

**2017**

# Progress Report

**U.S. Department of Energy**

**Joint Genome Institute**

Cover Photo:

One of the world's largest living organisms can be found in a forest, but its presence can only be detected by the appearance of fruiting bodies, better known as honey mushrooms, around infected trees. *Armillaria* fungi are among the most devastating fungal pathogens; the fungus *Armillaria ostoyae* was sequenced and analyzed by the JGI to better understand the evolution of *Armillaria's* abilities to spread and infect, and effectively break down all components of plant cell walls.

(Armand Robichaud, Flickr CC BY-NC 2.0)

Impact Section Case Study credits (in order of appearance):

*Laccaria bicolor* (Francis Martin, INRA); *Postia placenta* (Courtesy of Tom Kuster, FPL); *Scheffersomyces stipitis* (Courtesy of Tom Jeffries, University of Wisconsin, Madison); and, *Daldinia eschscholzii* (Petra Gloyn CC BY-NC-ND 2.0)

All photography, unless otherwise noted, by Marilyn Chung, Berkeley Lab.





Table of  
Contents



1	JGI Mission
2	Director's Perspective
8	Achieving the DOE Mission
10	Organizational Structure
12	Impact 2017
18	Case Study: 10 Years of Mycosm: Approaching 1,000 Fungal Genomes
20	Science: Year in Review
21	Discovery
30	Bioenergy
40	Biogeochemistry
48	Computational Infrastructure
50	Appendices
51	Appendix A: Acronyms at a Glance
53	Appendix B: Glossary
56	Appendix C: 2017 User Programs Supported Proposals
65	Appendix D: Advisory and Review Committee Members
67	Appendix E: 2017 Genomics of Energy and Environment Meeting
69	Appendix F: 2017 Publications



Aerial view looking toward Berkeley Lab (middle left),  
with JGI near the base of Mount Diablo. (*Jeremiah  
Higgins, Unsplash*)

# JGI Mission



The U.S. Department of Energy (DOE) Joint Genome Institute's (JGI) mission is to serve the diverse scientific community as a national user facility, by making available large-scale genomics and analysis of plants, microbes, and communities of microbes. This addresses the DOE mission goal of harnessing science and technology to address energy and environmental challenges.



Nigel Mouncey, *Director, DOE Joint Genome Institute*

# Director's Perspective

## Continuing Evolution

In 2017, the JGI continued its evolution towards an Integrative Genome Science User Facility that provides a diverse scientific community with access to state-of-the-art technologies and scientific expertise, enabling biological discoveries and applications in the mission areas of bioenergy, nutrient cycling, and biogeochemistry.

The JGI offers a suite of capabilities unique in their ability and scale to advance energy and environmental science. Collectively, the JGI's investments in infrastructure and data management, computational tool development, and cutting-edge approaches to data mining have paved the way to translate large-scale and diverse data into biological discoveries. The JGI identifies grand scientific challenges and provides the leadership, tools, expertise and resources to address them. I joined as the JGI's new Director in March 2017, bringing with me almost 20 years' experience in conducting and managing research programs and research and development organizations in the industrial biotechnology private sector, most recently at Dow AgroSciences. It is the tremendous science, expertise and the ability to advance scientific knowledge that attracted me to the JGI and Lawrence Berkeley National Laboratory's (Berkeley Lab) Biosciences Area.

I would like to recognize the outstanding leadership of Axel Visel, who served as Interim Director from March 2016 until my arrival.

Over the last year, the JGI has continued to successfully grow its user and science programs, in alignment with its scientific directions. We are developing a wide range of large-scale experimental and computational capabilities to link sequence to biological insights whose targets are largely driven by systematic (phylogenetic) and problem-focused (e.g., plant-microbe interactions) approaches, as well as large-scale functional studies. A major continued thrust at the JGI is to provide integrated capabilities that, for example, allow users to mine genomic databases, design and synthesize constructs, introduce these into suitable hosts, and determine gene/pathway function through metabolomics. Seamless integration of these capabilities provides users with a powerful suite of technologies that offer value far beyond their individual components. Significant progress has been made towards the two- and five-year milestones defined in the 2016 Strategic Planning Update, with 63% of the two-year milestones now completed.

As a User Facility, in fiscal year (FY) 2017 the JGI served a worldwide community of 1,598 users from academia, government, and industry. In support of their projects, we produced nine new and 2,646 resequenced plant genomes, 194 annotated fungal genomes, and 1,826 microbial genomes (including single-cell draft genomes), as well as 1,291 metagenomic data sets. As we continue to advance our functional genomics, we have generated 1,523 plant transcriptomic and 511 metatranscriptomic datasets. In total, the JGI produced a record sequencing output of 178 trillion bases of sequence data in FY17 alone. Additionally, our DNA Synthesis Science Program produced six million bases of DNA. Data from the sequencing projects are disseminated to the broader scientific community via the respective Science Program data portals, as well as the JGI's genome portal. These data platforms continue to attract new users: more than 11,000 new portal users accessed our data portals in FY17, and more than 100,000 visitors came for data access and analysis.

The demand for our user programs is also increasing. In 2017, we received 76 proposals submitted to the large-scale Community Science Program (CSP), as well as 153 proposals in response to the small-scale microbial/metagenome and DNA synthesis CSP calls, and 51 proposals to the joint Facilities Integrating Collaborations for User Science (FICUS) call with the Environmental Molecular Sciences Laboratory (EMSL) at Pacific Northwest National Laboratory (PNNL). FICUS is a user program collaboratively offered by the JGI with, as of 2017, two other national user facilities: EMSL and the National Energy Research Scientific Computing Center (NERSC) at Berkeley Lab. This Microbiome Data Science FICUS provides for the first time a service that is based on data rather than traditional products (e.g., sequencing). Across all of these programs, a total of 108 proposals were approved and accepted after peer review and are now in progress.

The impact of the research performed through all of these accepted proposals is reflected by the large number of scientific publications originating from the JGI and its users. In 2017, the JGI contributed to 147 peer-reviewed publications, including 17 in the most selective journals, *Science* and *Nature*. Seven of our scientists were included in the annual Highly Cited Researchers list by Clarivate Analytics (formerly Thomson Reuters), and several of our scientists serve on editorial boards of leading scientific journals. Selected scientific highlights published in 2017 included:

- **Toward More Effective Carbon Fixation.** By tapping the DNA synthesis expertise of the JGI, a team from the Max Planck Institute (MPI) for Terrestrial Microbiology has reverse engineered a biosynthetic pathway for more effective carbon fixation. This novel pathway is based on a new CO<sub>2</sub>-fixing enzyme that is nearly 20 times faster than the most prevalent enzyme in nature responsible for capturing CO<sub>2</sub> in plants by using sunlight as energy. Successfully reconstituting a synthetic enzymatic network in a test tube for the conversion of CO<sub>2</sub> into organic products could lead to potential future applications, such as artificial photosynthesis (*Science*, November 18, 2016).
- **Antarctic Adaptations in Diatoms.** To learn more about how the Antarctic alga *Fragilariopsis cylindrus* adapted to its extremely cold environment, a team led by University of East Anglia (UEA) scientists conducted a comparative genomic analysis involving three diatoms by tapping the JGI's sequencing and annotation expertise. The paper was recently named a Fast-Breaking Paper for Biology and Biochemistry by Clarivate Analytics (*Nature*, January 16, 2017).
- **Aspergillus Diversity for Industrial Applications.** An international team including JGI researchers reported sequencing the genomes of 10 novel *Aspergillus* species, more than doubling the number of *Aspergillus* species sequenced to date. Comparing the newly sequenced genomes with the eight other sequenced *Aspergillus* species provided the first ever genus-wide view. The consortium found that *Aspergillus* has a greater genomic and functional diversity than previously understood, broadening the range of potential applications for the fungi considered one of the most important workhorses in biotechnology (*Genome Biology*, February 14, 2017).
- **Major Gene Expression Regulator in Fungi.** Just four letters— A, C, T, and G — make up an organism's genetic code. Changing a single letter, or base, can lead to changes in protein structures and functions, affecting an organism's traits. In addition, though, subtler changes can and do happen, involving modifications of the DNA bases themselves. A JGI-led team reported the prevalence of 6-methyladenine (6mA) modifications in the earliest branches of the fungal kingdom, and proposed the role of 6mA in gene regulation (*Nature Genetics*, May 8, 2017).
- **Fungal Enzyme Complexes to Break Down Cellulose.** A team led by researchers at the University of California (UC), Santa Barbara, has found for the first time that early lineages of fungi can form complexes of enzymes called "cellulosomes" capable of degrading plant biomass. Consolidating these enzymes, in effect into protein assembly lines, makes them team up to work more efficiently than they would as individuals. Through the FICUS JGI-EMSL program, the team found cellulosomes in anaerobic gut fungi that attack plant biomass as a cluster of enzymes (*Nature Microbiology*, May 26, 2017).

- **IMG/M Database Helps Fill Previously Unknown Protein Structures.** The Pfam database contains close to 15,000 protein families: groups of proteins that share an evolutionary origin. A team led by the University of Washington's David Baker reported that structural models have been generated for 614 (12%) of the protein families that had previously had no structural information available. This accomplishment was made possible through a collaboration in which the Baker lab's protein structure prediction server, Rosetta, analyzed the publicly available metagenomic sequences on the JGI's Integrated Microbial Genomes (IMG) system (*Science*, January 20, 2017).
- **1003 Microbial Genomes Released.** Culminating nearly a decade's worth of work, JGI's Prokaryotic Super Program head Nikos Kyrpides and his team reported the release of 1,003 phylogenetically diverse bacterial and archaeal reference genomes — the single largest release to date. With the release of this high-quality genomic information, the JGI provides a wealth of new sequences that will be invaluable to scientists interested in experiments such as characterizing biotechnologically relevant secondary metabolites or studying enzymes that work under specific conditions (*Nature Biotechnology*, June 12, 2017).
- **Novel Group of Giant Viruses Discovered.** Giant viruses were first discovered in 2003, and a handful of other giant virus groups have been found since. JGI scientists' report of a novel group of giant viruses — dubbed "Klosneuviruses" — with a more complete set of translation machinery genes than any other virus known to date resolved the debate over viruses as a possible fourth domain of life by demonstrating that eukaryote translation-encoding genes had been taken up by and incorporated into the viral genomes (*Science*, April 7, 2017).
- **Minimum Metadata Standards.** As genomic data production has ramped up over the past two decades and is being generated on various platforms around the world, scientists have worked together to establish definitions for terms such as "draft assembly" and data collection standards that apply across the board. An international team led by JGI researchers has developed standards for the minimum metadata to be supplied with single amplified genomes and metagenome-assembled genomes submitted to public databases (*Nature Biotechnology*, August 8, 2017).

To continue enabling these types of consequential scientific discoveries into the future, we invest in an expert workforce and develop genomic technologies that expand frontiers. Over the last year, we have brought into production higher throughput and capacity short-read instrumentation and more robust single-molecule real time long read technology, and continue to develop nanopore-based sequencing technologies. For functional genomics, we have established new capabilities in experimental transcription factor binding motif detection (DAP-Seq), chromatin

At the IGB groundbreaking (Left to Right): Berkeley Lab Director Michael Witherell, DOE's Sharlene Weatherwax, JGI's Nigel Mouncey, KBase's Adam Arkin, UC's Kimberly Budil, and Berkeley Lab's Jay Keasling.



conformation analysis (ChIP-Seq), and virus-based sequencing. In single-cell genomics, we have expanded into eukaryotic *de novo* genome assembly and the binning of cells containing similar transcriptomes. We continue to provide users with capabilities to determine gene and pathway function through expansion of our DNA synthesis science and metabolomics platforms. These capabilities allow researchers to delve deeper into genomes and maintain the JGI's relevance through the continuous addition of novel, pioneering approaches available to users.

