat Commun.* 2015 Feb 27;6:6372. doi: 10.1038/ncomms7372.

Matsuda Y et al. A New Nomenclature of *Xenopus laevis* Chromosomes Based on the Phylogenetic Relationship to *Silurana/Xenopus tropicalis*. *Cytogenet Genome Res.* 2015 Aug;145:187-191.

Mazur A et al. High-quality permanent draft genome sequence of *Rhizobium leguminosarum* bv. viciae strain GB30; an effective microsymbiont of *Pisum sativum* growing in Poland. *Stand Genomic Sci.* 2015 Jul 16;10:36. doi: 10.1186/s40793-015-0029-6.

McCarthy S et al. Complete Genome Sequence of *Sulfolobus solfataricus* Strain 98/2 and Evolved Derivatives. *Genome Announc.* 2015 May 28;3(3). pii: e00549-15. doi: 10.1128/genomeA.00549-15.

McIlroy SJ et al. High quality draft genome sequence of *Meganema perideroedes* str. Gr1T and a proposal for its reclassification to the family Meganemaceae fam. nov. *Stand Genomic Sci.* 2015 Feb 27;10:23. doi:10.1186/s40793-015-0013-1.

McTaggart TL et al. Draft genomes of two strains of flavobacterium isolated from lake washington sediment. *Genome Announc.* 2015 Feb 19;3(1). pii: e01597-14. doi: 10.1128/genomeA.01597-14.

McTaggart TL et al. Draft Genome of Pseudomonas sp. Strain 11/12A, Isolated from Lake Washington Sediment. *Genome Announc.* 2015 Feb 19;3(1). pii: e01587-14. doi: 10.1128/genomeA.01587-14.

McTaggart TL et al. Draft Genome of Janthinobacterium sp. RA13 Isolated from Lake Washington Sediment. *Genome Announc.* 2015 Feb 12;3(1). pii: e01588-14. doi: 10.1128/genomeA.01588-14.

McTaggart TL et al. Draft genome sequences of five new strains of methylophilaceae isolated from lake washington sediment. *Genome Announc.* 2015 Feb 5;3(1). pii: e01511-14. doi: 10.1128/genomeA.01511-14.

Medema MH et al. Minimum Information about a Biosynthetic Gene cluster. *Nat Chem Biol.* 2015 Aug 18;11(9):625-631. doi: 10.1038/nchembio.1890.

Meier-Kolthoff JP et al. Complete genome sequence of DSM 30083T, the type strain (U5/41T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. *Stand Genomic Sci.* 2014 Dec 8. doi:10.1186/1944-3277-9-2

Mhamdi R et al. High-quality permanent draft genome sequence of *Ensifer meliloti* strain 4H41, an effective salt- and drought-tolerant microsymbiont of *Phaseolus vulgaris*. *Stand Genomic Sci.* 2015 Jul 2;10:34. doi:10.1186/s40793-015-0005-1.

Min B et al. Genome sequence of a white rot fungus *Schizopora paradoxa* KUC8140 for wood decay and mycoremediation. *J Biotechnol.* 2015 Oct 10;211:42-3. doi: 10.1016/j.jbiotec.2015.06.426. Epub 2015 Jul 15.

Muchero W et al. High-resolution genetic mapping of allelic variants associated with cell wall chemistry in *Populus*. *BMC Genomics.* 2015 Jan 23;16(1):24.

Mukherjee S et al. Large-scale contamination of microbial isolate genomes by Illumina PhiX control. *Stand Genomic Sci.* 2015 Apr 17;10:18. doi:10.1186/1944-3277-10-18.

Mukherjee S et al. High quality draft genome sequence and analysis of *Pontibacter roseus* type strain SRC-1T (DSM 17521T) isolated from muddy waters of a drainage system in Chandigarh, India. *Stand Genomic Sci.* 2015 Feb 9;10:8. doi: 10.1186/1944-3277-10-8.