The JGI Science Programs also develop novel approaches and resources relevant to our mission, showcasing these through high-visibility publications such as the ones cited earlier. These initial demonstration studies serve as an inspiration to our user base to adopt new JGI capabilities and have proven invaluable in developing the JGI user communities. For example, we continue our leadership role in microbial diversity research by expanding the tree of life through exploiting differential coverage binning of deeply sequenced metagenomes to identify new phyla. Additionally, the Plant Program, led by Jeremy Schmutz and largely executed at our partner institute, HudsonAlpha Institute for Biotechnology in Huntsville, Alabama, continues to be one of the world's leading plant genomics programs. Through this long-standing partnership, the JGI generates extensively annotated *de novo* and improved genomes of JGI Flagship plants and comparator species, performs large-scale population genetics and gene expression studies to link genes to bioenergy-relevant traits, and characterizes evolutionary genome dynamics and their influence on functional adaptations through pan-genome analyses.

Today's ability to generate data at a scale and pace that outweighs the capacity to manage, analyze, and interpret data requires scalable computational infrastructure and tools. NERSC and the JGI have developed a deeper partnership that has led to the successful implementation of several computing initiatives, including migration of pipelines to Docker for increased portability, consolidation of workflow management systems, and extensive training on emerging computing technologies. The DOE Office of Science has a major initiative, the Exascale Computing Project (ECP), dedicated to developing an exascale ecosystem to advance scientific discovery. The JGI partners with the Berkeley Lab Computational Research Division and participates in ECP through the ExaBiome project. The main aim of this work is to adapt key elements of the JGI computing workload to run at exascale (a billion billion calculations per second), so the JGI analysis can advance in step with the enormous volumes of data that will be generated. In 2017, we also completed the automated GenBank submission system's development to release our data to NCBI.

With the wealth and breadth of data being generated, the JGI continues to develop the tools and systems for users to extract new insights. For example, the number of registered users in our IMG system doubled (surpassed 18,000 registrations from 102 countries), as has the number of metagenome datasets integrated in IMG. We have now launched an institutional effort to overhaul IMG (with the working title IMG-Next), which is making good progress towards providing users with improved pipelines and interfaces, and providing scalable database architecture to accommodate and enable the analysis of rapidly growing amounts of microbial genome and microbiome data.

The Emerging Technologies Opportunity Program (ETOP) was designed to facilitate JGI user access to the latest sequence-to-function capabilities developed by external groups collaborating with the JGI. Excellent progress has been made in our ongoing projects, some examples of which are detailed in the "Science: Year in Review" section. This year, we also launched a new ETOP call for proposals related to rapid-prototyping systems to rapidly and at scale determine gene/pathway function.

As a DOE User Facility, the JGI must engage actively with users. Our 12th Annual Genomics of Energy and Environment Meeting filled our Walnut Creek meeting space to capacity. As a result, the meeting will be held in San Francisco in 2018. We had an exciting lineup of speakers, with keynote addresses from Tobias Erb of the Max Planck Institute (MPI) for Terrestrial Ecology and Titus Brown of the University of California, Davis (see Appendix E.) The JGI also hosted the NeLLi (New Lineages of Life) and Microbial and Plant Systems Modulated by Secondary Metabolites meetings, and the first National Microbiome Data Collaborative workshop. The Microbial Genomics and Metagenomics (MGM) workshops marked their tenth year in 2017, and along with that, the milestone of training more than 1,000 scientists from 55 countries in using the IMG resources.

We are also expanding our communication and outreach efforts to attract new users from both the academic and private sectors, and in 2017 formed a Business Development Team to lead a new Industry Engagement Program. This program has already had discussions with more than 50 companies and launched its first sponsored research project with an industrial partner at JGI.

The foundations of building and sustaining high-performing teams at the JGI are extensive training programs (especially for safety), the Performance Management Process, and people development opportunities. The JGI has an exceptionally highly skilled workforce motivated to deliver science and expertise to our users. I have established a new JGI Leadership Team (JLT) comprised of the JGI Director; Deputies for User Programs, Science Programs, Genomic Technologies, and Operations; Chief Informatics Officer, Senior Communications and Outreach Manager, and Business Development and Program Management Office Leader. The JLT has continued to emphasize the safety and wellness of our staff as an institutional priority, and our outstanding safety record has had a positive impact on our overall operational efficiency and productivity.

A core JGI value is diversity and inclusion (D&I), and our culture of inclusion helps us attract, retain, and develop outstanding talent, thereby directly promoting our scientific excellence. Our D&I Working Group is very active in promoting monthly social events, arranging talks and videos to stimulate discussion, and organizing trainings (e.g., to raise awareness of implicit bias). Additionally, the JGI is committed to ensuring that its management teams, advisory bodies, and speakers at JGI meetings and review panels represent the full demographic and scientific diversity of the JGI and its user community. We launched the OneJGI initiative in 2017 to enhance collaboration, trust, and respect at the JGI by revisiting and communicating the JGI's core values and norms and ensuring open and transparent communication.

Finally, the first step in realizing Berkeley Lab's vision to co-locate Biosciences activities and organizations at the main Berkeley Lab site is the JGI's new home, the Integrative Genomics Building (IGB). A groundbreaking ceremony was held at the end of January 2017 to kick off the construction phase of this project. Excellent progress has been made on the construction, with foundations completed and the first-floor structure erected. We aim to move into this building in the first half of 2019.

We are very excited about this move, and the expected scientific and operational productivity gains for the JGI and the DOE Systems Biology Knowledgebase (KBase), as well as related research programs in the Biosciences Area. Achieving systems biology's promise to understand and solve basic energy and environment problems requires the ability to iteratively predict the behavior of organisms and biological systems, then test those predictions and adjust models accordingly. To fulfill this promise, we have strengthened our partnership with KBase, with the aim of enabling seamless interactions to provide our users with access to KBase tools and data that go beyond what is offered through JGI resources.

In summary, 2017 was another highly productive year for the JGI, and we continue to offer our global user community a unique suite of state-of-the-art capabilities essential to enable the advancement of energy and environmental science and address scientific grand challenges. As the JGI continues its evolution into an *Integrative Genome Science User Facility* that is the leading center worldwide for energy and environmental genomics, our science, our capabilities, our users, and our OneJGI culture all position us for growing JGI's future success. This will be the guiding principle as we conduct a refresh of our strategic plan in 2018 that will guide the JGI to meet the future scientific needs of its broad and diverse users.



**Nigel J. Mouncey, DPhil**  
*Director, DOE Joint Genome Institute*



Achieving the  
**DOE Mission**



The Joint Genome Institute (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 plant biomass feedstocks for biofuels by exploiting genomic knowledge of plants, microbes, fungi, and microbial communities.
- Gain insights into biogeochemical processes controlling the cycling of carbon, nitrogen, and key nutrients in environments and the mobility of heavy metals and radionuclides at contaminated sites for which DOE has stewardship responsibilities.

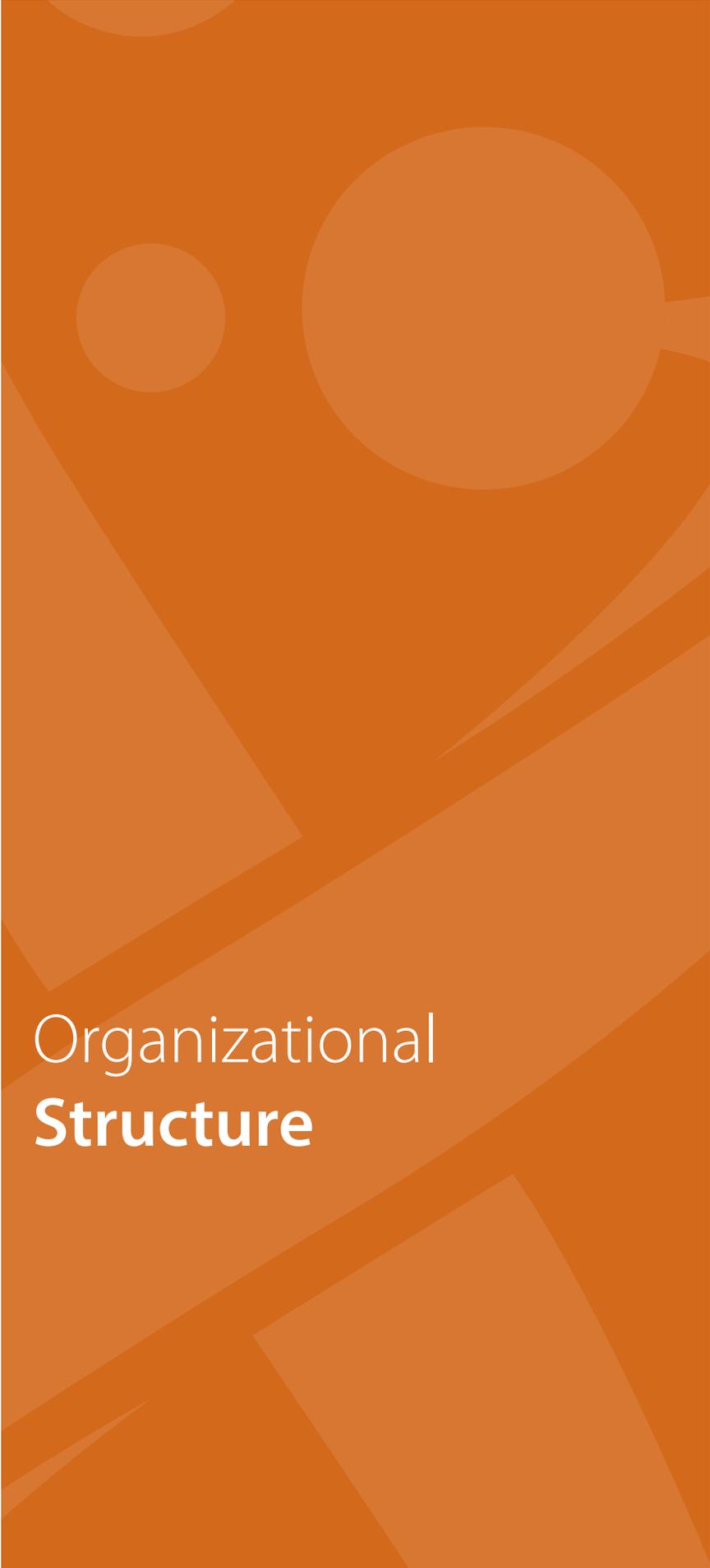
## Bioenergy

The United States is the world's largest consumer of petroleum, with most 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 the existing fuel distribution infrastructure. Sequencing projects at the JGI that contribute to meeting this goal focus on at least one of three categories: terrestrial plant feedstocks for biofuel production and their associated microbial communities (microbiomes); fungi, microbes, and microbial communities that break down the lignin and cellulose in plant walls; and organisms that convert lignocellulose-derived sugars or lignols into biofuels or other bioproducts currently produced from petroleum, such as plastics.