Muñoz-Amatriaín M et al. Sequencing of 15,622 gene-bearing BACs clarifies the gene-dense regions of the barley genome. *Plant J.* 2015 Oct;84(1):216-227. doi: 10.1111/tpj.12959. Epub 2015 Sep 21.

Ngan CY et al. Lineage-specific chromatin signatures reveal a regulator of lipid metabolism in microalgae. *Nat Plants.* 2015 Aug;1:15107. doi:10.1038/nplants.2015.107

Nobu MK et al. Phylogeny and physiology of candidate phylum 'Atribacteria' (OP9/JS1) inferred from cultivation-independent genomics. *ISME J.* 2015 Jun 19. doi: 10.1038/ismej.2015.97. [Epub ahead of print]

Nobu MK et al. Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor. *ISME J.* 2015 Aug;9(8):1710-22. doi: 10.1038/ismej.2014.256.

Ntougias S et al. High-quality permanent draft genome sequence of the extremely osmotolerant diphenol degrading bacterium *Halotalea alkalilenta* AW-7(T), and emended description of the genus Halotalea. *Stand Genomic Sci.* 2015 Aug 13;10:52. doi: 10.1186/s40793-015-0052-7.

O'Connor RM et al. Gill bacteria enable a novel digestive strategy in a wood-feeding mollusk. *Proc Natl Acad Sci* USA. 2014 Nov 25;111(47):E5096-104. doi: 10.1073/pnas.1413110111.

Olsen MT et al. The molecular dimension of microbial species: 3. Comparative genomics of Synechococcus strains with different light responses and *in situ* diel transcription patterns of associated putative ecotypes in the Mushroom Spring microbial mat. *Front Microbiol.* 2015 Jun 23;6:604. doi: 10.3389/fmicb.2015.00604.

Paul BG et al. Targeted diversity generation by intraterrestrial archaea and archaeal viruses. *Nat Commun.* 2015 Mar 23;6:6585. doi: 10.1038/ncomms7585.

Petit E et al. Genome and Transcriptome of *Clostridium phytofermentans*, Catalyst for the Direct Conversion of Plant Feedstocks to Fuels. *PLoS One.* 2015 Jun 2;10(6):e0118285. doi: 10.1371/journal.pone.0118285.

Piao H et al. Insights into the bacterial community and its temporal succession during the fermentation of wine grapes. *Front Microbiol.* 2015 Aug 18;6:809. doi: 10.3389/fmicb.2015.00809.

Plett JM et al. The Mutualist *Laccaria bicolor* Expresses a Core Gene Regulon During the Colonization of Diverse Host Plants and a Variable Regulon to Counteract Host-Specific Defenses. *Mol Plant Microbe Interact.* 2015 Mar;28(3):261-73. doi: 10.1094/MPMI-05-14-0129-FI.

Ramsay BD et al. High-Quality Draft Genome Sequence of *Desulfovibrio carbinoliphilus* FW-101-2B, an Organic Acid-Oxidizing Sulfate-Reducing Bacterium Isolated from Uranium(VI)-Contaminated Groundwater. *Genome Announc.* 2015 Mar 12;3(2). pii: e00092-15. doi: 10.1128/genomeA.00092-15.

Ray J et al. Complete Genome Sequence of *Cupriavidus basilensis* 4G11, Isolated from the Oak Ridge Field Research Center Site. *Genome Announc.* 2015 May 14;3(3). pii: e00322-15. doi: 10.1128/genomeA.00322-15.

Reddy TB et al. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Res.* 2015 Jan;43(Database issue):D1099-106. doi: 10.1093/nar/gku950.

Reeve W et al. High-Quality draft genome sequence of the Lotus spp. microsymbiont *Mesorhizobium loti* strain CJ3Sym. *Stand Genomic Sci.* 2015 Aug 14;10:54. doi:10.1186/s40793-015-0049-2 .

Reeve W et al. A Genomic Encyclopedia of the Root Nodule Bacteria: assessing genetic diversity through a systematic biogeographic survey. *Stand Genomic Sci.* 2015 Feb 9;10:14. doi: 10.1186/1944-3277-10-14.

Reeve W et al. Genome sequence of the *Lotus corniculatus* microsymbiont Mesorhizobium loti strain R88B. *Stand Genomic Sci.* 2014 Dec 8. doi: 10.1186/1944-3277-9-3

Rellán-Álvarez R et al. GLO-Roots: an imaging platform enabling multidimensional characterization of soil-grown root systems. *eLife.* 2015 Aug 19;10.7554/eLife.07597

Rooney EA et al. Draft Genome Sequence of the Cellulolytic and Xylanolytic Thermophile *Clostridium clariflavum* Strain 4-2a. *Genome Announc.* 2015 Jul 23;3(4). pii: e00797-15. doi: 10.1128/genomeA.00797-15.

Roux S et al. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. *eLife.* 2015 Jul 22;4. doi: 10.7554/eLife.08490.

Sakamoto M et al. High quality draft genome sequence of *Bacteroides barnesiae* type strain BL2(T) (DSM 18169(T)) from chicken caecum. *Stand Genomic Sci.* 2015 Aug 2;10:48. doi: 10.1186/s40793-015-0045-6. eCollection 2015.

Salazar G et al. Global diversity and biogeography of deep-sea pelagic prokaryotes. *ISME J.* 2015 Aug 7. doi: 10.1038/ismej.2015.137. [Epub ahead of print]

Satoh N et al. Chordate evolution and the three-phylum system. *Proc Biol Sci.* 2014 Nov 7;281(1794): 20141729. doi: 10.1098/rspb.2014.1729.

Scheuner C et al. Complete genome sequence of *Planctomyces brasiliensis* type strain (DSM 5305T), phylogenomic analysis and reclassification of Planctomycetes including the descriptions of Gimesia gen. nov., Planctopirus gen. nov. and Rubinisphaera gen. nov. and emended descriptions of the order Planctomycetales and the family Planctomycetaceae. *Stand Genomic Sci.* 2014 Dec 8. doi: 10.1186/1944-3277-9-10

Schicklberger M et al. Draft Genome Sequence of *Raoultella terrigena* R1Gly, a Diazotrophic Endophyte. *Genome Announc.* 2015 Jun 11;3(3). pii: e00607-15. doi: 10.1128/genomeA.00607-15.

Scholz M et al. Improved assemblies using a source-agnostic pipeline for MetaGenomic Assembly by Merging (MeGAMerge) of contigs. *Sci Rep.* 2014 Oct 1;4:6480. doi: 10.1038/srep06480.

Schwessinger B et al. Transgenic Expression of the Dicotyledonous Pattern Recognition Receptor EFR in Rice Leads to Ligand-Dependent Activation of Defense Responses. *PLoS Pathog.* 2015 Mar 30;11(3):e1004809. doi: 10.1371/journal.ppat.1004809. eCollection 2015 Mar.

Septer AN et al. Bright luminescence of *Vibrio fischeri* aconitase mutants reveals a connection between citrate and the Gac/Csr regulatory system. *Mol Microbiol.* 2015 Jan;95(2):283-96. doi: 10.1111/mmi.12864.

Ševčíková T et al. Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Sci Rep.* 2015 May 28;5:10134. doi: 10.1038/srep10134.

Sharma MK et al. Targeted Switchgrass BAC Library Screening and Sequence Analysis Identifies Predicted Biomass and Stress Response-Related Genes. *Bioenerg. Res.* 2015 Aug 30:1-14. doi:10.1007/s12155-015-9667-1

Sharon I et al. Accurate, multi-kb reads resolve complex populations and detect rare microorganisms. *Genome Res.* 2015 Apr;25(4):534-43. doi: 10.1101/gr.183012.114.