## Biogeochemistry

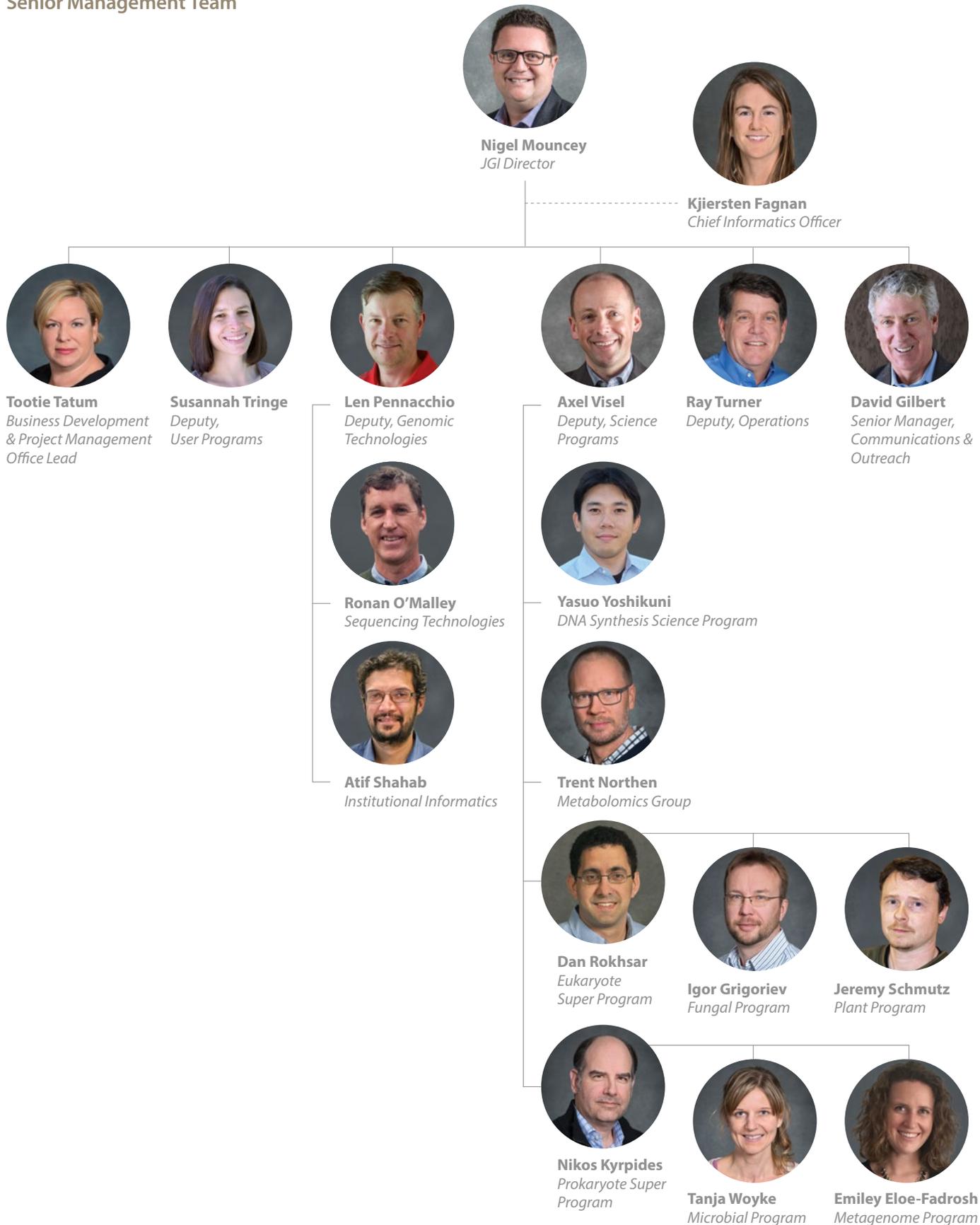
Many DOE-relevant environmental processes are controlled by complex, interconnected biogeochemical reactions. The JGI engages in projects that can couple a genome-enabled understanding of biological processes in the context of the physical, chemical, and geochemical processes controlling the cycling and fate of key elements in environments affecting BER's energy and environmental missions. Microbes and microbial communities of interest to the JGI as targets for sequencing include those involved in terrestrial carbon, nitrogen, phosphorus, sulfur, and other macronutrient cycles that affect sustainable bioenergy crop growth or global carbon cycling. Others include 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. As microbes constitute the largest component of Earth's biodiversity and biomass, understanding how they metabolize these elements and how environmental changes affect these processes is crucial.

With the support of BER through the Subsurface Systems Scientific Focus Area (SFA) 2.0, Berkeley Lab's Earth and Environmental Sciences Area (EESA) and their JGI collaborators have conducted major investigations at the field site in Rifle, Colorado to facilitate integrated, field-based subsurface biogeochemical and microbial genomics research relevant to uranium mobility to improve the predictive understanding of subsurface flow and transport relevant to metal and radionuclide contaminants. *(Roy Kaltschmidt, Berkeley Lab)*

The background of the slide is a solid orange color. It features several large, semi-transparent geometric shapes in various shades of orange, including circles and triangles, which are layered and overlap each other. The shapes are positioned in the upper and lower portions of the slide, leaving a clear space for the text in the center.

# Organizational **Structure**

## Senior Management Team



The image features a dark grey background with several overlapping circles in various shades of blue and grey. A prominent vertical bar in a medium blue color runs through the center of the page. The text 'Impact 2017' is positioned in the lower-left area of this bar.

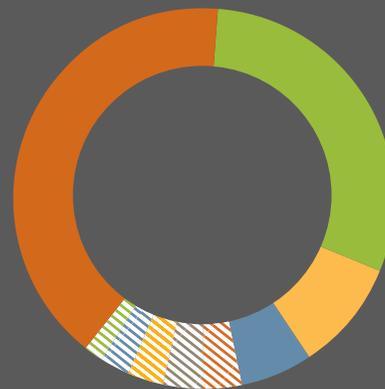
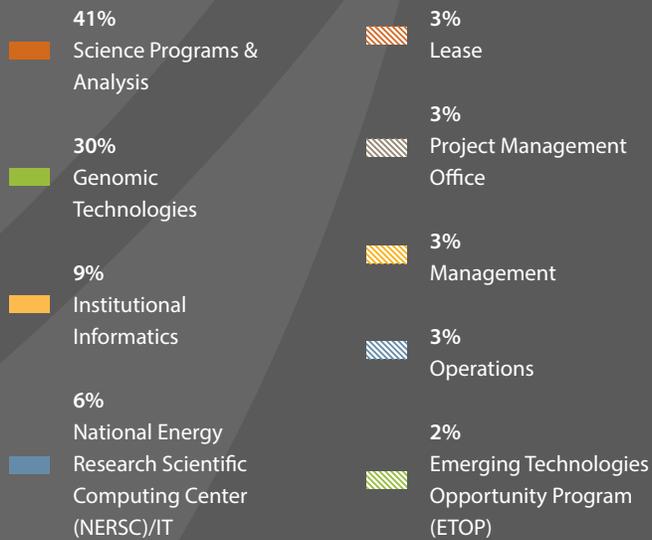
Impact **2017**

### Primary Users **Fiscal Year 2017**

This category captures the primary users of the JGI, which include PIs and their collaborators on all user projects that were active during FY 2017. 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 JGI or funded through the JGI's partner subcontracts.



### Spending Profile **FY2017**



**Users on the Map: 1,598**

Academic **1,130**  
 Other **316**  
 DOE National Laboratory **134**  
 Company **18**

**North America**  
**1,195**  
 United States  
**1,113**

**Europe**  
**299**

**12** **South America**

**North America 1,195**

**South America 12**

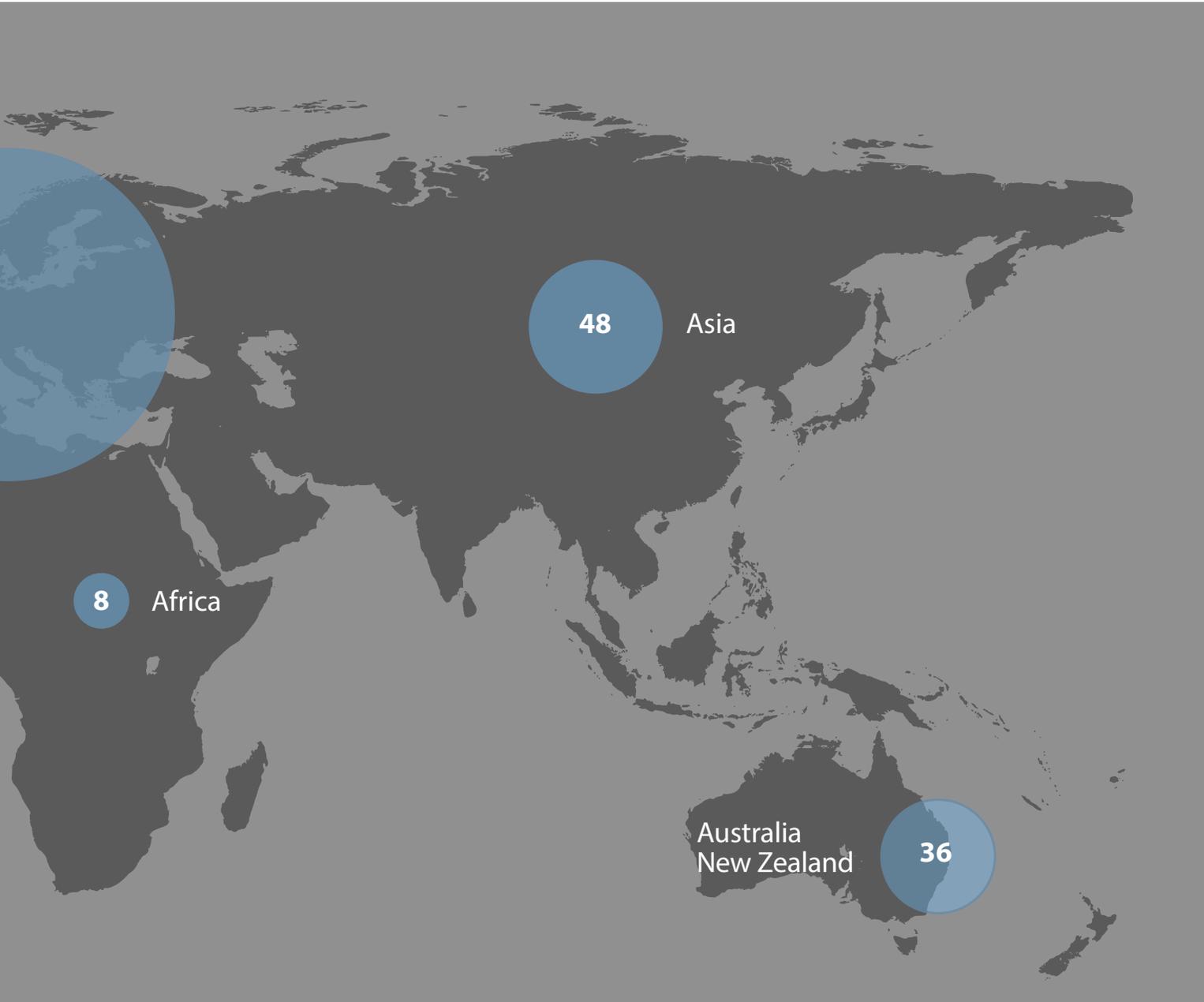
**Europe 299**

United States 1,113  
 Canada 75  
 Mexico 6  
 Panama 1

Brazil 9  
 Chile 2  
 Peru 1

Austria 23  
 Belgium 7  
 Czech Republic 3  
 Denmark 6  
 Finland 11  
 France 35  
 Germany 64  
 Greece 2  
 Hungary 1  
 Iceland 1  
 Italy 14

Netherlands 17  
 Norway 6  
 Poland 1  
 Portugal 3  
 Russian Federation 4  
 Slovenia 1  
 Spain 36  
 Sweden 12  
 Switzerland 8  
 Turkey 2  
 United Kingdom 42



Africa		Asia		Australia & New Zealand	
	8		48		36
Egypt	1	China	10	Australia	30
South Africa	7	India	7	New Zealand	6
		Israel	5		
		Japan	17		
		Korea, Republic of	1		
		Malaysia	1		
		Singapore	1		
		Taiwan	6		

## Users of JGI Tools & Data

The JGI produces high-quality data that are made available to the community through our data portals. These investigators are not included in the primary user count because their projects were not conducted as part of JGI's user programs. When users log in to the systems and download data, this activity is tracked in order to help

us understand which data sets are of greatest value to our Users. Additionally, the JGI's data management system is able to restore data from HPSS upon user request and this activity is logged. In 2017, JGI users downloaded more than 1.8 million data sets and 1 PB of data that were restored through the JAMO system.

### Workshops and Meetings

Workshops and Meetings	Participants
Genomics of Energy & Environment 12th Annual User Meeting	512
Other Workshops	1,397

User Meeting Workshop	Total Attendees	External Only
Fungal	51	38
Genomic Technologies	95	65
KBase	64	44
Synthetic Biology	55	31
NERSC	44	29
Metabolomics	75	51
<b>TOTAL</b>	<b>258</b>	<b>258</b>

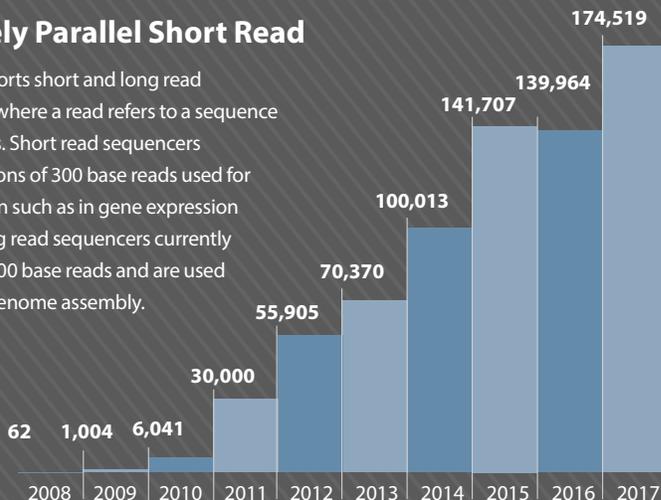
Date	Conferences	# of Attendees
Sept. 18-22, 2017	Microbial Genomics and Metagenomics (MGM) 25	43
July 24-26, 2017	2nd Microbial / Plant Systems Modulated by Secondary Metabolites	120
July 16-19, 2017	MSA JGI Fungal Program/MycoCosm Workshop	60
June 24-28, 2017	American Plant Biology talk and joint outreach booth with Kbase	50
June 5-8, 2017	CROPS: Phytozome Workshop	62
June 1-5, 2017	ASM Microbe MGM Workshop	30
June 1-5, 2017	ASM Microbe National Microbiome Data Collaborative (NMDC)	300
April 5-6, 2017	NeLLi 2017: From New Lineages of Life to New Functions Workshop	79
Mar. 14-19, 2017	29th Fungal Genetics Conference	200
Feb. 27- Mar. 3, 2017	MGM-24	45
Jan. 17, 2017	PAG JGI & KBase Plant Sciences Workshop	100
Jan. 17, 2017	PAG XXVI Fungal Workshop	50
<b>TOTAL ATTENDEES</b>	<b>1,139</b>	<b>1,139</b>

## Sequence Output

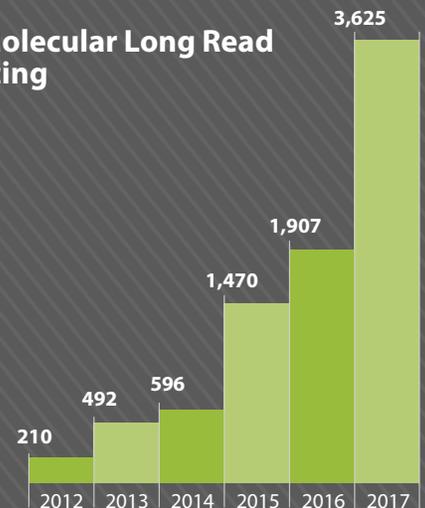
(in billions of bases or GB)

### Massively Parallel Short Read

The JGI supports short and long read sequencers, where a read refers to a sequence of DNA bases. Short read sequencers produce billions of 300 base reads used for quantification such as in gene expression analysis. Long read sequencers currently average 12,000 base reads and are used for *de novo* genome assembly.

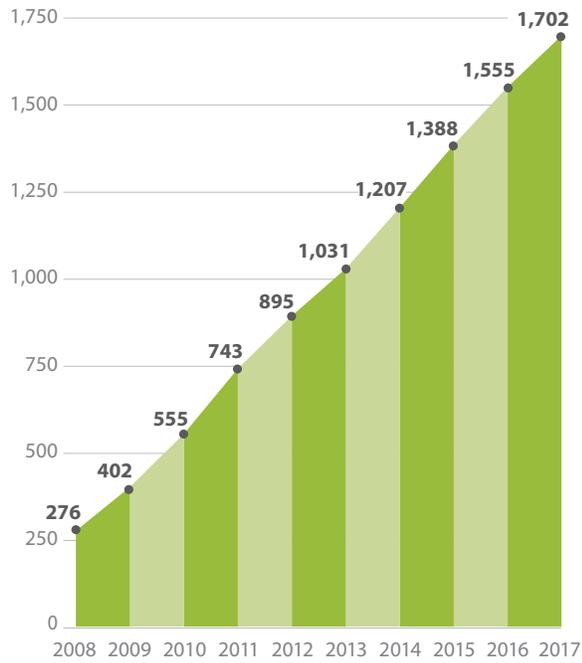


### Single Molecular Long Read Sequencing



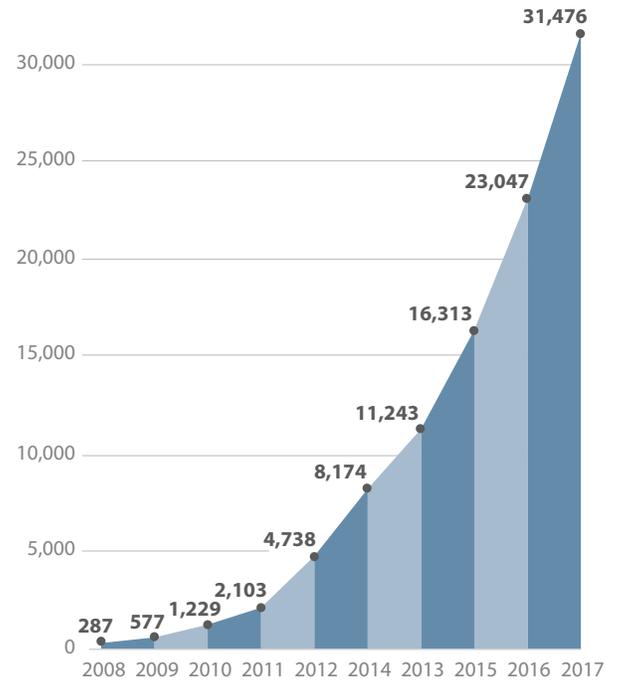
Combined short read and long read totals per year give JGI's annual sequence output.

### Cumulative Number of Scientific Publications

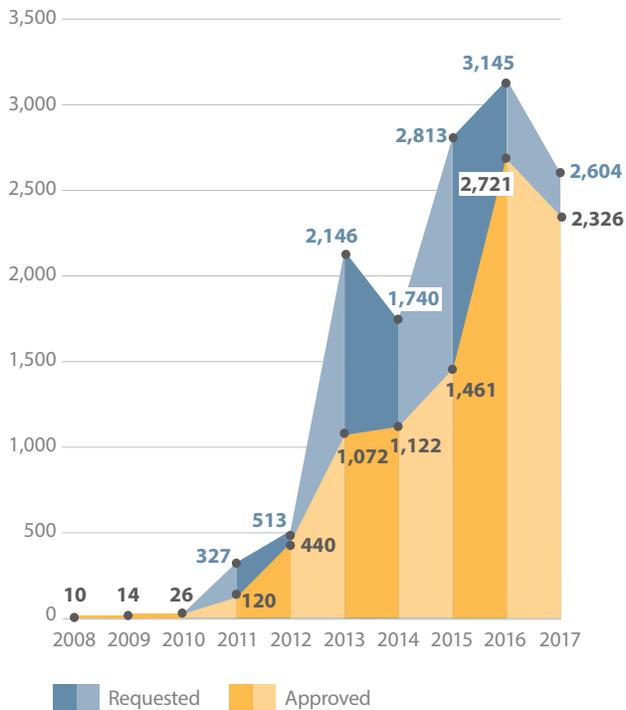


Papers published 2013-2015 were cited by researchers over 5,200 times in 2016.

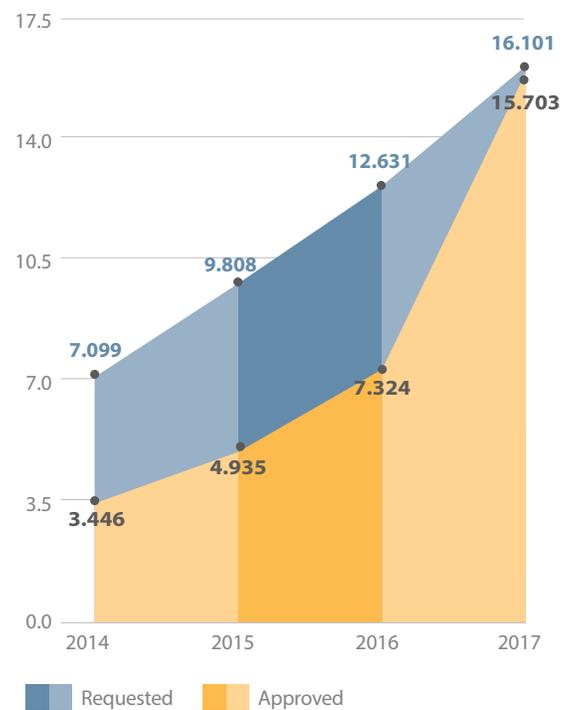
### Cumulative Number of Projects Completed



### User Demand: CSP Single-Cell Genomes Requested & Approved



### User Demand: Synthetic Biology Requested & Approved (Mb)



## Case Study: 10 Years of MycoCosm: Approaching 1,000 Fungal Genomes

MycoCosm is the JGI's web-based repository of fungal resources. With nearly 1,000 genomes, it is the world's largest interactive collection of fungal genomes. MycoCosm started with the integration of six genome projects for six plant pathogenic Dothideomycetes; the JGI held jamborees in 2007–2008 that brought together the six labs responsible for the individual genome projects. This historic exercise in community building and data integration led to the first large-scale fungal comparative genomics project at JGI, and a series of joint publications.