Sharp CE et al. Draft Genome Sequence of the Moderately Halophilic Methanotroph *Methylohalobius crimeensis* Strain 10Ki. *Genome Announc.* 2015 Jun 11;3(3). pii: e00644-15. doi: 10.1128/genomeA.00644-15.

Shetty A et al. Complete genome sequence of the phenanthrene-degrading soil bacterium *Delftia acidovorans* Cs1-4. *Stand Genomic Sci.* 2015 Aug 15;10:55. doi:10.1186/s40793-015-0041-x. eCollection 2015.

Siegel JB et al. Computational protein design enables a novel one-carbon assimilation pathway. *Proc Natl Acad Sci* USA. 2015 Mar 24;112(12):3704-9. doi: 10.1073/pnas.1500545112.

Silva LP et al. Exometabolomics and MSI: deconstructing how cells interact to transform their small molecule environment. *Curr Opin Biotechnol.* 2015 Aug;34:209-16. doi: 10.1016/j.copbio.2015.03.015.

Simirenko L et al. The Joint Genome Institute's synthetic biology internal review process. *Journal of Responsible Innovation.* 2015 Jan 27;2(1):133-136. doi: 10.1080/23299460.2014.1002058

Simmons CW et al. Metatranscriptomic analysis of lignocellulolytic microbial communities involved in high-solids decomposition of rice straw. *Biotechnol Biofuels.* 2014 Dec 31;7(1):495. doi: 10.1186/s13068-014-0180-0. eCollection 2014.

Singan V. Precision Drug Targeting: Understanding the Intracellular Transport Mechanism. *J Investig Genomics.* 2014 Nov 4;2(1):00011. doi: 10.15406/jig.2015.02.00011

Singh R et al. Complete Genome Sequence of an Evolved *Thermotoga maritima* Isolate. *Genome Announc.* 2015 May 28;3(3). pii: e00557-15. doi: 10.1128/genomeA.00557-15.

Smalley NE et al. Functional and genomic diversity of methylotrophic Rhodocyclaceae: description of the new species *Methyloversatilis discipulorum* sp. nov. *Int J Syst Evol Microbiol.* 2015 Jul;65(7):2227-33. doi: 10.1099/ijs.0.000190.

Smedley D et al. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* 2015 Jul 1;43(W1):W589-98. doi: 10.1093/nar/gkv350.

Song Q et al. SNP Assay Development for Linkage Map Construction, Anchoring Whole Genome Sequence and Other Genetic and Genomic Applications in Common Bean. *G3* (Bethesda). 2015 Aug 28. pii: g3.115.020594. doi: 10.1534/g3.115.020594.

Sun CL et al. Metagenomic reconstructions of bacterial CRISPR loci constrain population histories. *ISME J.* 2015 Sep 22. doi: 10.1038/ismej.2015.162. [Epub ahead of print]

Tennessen K et al. ProDeGe: a computational protocol for fully automated decontamination of genomes. *ISME J.* 2016 Jan;10(1):269-72. doi: 10.1038/ismej.2015.100. Epub 2015 Jun 9.

Tian R et al. High-quality permanent draft genome sequence of Bradyrhizobium sp. Ai1a-2; a microsymbiont of *Andira inermis* discovered in Costa Rica. *Stand Genomic Sci.* 2015 Jun 14;10:33. doi: 10.1186/s40793-015-0007-z.

Tian R et al. High-quality permanent draft genome sequence of Bradyrhizobium sp. Tv2a.2, a microsymbiont of *Tachigali versicolor* discovered in Barro Colorado Island of Panama. *Stand Genomic Sci.* 2015 May 17;10:27. doi:10.1186/s40793-015-0006-0.

Tian R et al. High-quality permanent draft genome sequence of Bradyrhizobium sp. Th.b2, a microsymbiont of *Amphicarpaea bracteata* collected in Johnson City, New York. *Stand Genomic Sci.* 2015 May 16;10:24. doi:10.1186/s40793-015-0008-y.