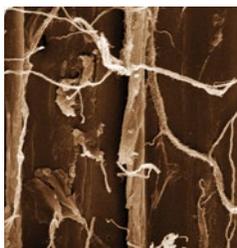
MycoCosm became an official JGI community resource in 2010 and, due to near exponential growth, is approaching an important milestone of 1,000 fungal genomes. Through its community building framework, there were 8,500 registered users and 130,000 distinct web visitors in 2017 alone.

JGI has transformed mycology into genome-based science. MycoCosm enabled the transition from a single-genome analysis — sequencing the first mycorrhiza, the first wood decayer, and the first of many other targeted DOE mission-relevant groups — to comparative genomics.



**Plant-Mycorrhizal Associations.** Over 80% of plant species depend on mycorrhizae. The first genomes of each major type of mycorrhizae — arbuscular (*Rhizophagus irregularis*), ericoid (*Oidiodendron maius*), orchid (*Tulasnella calospora*, learn more in the “Science: Year in Review” section), and ectomycorrhizae (*Laccaria bicolor*) — were sequenced and annotated by the JGI. The sequencing and analysis of a larger group of mycorrhizal fungi identified similar but independently evolved symbiotic toolkits, demonstrating their convergent evolution in mycorrhizal fungi across the entire Fungal Tree of Life. The JGI supports studies of fungal interactions

with plants in several model plant-fungal systems; for example, the genome of the dominant ectomycorrhizal *Cenococcum geophilum* suggested clues to plant drought tolerance.



**Lignocellulose Degradation by Filamentous Fungi.** Lignin is the second-most important component of plant biomass after cellulose. Together the two biopolymers support plant growth and protect plants from microbial attack. The first genomes of a white rot fungus (*Phanerochaete chrysosporium*) and a brown rot fungus (*Postia placenta*) were published in 2004 and 2009, respectively, and demonstrated dramatic differences in the fungi's approaches to lignocellulose degradation. In 2012, the JGI published a comparative genomics analysis of 12 wood decay fungi and traced the evolution of class II peroxidases, lignin-degrading enzymes that distinguish white rot from brown rot fungi. Further analysis suggested

a continuum rather than a dichotomy between the white-rot and brown-rot modes of wood decay, and identified several gene families predictive of wood decay mode.



**Sugar-Fermenting Yeasts.** Sugar fermentation is a critical step in converting sugars extracted from plant biomass into biofuels and bio-based products. Plant biomass is composed of a variety of sugars, but the model yeast *Saccharomyces cerevisiae* grows primarily on glucose. The genome of the xylose-fermenting yeast *Scheffersomyces stipites*, sequenced by the JGI in 2007, revealed genes later used in engineering xylose-fermenting *S. cerevisiae* strains. In 2016, the JGI sequenced and analyzed diverse Ascomycete yeasts growing on other types of sugars to produce an expanded catalog of genes, enzymes, and pathways. This yeast project also discovered a genetic code change in *Pachysolen tannophilus* where the codon CUG encodes the amino acid alanine instead of the expected leucine or serine seen in nearly all other eukaryotic nuclear genomes.

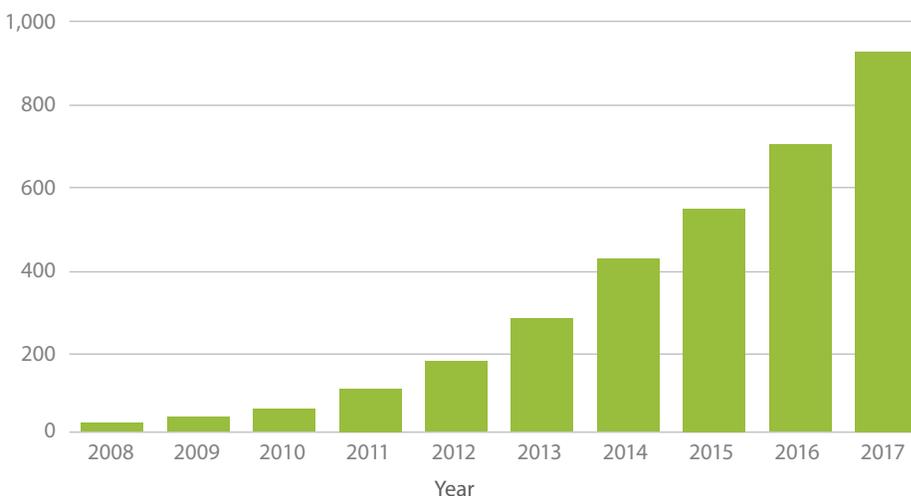


**Genomics of Consolidated Bioprocessing.** Most state-of-the-art platforms for the bioconversion of plant biomass into biofuels involve multiple costly steps of biomass pretreatment and saccharification, challenging the economic feasibility of these technologies. An alternative strategy, consolidated bioprocessing (CBP), aims to combine all these steps into a single unit operation. In collaboration with the Joint BioEnergy Institute (JBEI), the JGI sequenced and analyzed four endophytes (*Hypoxyylon* and *Daldinia* spp.) to assess their potential to produce “mycodiesel” (volatile organic compounds with properties similar to fossil fuels) while growing on lignocellulosic

plant and agricultural residues. Their ability to simultaneously deconstruct lignocellulose and convert it into mycodiesel makes them candidate CBP hosts for biofuel production.

More recent Community Science Program (CSP) projects have significantly advanced MycoCosm’s development. The production of fungal multi-omics datasets for functional genomics is pioneered by fungal miniENCODE (Encyclopedia of DNA Elements) projects. The ongoing JGI 1000 Fungal Genomes project, aimed at producing at least one reference genome for every known family of fungi, is one of the world’s largest genomics studies of fungal diversity. These and similar projects not only ensure further growth of MycoCosm, but also transformed the portal from a static data repository resource into a dynamic platform for the user community to analyze sequenced genomes and nominate new species for sequencing to fill gaps in the Fungal Tree of Life.

### Growth in number of genomes in **MycoCosm**





Among the JGI's major initiatives for the Plant Program is focusing computational and experimental efforts to move beyond sequence to function, ultimately depositing validated accurate data sets into KBase. To that end, the Plant Functional Genomics team uses the model grass *Brachypodium distachyon* and other plants to develop and utilize resources to understand traits such as genome organization and regulation, abiotic stress tolerance and the molecular basis of perenniality.

## Science: A Year in Review



# Discovery

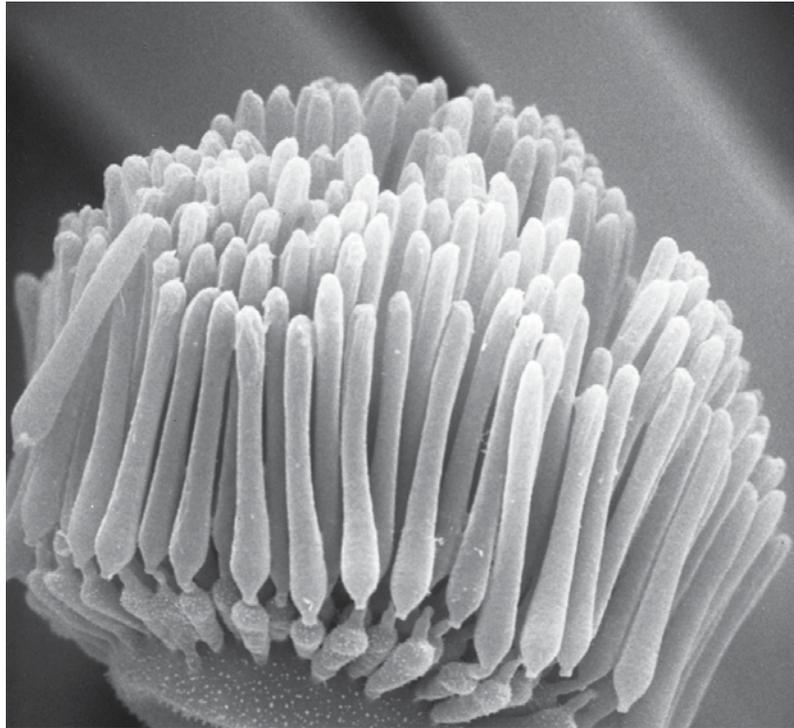
---

## BOOSTing the DNA Sequence Design-and-Build Process

The ability to design and manufacture synthetic DNA has opened tremendous possibilities in genomic research. In addition to providing access to samples that are hard to find in nature (as well as crafting genomic sequences not known to occur in the natural world), manufacturing DNA enables scientists to test any sequence in a wide variety of contexts and environments.

BioCAD/CAM software tools help researchers design sequences that can be critical to discovering new solutions for energy and the environment, but so far the software has not been able to automatically fix problematic sequences. This delays synthesis of the designed DNA and the subsequent transition to the manufacturing process.

To solve this problem, JGI researchers Ernst Oberortner, Jan-Fang-Cheng, Nathan Hillson, and Sam Deutsch developed a suite of build-optimization software tools (BOOST) to automate the synthetic DNA design process—and do away with the trial-and-error process scientists currently utilize to determine a sequence that can be synthesized. As reported December 6, 2016, in *ACS Synthetic Biology*, BOOST can automatically detect “difficult” sequences (of nucleotides) and redesign them for DNA synthesis, addressing DNA sequences with certain problematic characteristics (e.g., extreme %GC content, sequence patterns, and repeats), which decrease the success rate of DNA synthesis. By improving the design and manufacture of synthetic DNA through enhanced tools, scientists can accelerate gene discovery and gene characterization toward practical applications for energy and the environment. The BOOST suite of tools is available as a web application at <https://boost.jgi.doe.gov>.



*Linderina pennispora*, a fungus belonging to the earliest branches of the fungal kingdom. (ZyGoLife Research Consortium, Flickr, CC BY-SA 2.0)

## Finding a New Major Gene Expression Regulator in Fungi

Just four letters — A, C, T, and G — make up an organism’s genetic code. Changing a single letter, or base, can lead to changes in protein structures and functions, affecting an organism’s traits. In addition, though, subtler changes can and do happen, involving modifications of the DNA bases themselves. The best-known example of this kind of change is a methylation of the base cytosine at the fifth position on its carbon ring (5mC). In eukaryotes, a less well-known modification involves adding a methyl group to base 6 of adenine (6mA).

As reported May 8, 2017, in *Nature Genetics*, a team led by JGI scientists described the prevalence of 6mA modifications in the earliest branches of the fungal kingdom. Though fungi have existed for a billion years and collectively are capable of degrading nearly all naturally occurring polymers and even some human-made ones, most of the species that have been studied belong to just two phyla: the Ascomycota and Basidiomycota. The remaining six groups of fungi are classified as “early diverging lineages,” the earliest branches in fungal genealogy. They comprise a little-explored realm of fungi, providing a repertoire of important and valuable gene products for DOE bioenergy and environment missions.

“By and large, early-diverging fungi are very poorly understood compared to other lineages. However, many of these fungi turn out to be important in a variety of ways,” said study first author and JGI analyst Stephen Mondo.

Many of the fungal genomes used in the study were sequenced as part of the JGI’s 1000 Fungal Genomes initiative aimed at producing at least one reference genome for every family of fungi. For the study, the team used 16 fungal genomes sequenced at the JGI using the Pacific Biosciences sequencing platform. JGI scientists took advantage of this sequencing platform to explore epigenetic (5mC, 6mA) modifications. They discovered very high levels of 6mA in fungi, where up to 2.8% of all adenines were methylated, and confirmed these findings using multiple independent methods. The previous record holder for genomic 6mA, noted Mondo, is the alga *Chlamydomonas reinhardtii* (sequenced and annotated by the JGI), in which just 0.4% of adenines were methylated.

“This is one of the first direct comparisons of 6mA and 5mC in eukaryotes, and the first 6mA study across the fungal kingdom,” said JGI Fungal Program head and senior author Igor Grigoriev. “6mA has been shown to have different functions depending on the organism. For example, in animals it is involved in suppressing transposon activity, while in algae it is positively associated with gene expression. Our analysis has shown that 6mA modifications are associated with expressed genes and are preferentially deposited based on gene function and conservation, revealing 6mA as a marker of expression for important functionally-relevant genes. The discovery of DNA methylation in early diverging fungi helps the research community better understand regulation of genes that encode the parts for bio-based economy and bioenergy applications.”

The team found that, in addition to 6mA performing what seems to be the opposite role of 5mC (which suppresses expression), the presence of 5mC and 6mA are inversely correlated. Mondo described the identification and proposed role of 6mA in early diverging fungi at the 12th Annual Genomics of Energy and Environment Meeting. Watch his talk on the JGI's YouTube channel at <http://bit.ly/JGI2017Mondo>.

---

## New Technology to Access Microbial Dark Matter

New technologies that can be harnessed to illuminate so-called microbial “dark matter,” the majority of the planet’s microbial diversity that remains uncultivated, have many benefits. The genes and metabolic functions of these yet-undiscovered microbes enable them to live in a wide range of environments, and could have potential applications in fields ranging from bioenergy to biotechnology to environmental research.



Samples used to demonstrate the efficacy of the new technology were taken from hot springs at Yellowstone National Park. (Image by Paul Blainey, Christina Mork and Geoffrey Schiebinger)

More than 50,000 microbial genome sequences are in the JGI's Integrated Microbial Genomes & Microbiomes (IMG/M) publicly accessible database, and many of them have been uncovered through the use of metagenomic sequencing and single-cell genomics. Despite their utility, these techniques have limitations: single-cell genome amplifications are time-consuming and often incomplete, and shotgun metagenomics sequencing generally works best if the environmental sample is not too complex.

A team of Stanford University researchers reported the development of a microfluidics-based, mini-metagenomics approach to mitigate these challenges on July 5, 2017, in *eLife*. The technique starts with reducing the environmental sample's complexity by separating it using microfluidics into 96 subsamples of five to 10 cells. Then the genomes in the few cells in each subsample are amplified, and libraries are created for sequencing of these mini-metagenomes. The smaller subsamples can be held to single-cell resolution for statistical analyses. Co-occurrence patterns from many subsamples can also be used to perform sequence independent genome binning. The technology was developed through resources provided by the JGI's Emerging Technologies Opportunity Program (ETOP), which was launched in 2013.

The team validated the technique using a synthetic microbial community, and then applied it to samples from the Bijah and Mound hot springs at Yellowstone National Park. Among the team's findings was that the microbes at Mound Spring had higher potential to produce methane than the microbes from Bijah Spring. The team also identified a microbial genome from Bijah Spring that could reduce nitrite to nitrogen. Applying this new technology to additional sample sites will add to the range of hitherto uncharacterized microbial capabilities with potential DOE mission applicability.

---

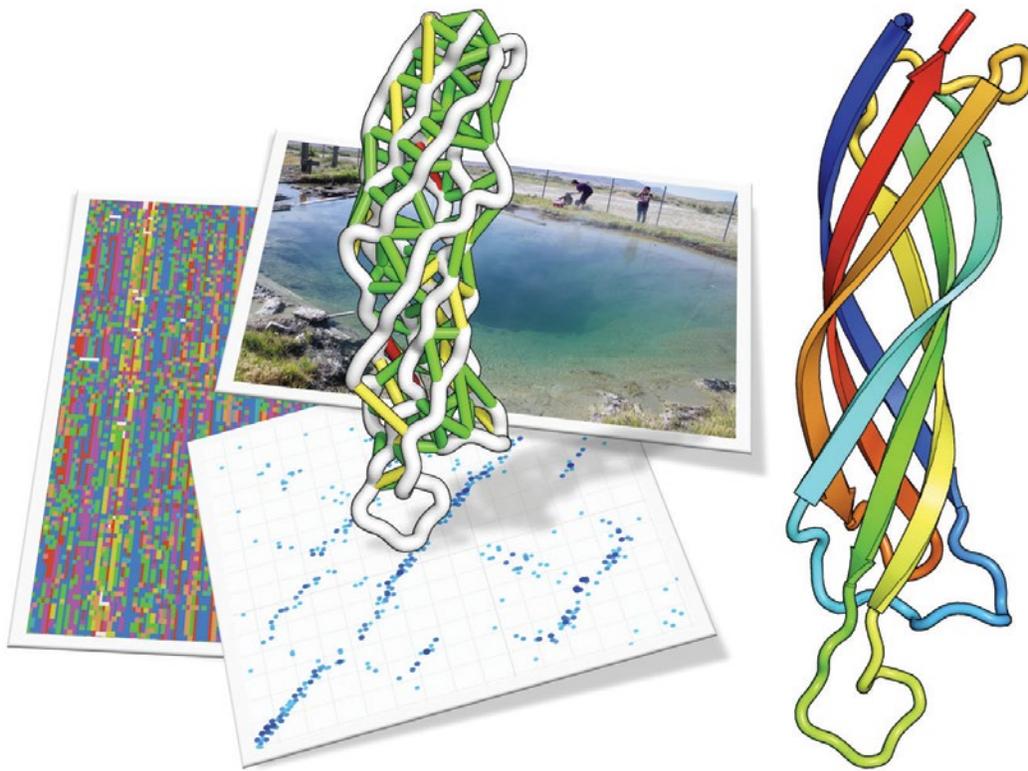
## IMG/M Database Helps Fill Previously Unknown Protein Structures

Proteins largely form a cell's structures and carry out its functions. They control growth and influence mobility, serve as catalysts, and transport or store other molecules. On paper, the one-dimensional amino acid sequence may seem meaningless, but when viewed in three dimensions, a protein's structure, and particularly the way it folds, determines its functions.

The Pfam database contains close to 15,000 protein families: groups of proteins that share an evolutionary origin. For nearly a third of these protein families, at least one protein in each family already has an experimentally determined structure. For another third of the protein families, comparative models could be built with some degree of confidence. For the final third of the protein families in the database, however, no structural information exists.

In the January 20, 2017, issue of *Science*, a team led by the University of Washington's David Baker collaborated with JGI researchers to report that structural models have been generated for 614, or 12%, of the protein families that had previously had no structural information available. This accomplishment was made possible through a collaboration in which the Baker lab's protein structure prediction server, Rosetta, analyzed the metagenomic sequences publicly available on the JGI's IMG/M system. "That this could be accomplished using computational modeling methods was not at all apparent five years ago," the team noted in its paper.

Armed with genome sequences, researchers like Baker have been able to identify sets of amino acids that evolve simultaneously, even though they are nowhere near each other on the unfolded chain. JGI Prokaryote Super Program head Nikos Kyrpides said the collaboration between the Baker lab and the JGI enabled the team to come up with a powerful way of predicting structures and structural alignments. "Such efforts were previously restricted on protein families generated from sequences found on the isolate genome only. These genomes comprise about 200 million sequences. As expected, by harnessing the 5 billion assembled metagenome sequences available on our IMG/M database, we were able to dramatically increase the coverage of many of the known protein families. Efforts like this one heavily depend on the availability of assembled metagenomics sequences, which is an advantage the JGI brings to the table with our high-quality assemblies. The application of such tools on our data provides a great example of how the larger community can utilize JGI resources for discovery."



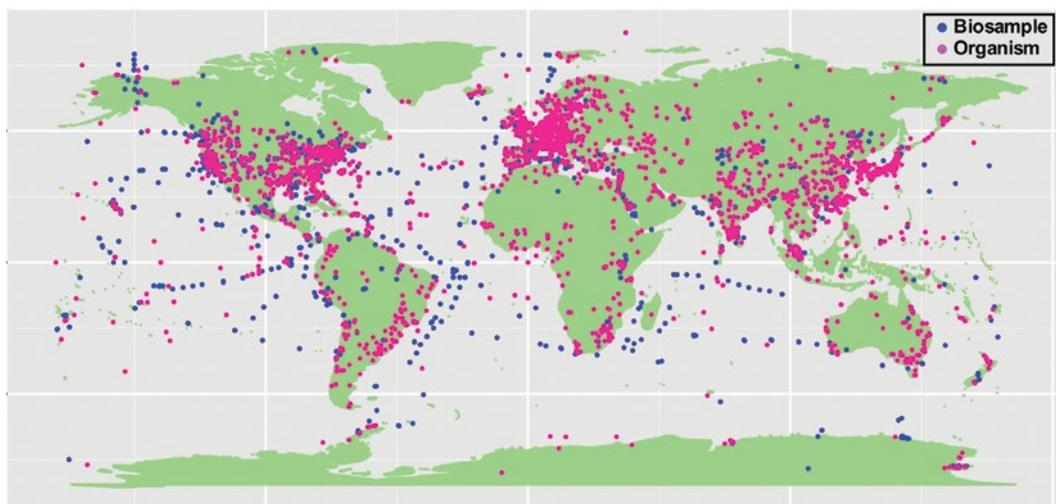
From sample to structure. **Top:** Researchers gathering samples from Great Boiling Spring in Nevada. **Left:** A snapshot of aligned metagenomic sequences. Each row is a different sequence (the different colors are the different amino acid groups). Each position (or column) is compared to all other positions to detect patterns of co-evolution. **Bottom:** The strength of the top co-evolving residues is shown as blue dots, and are represented as colored lines on the structure above. The goal is to make a structure with as many of these contacts as possible. **Right:** A cartoon of the protein structure predicted. The protein domain shown is from Pfam DUF3794; this domain is part of a Spore coat assembly protein SafA. (Image of Great Boiling Spring by Brian Hedlund, UNLV. Protein structure and composite image by Sergey Ovchinnikov, UW.)

## IMG/M Platform Debuts Database of Global Viral Diversity

Providing high-quality, publicly accessible sequence data goes hand-in-hand with developing and maintaining the databases and tools that the research community can employ to help answer scientific questions. In the "Database" issue of the journal *Nucleic Acids Research*, released January 1, 2017, JGI researchers reported on the latest updates to several publicly accessible databases and computational tools that benefit the global community of microbial researchers. One report focused on a new database dedicated to global viral diversity called IMG/VR (<https://img.jgi.doe.gov/vr/>), which followed on the heels of a recent JGI viral diversity study in *Nature*.