Tisa LS et al. Draft Genome Sequence of Frankia sp. Strain DC12, an Atypical, Noninfective, Ineffective Isolate from *Datisca cannabina*. *Genome Announc.* 2015 Aug 6;3(4). pii: e00889-15. doi: 10.1128/genomeA.00889-15.

Tremblay J et al. Primer and platform effects on 16S rRNA tag sequencing. *Front Microbiol.* 2015 Aug 4;6:771. doi: 10.3389/fmicb.2015.00771.

Tripp HJ et al. Toward a standard in structural genome annotation for prokaryotes. *Stand Genomic Sci.* 2015 Jul 25;10:45. doi: 10.1186/s40793-015-0034-9.

Tschitschko B et al. Antarctic archaea-virus interactions: metaproteome-led analysis of invasion, evasion and adaptation. *ISME J.* 2015 Sep;9(9):2094-107. doi: 10.1038/ismej.2015.110.

Varghese NJ et al. Microbial species delineation using whole genome sequences. *Nucleic Acids Res.* 2015 Aug 18;43(14):6761-71. doi: 10.1093/nar/gkv657.

Vishnivetskaya TA et al. Draft genome sequences of 10 strains of the genus exiguobacterium. *Genome Announc.* 2014 Oct 16;2(5). pii: e01058-14. doi: 10.1128/genomeA.01058-14.

Walworth N et al. Trichodesmium genome maintains abundant, widespread noncoding DNA in situ, despite oligotrophic lifestyle. *Proc Natl Acad Sci* USA. 2015 Apr 7;112(14):4251-6. doi: 10.1073/pnas.1422332112.

Wang X et al. Bayesian Spatial-Temporal Modeling of Ecological Zero-Inflated Count Data. *Stat Sin.* 2015 Jan;25(1):189-204. doi: 10.5705/ss.2013.212w

Wetmore KM et al. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *MBio.* 2015 May 12;6(3):e00306-15. doi: 10.1128/mBio.00306-15.

Whitman WB et al. Genomic Encyclopedia of Bacterial and Archaeal Type Strains, Phase III: the genomes of soil and plant-associated and newly described type strains. *Stand Genomic Sci.* 2015 May 17;10:26. doi:10.1186/s40793-015-0017-x.

Woo HL et al. Complete genome sequence of the lignin-degrading bacterium Klebsiella sp. strain BRL6-2. *Stand Genomic Sci.* 2014 Dec 8. doi: 10.1186/1944-3277-9-19

Woyke T et al. Evolution. Searching for new branches on the tree of life. *Science.* 2014 Nov 7;346(6210):698-9. doi: 10.1126/science.1258871.

Woyke T et al. Function-driven single-cell genomics. *Microb Biotechnol.* 2015 Jan;8(1):38-9. doi: 10.1111/1751-7915.12247.

Yang X et al. A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. *New Phytol.* 2015 Aug;207(3):491-504. doi: 10.1111/nph.13393.

Yassin AF et al. High quality draft genome sequence of *Corynebacterium ulceribovis* type strain IMMIB-L1395 T (DSM 45146 T). *Stand Genomic Sci.* 2015 Aug 5, 10:50. doi:10.1186/s40793-015-0036-7

Yates R et al. High-quality permanent draft genome sequence of *Rhizobium sullae* strain WSM1592; a *Hedysarum coronarium* microsymbiont from Sassari, Italy. *Stand Genomic Sci.* 2015 Jul 24;10:44. doi: 10.1186/s40793-015-0020-2. eCollection 2015.