Additional articles in the same issue describe updates to several publicly accessible, interactive databases since the last set of reports was published in 2014. For example, as of July 2016, 47,516 archaeal, bacterial, and eukaryotic genomes were in the Integrated Microbial Genomes with Microbiome Samples (IMG/M: <https://img.jgi.doe.gov/m/>) system, with researchers noting that number "represents an over 300% increase since September 2013." Another paper concerns the Genomes Online Database (GOLD: <https://gold.jgi.doe.gov/>), a manually curated data management system that catalogs sequencing projects with associated metadata from around the world. A fourth paper focuses on the Integrated Microbial Genomes Atlas of Biosynthetic gene

Geographic distribution of GOLD biosamples and organisms with organism location of isolation marked in pink while biosample location of collection is denoted with blue dots. (Image from Mukherjee et al. *Nucl. Acids Res.* 2016;nar.gkw992)



Clusters (IMG-ABC: <https://img.jgi.doe.gov/abc/>). Launched in 2015, IMG-ABC allows researchers to search for biosynthetic gene clusters and secondary metabolites. The latest update now incorporates ClusterScout, a tool for targeted identification of custom biosynthetic gene clusters across several thousand isolate microbial genomes, and a new search capability.

Microbes play key roles in maintaining the planet's biogeochemical cycles. Viruses, thought to outnumber microbes by 10-fold, exert major influences on microbial survival and community interactions. Advances in sequencing technologies have generated vast amounts of data about viruses, requiring tools to manage and interpret the information. These updates focus on database analytical tools for microbial genomics and viruses relevant to DOE bioenergy and environment missions.

## Minimum Information Standards for Uncultivated Microbial Genomes

During the Industrial Revolution, the establishment of standards allowed builders to produce supplies in bulk, maintain production quality, and fuel interstate commerce. The importance of standards is dramatically illustrated when they don't exist or are not commonly accepted.

More than a century after the Industrial Revolution, advances in DNA sequencing technologies have caused similarly dramatic shifts in scientific research. One aspect is studying the planet's biodiversity. An international team led by JGI researchers has developed standards for the minimum metadata to be supplied with single amplified genomes (SAGs) and metagenome-assembled genomes (MAGs) submitted to public databases. The team's work was published August 8, 2017, in *Nature Biotechnology*.

"Over the last several years, single-cell genomics has become a popular tool to complement metagenomics," said study senior author Tanja Woyke, head of the JGI Microbial Program. "Starting in 2007, the first single-cell genomes from environmental cells appeared in public databases, and they are draft assemblies with fluctuations in the data quality. Metagenome-assembled genomes have similar quality challenges. For researchers who want to conduct comparative analyses, it's really important to know what goes into the analysis. Robust comparative genomics relies on extensive and correct metadata."

Microbes play crucial roles in regulating global cycles involving carbon, nitrogen, and phosphorus among others, but many of these microorganisms remain uncultured and unknown. Learning more about this "microbial dark matter" involves sequencing metagenomes or the amplified DNA of single cells, then bioinformatically

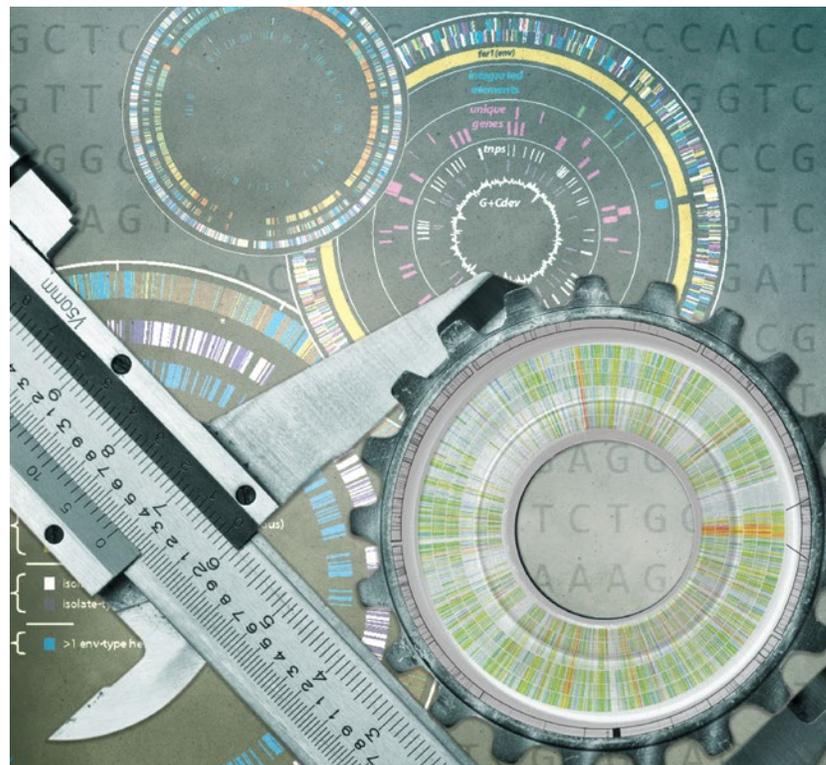
identifying genomes. As genomic data production has ramped up over the past two decades and is being generated on various platforms around the world, scientists have worked together to establish definitions for terms such as “draft assembly” and data collection standards that apply across the board.

In their paper, Woyke and her colleagues proposed four categories of genome quality. Low-Quality Drafts would be less than 50% complete, with minimal review of the assembled fragments and less than 10% contaminated with non-target sequence. Medium-Quality Drafts would be at least 50% complete, with minimal review of the assembled fragments and less than 10% contamination. High-Quality Drafts would be more than 90% complete with the presence of the 23S, 16S, and 5S rRNA genes, as well as at least 18 tRNAs, and with less than 5% contamination. The Finished Quality category is reserved for single contiguous sequences without gaps and less than 1 error per 100,000 base pairs.

Moving from a proposal in print to implementation requires community buy-in. Woyke and her colleagues conceived of the minimum metadata requirements for SAGs and MAGs as extensions to existing metadata standards for sequence data, referred to as “MIxS,” developed and implemented by the Genomic Standards Consortium (GSC) in 2011. The GSC is an open-membership working body that ensures the research community and the data repositories are engaged in the standards development process. By working directly with these groups, the GSC can assist both large-scale data submitters and databases to align with the MIxS standard and submit compliant data.

“I think it helps that the people developing standards are the people conducting the studies,” said GSC President and study co-author Lynn Schriml of the Institute of Genome Sciences at the University of Maryland School of Medicine. “We have a vested interest in the data. Developing these novel metadata standards enables researchers to consistently report the most critical metadata for analysis. Capturing data using controlled vocabularies facilitates data consistency, thus making the databases richer and reusable.”

As the number of single-cell genomes and metagenome-assembled genomes in public databases continues to rise, developing minimum metadata standards enables researchers to consistently perform robust comparative analyses. (Zosia Rostomian, Berkeley Lab Creative Services)



## Mining Metagenomics Data for New CRISPR-Cas Systems

In microbes, CRISPR-Cas systems grant adaptive immunity, and these gene-editing tools are the bases of versatile technologies revolutionizing research. Thus far, CRISPR-Cas technology has been based only on systems from isolated bacteria. In a study led by longtime JGI collaborator Jill Banfield of UC Berkeley and Berkeley Lab's Earth & Environmental Sciences Area (EESA) published in *Nature* on December 16, 2016, researchers reported discovering a CRISPR-Cas9 system in archaea, as well as two previously unknown simple CRISPR-Cas systems in uncultivable bacteria. To identify these new systems, Banfield and her colleagues harnessed more than a decade's worth of metagenomic data from samples sequenced and analyzed by the JGI.

Microbes play key roles in the planet's cycles, and characterizing them helps researchers work towards solutions for energy and environmental challenges. Examining environmental microbial communities has provided access to an unprecedented diversity of genomes and CRISPR-Cas systems that have many applications, including biological research. By using the combined computational-experimental approach that was successful in this study, nearly all environments where life exists can be investigated. The researchers analyzed 155 million protein coding genes from uncultivated microbial communities, leading to the novel discoveries. The CasX and CasY proteins were found in bacteria from groundwater and sediment samples. The archaeal Cas9 were identified in samples taken from the Iron Mountain Mine Superfund site as part of Banfield's pioneering metagenomics work with the JGI. Both CasX and CasY are among some of the most compact systems ever identified. Banfield's team noted that this application of metagenomics gives a significant validation to studies of CRISPR-Cas proteins using living organisms.



The newly discovered CRISPR-CasY system was found in bacteria from deep underground at Utah's Crystal Geyser.

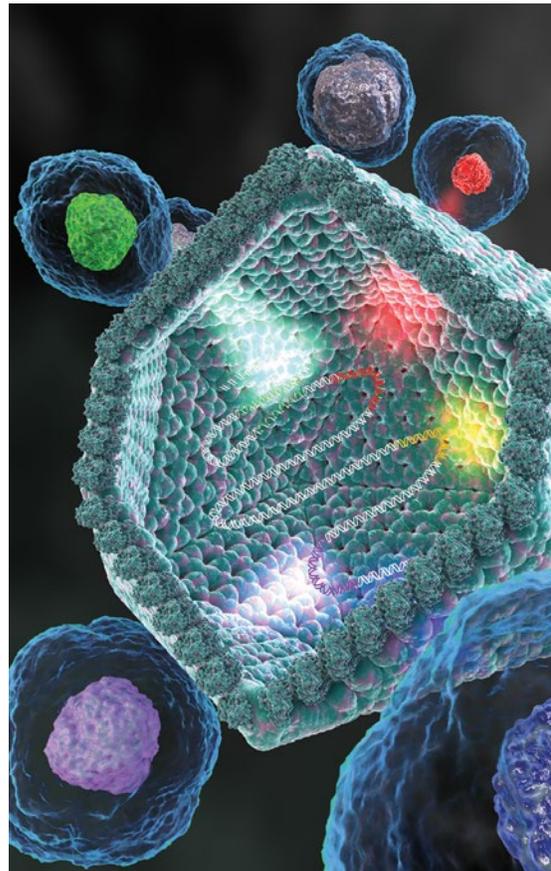
## Novel Group of Giant Viruses Discovered

Viruses are everywhere; their population is estimated to surpass the number of stars in the Milky Way. Giant viruses are characterized by disproportionately large genomes and virions that house the viruses' genetic material. They can encode several genes potentially involved in protein biosynthesis, a unique feature which has led to diverging hypotheses about the origins of these viruses.

As reported April 7, 2017, in *Science* by a JGI-led team that included collaborators at the National Institutes of Health, the University of Vienna, and CalTech, the discovery of a novel group of giant viruses (dubbed "Klosneuviruses") with a more complete set of translation machinery genes than any other virus known to date lends insights into viral evolution.

The predicted hosts for the Klosneuviruses are protists: single-celled microorganisms. And while their direct impacts on protists are not yet worked out, these giant viruses are thought to have a significant impact on these protists that help regulate the planet's biogeochemical cycles.

"The discovery presents virus evolution for us in new ways, vastly expanding our understanding of how many essential host genes viruses can capture during their evolution," said National Institutes of Health evolutionary and computational biologist Eugene Koonin, a study co-author whose lab collaborated with the JGI on analyzing the Klosneuvirus genome. "Since protein synthesis is one of the most prominent hallmarks of cellular life, it shows that these new viruses are more 'cell-like' than any virus anyone has ever seen before."