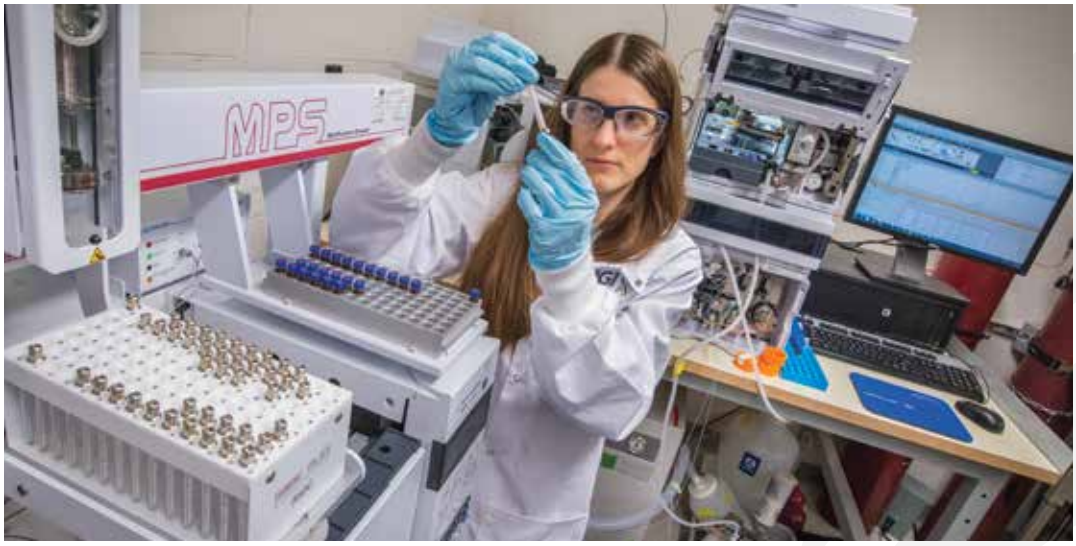
Young D et al. Degradation of Bunker C Fuel Oil by White-Rot Fungi in Sawdust Cultures Suggests Potential Applications in Bioremediation. *PLoS One.* 2015 Jun 25;10(6):e0130381. doi: 10.1371/journal.pone.0130381.

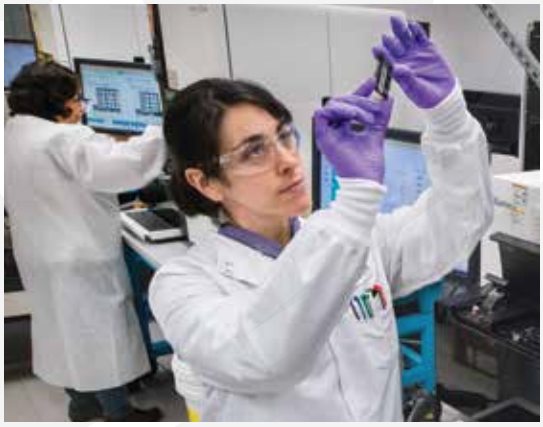
Youssef NH et al. In Silico Analysis of the Metabolic Potential and Niche Specialization of Candidate Phylum "Latescibacteria" (WS3). *PLoS One.* 2015 Jun 3;10(6):e0127499. doi: 10.1371/journal.pone.0127499.

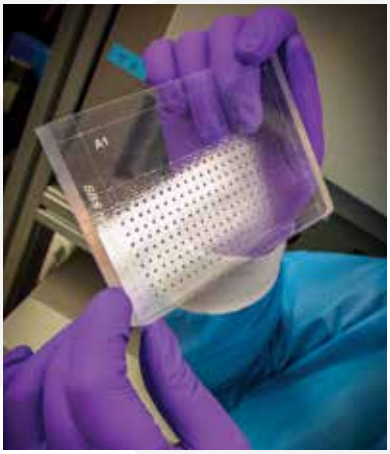
Yung MC et al. Tn-seq of *Caulobacter crescentus* under uranium stress reveals genes essential for detoxification and stress tolerance. *J Bacteriol.* 2015 Oct;197(19):3160-72. doi: 10.1128/JB.00382-15. Epub 2015 Jul 20.

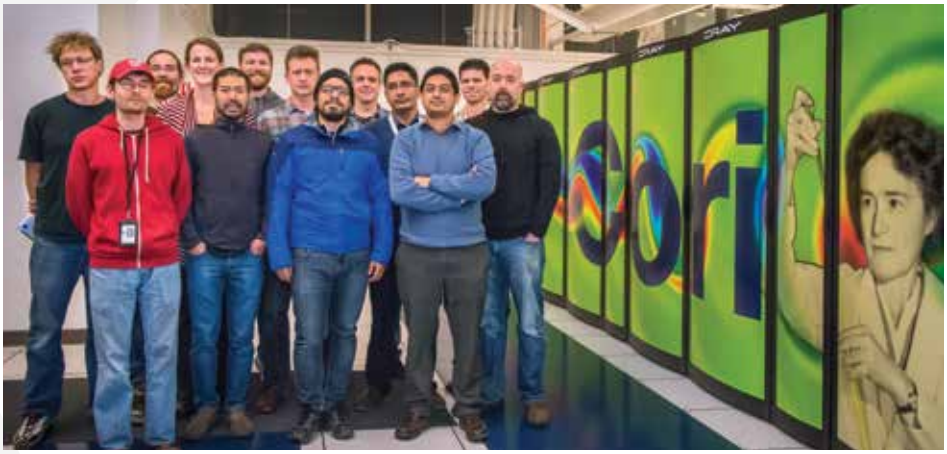
Zhou A et al. Rapid selective sweep of pre-existing polymorphisms and slow fixation of new mutations in experimental evolution of *Desulfovibrio vulgaris*. *ISME J.* 2015 Nov;9(11):2360-72. doi: 10.1038/ismej.2015.45. Epub 2015 Apr 7.

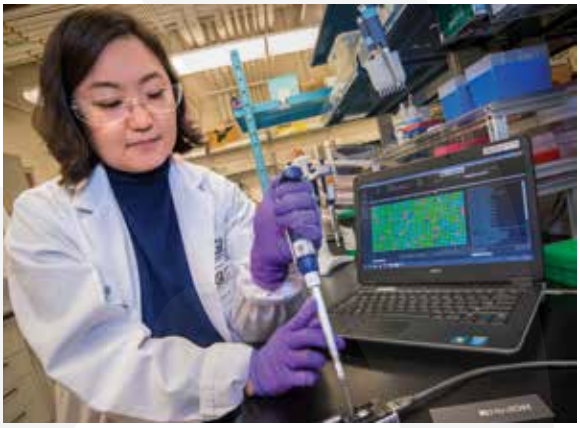
Zhou J et al. High-throughput metagenomic technologies for complex microbial community analysis: open and closed formats. *MBio.* 2015 Jan 27;6(1). pii: e02288-14. doi: 10.1128/mBio.02288-14.

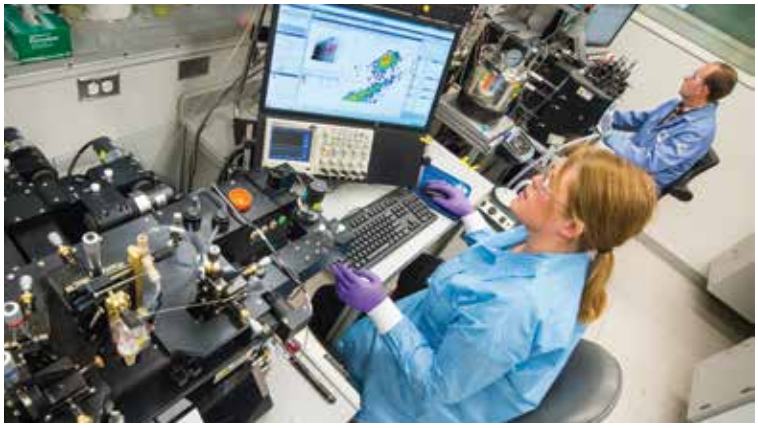












Comments?

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.

Massie Santos Ballon, DOE JGI Public Affairs, Editor

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

16-JGI-2520

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