Depiction of giant virus acquiring genes from different eukaryotic host cells. (Ella Maru studio, [www.scientific-illustrations.com](http://www.scientific-illustrations.com))

Scientists have been fascinated by giant viruses since 2003, when a group of French biologists discovered the Mimiviruses. Since then, a handful of other giant virus groups have been found. Giant viruses have the unique ability to encode proteins involved in translation (typically DNA to RNA to protein), piquing researchers' interests as to the giant viruses' origin. Two evolutionary hypotheses have emerged; one posits that giant viruses evolved from an ancient cell, perhaps one from an extinct fourth domain of cellular life. Another — a scenario championed by Koonin — presents the idea that giant viruses arose from smaller viruses.

The discovery of Klosneuvirus supports the latter idea, according to Tanja Woyke, JGI Microbial Program lead and senior author of the paper. "In this scenario, a smaller virus infected different eukaryote hosts and picked up genes encoding translational machinery components from independent sources over long periods of time through piecemeal acquisition," she said.

Frederik Schulz spoke about the team's discovery of the novel group of giant viruses at the 12th Annual Genomics of Energy and Environment Meeting. Watch his talk at <http://bit.ly/JGI2017Schulz>.

Sorghum is of interest to the DOE as a bioenergy crop because of its drought resistance and ability to thrive on marginal lands. JGI sequenced the sorghum reference genome (*Sorghum bicolor*), publishing it in 2009. The Epigenetic Control of Drought Response in Sorghum (EPICON), a five-year project funded by the DOE and involving JGI researchers, focuses on understanding how sorghum, a grass related to corn, thrives on limited resources, including water.



# Bioenergy

## A Gene that Influences Grain Yields in Grasses

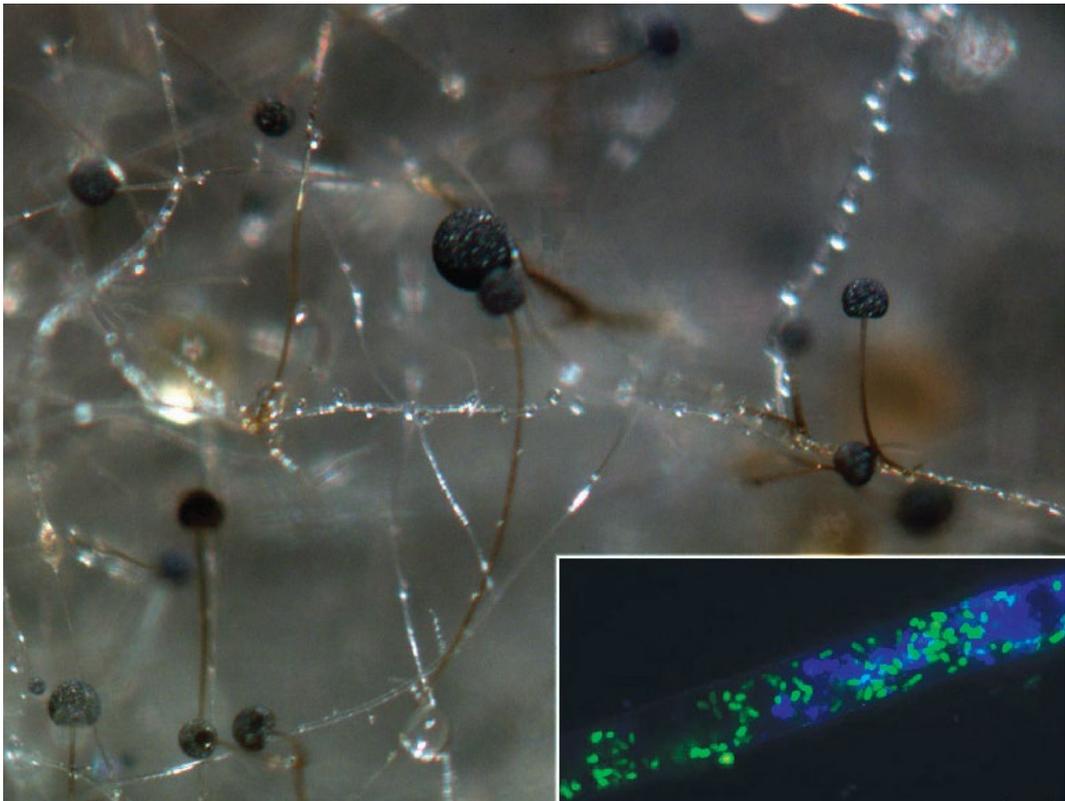
Gene discovery in maize, an important food and bioenergy crop, has been limited due to its large and complex genome. *Setaria viridis* has been proposed as a model crop for maize; *Setaria* species, among them green foxtail (*S. viridis*) and foxtail millet (*S. italica*), are related to several candidate bioenergy grasses including switchgrass and Miscanthus. They serve as model systems to study grasses that photosynthetically fix carbon from CO<sub>2</sub> through a water-conserving (C<sub>4</sub>) pathway. The genomes of both green foxtail and foxtail millet were sequenced and annotated through the JGI's CSP.

A team led by the Donald Danforth Plant Science Center's Tom Brutnell and including JGI researchers reported that it had identified genes that may play a role in flower development on the panicle of green foxtail. The panicle is the spear-shaped flowering cluster at the tip of each branch necessary for reproduction. The team's work was published April 18, 2017, in *Nature Plants*.

The team identified four sparse panicle recessive mutants, tagged *spp1* through *spp4*, that lead to panicles with reduced and uneven flower clusters. Focusing on the *spp1* mutation, they performed deep sequencing to specifically locate the genes that cause the mutation, narrowing their search down to a 1-million base sequence. They ultimately identified the *SvAUX1* gene in green foxtail as one critical for flower cluster development in green foxtail. Panicle development is critical for determining grain yield that is crucial to food crops as well as candidate crops for producing renewable and sustainable fuels. A homologous gene in maize was identified as playing a similar role, illustrating the value of model systems in finding genes involved in important properties in potential bioenergy-relevant plants.



Hyphae and asexual sporangia of *Rhizopus microsporus* (photo by Stephen Mondo). Inset shows a hypha with YFP-labeled *Burkholderia endobacteria* in green and DAPI-stained fungal nuclei in blue; scale bar, 5  $\mu\text{m}$ . (photo by Olga Lastovetsky)



## Mutualism and Lipid Metabolism in Fungi

To answer the challenge of producing renewable, sustainable alternative fuels, researchers aren't just looking at developing candidate bioenergy crops but are also reviewing other natural sources of energy-dense oils, such as fungi. To learn more about how bacteria interact with fungi in a symbiotic relationship to support the biochemical processes that could contribute to the development of alternate fuel sources, Cornell and JGI researchers used a model bacterial-fungal system to reveal the mechanism for lipid production in oil-producing, or oleaginous, fungi.

The widespread use of antimicrobials has led to the perception that fungi and bacteria are antagonists, but this is not always true. Researchers at Cornell University and the JGI reported on the mechanisms influencing the mutualistic relationship between the fungal plant pathogen *Rhizopus microsporus* and a *Burkholderia* endosymbiont in the *Proceedings of the National Academy of Sciences* on December 12, 2016. The work highlighted the complicated nature of fungal-bacterial relationships that can oscillate between mutualistic and antagonistic.

*Rhizopus* is a plant pathogen of crops including rice, sunflower, and maize, and it relies on the toxin produced by the endosymbiont bacteria. This symbiotic relationship was discovered in 2005, and researchers expect more examples of fungal-bacterial mutualisms to be found now that they are being actively sought out. The Cornell team cultivated the fungi and isolated the endobacteria while the JGI team sequenced, assembled, and annotated the host and non-host genomes as part of the JGI's 1000 Fungal Genomes project and other JGI Community Science Projects. The results indicate that the transition between symbiosis and antagonism between the bacteria and their fungal host is driven by a novel mechanism involving fungal lipid metabolism. "When lipid metabolism in the fungus is triggered and phosphatidic acid (PA)-producing enzymes are activated, the fungus and bacteria act as mutualists. When the PA-producing enzymes are inhibited, the bacteria and fungi act as antagonists," wrote Cornell's Teresa Pawlowska and her graduate student Olga Lastovetsky.

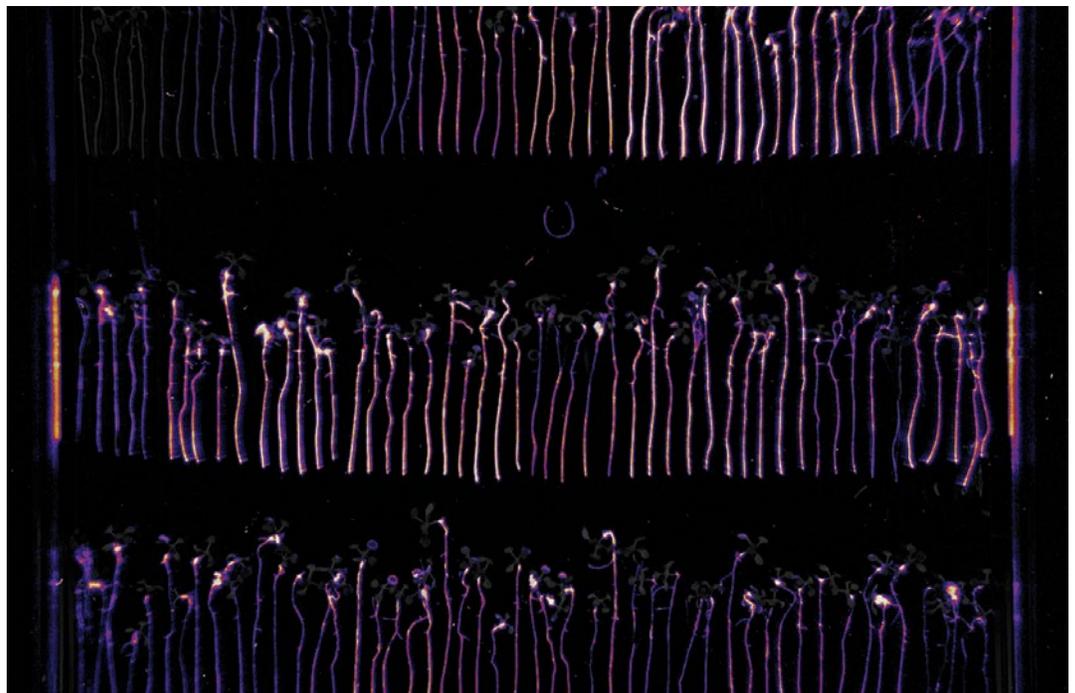
## A Genome-Wide Map of Bacterial Genes

A plant's health and development are influenced by microbes in the soil and in the rhizosphere where the plant roots interact with the soil, and the endophytes residing within the plant. By identifying the bacterial genes that alter how well microbes colonize a plant, researchers can develop targeted approaches to improve plant health and growth for a number of applications, including increased biomass yield for biofuel production. To better understand how microbes colonize the root environment, researchers from JGI and a team led by Jeff Dangl of the Howard Hughes Medical Institute at the University of North Carolina at Chapel Hill applied a genome-wide transposon mutagenesis approach on the model plant growth-promoting bacterium *Pseudomonas simiae*. The team used the model plant *Arabidopsis thaliana* as a host to generate a genome-wide map of bacterial genes that affect the efficacy of microbial colonization.

One of the principal challenges arising from rapid sequencing is the assignment of functions to new genes. The randomly barcoded transposon sequencing (RB TnSeq) approach used here can accelerate the association of new genes with characteristics and behaviors of importance to DOE missions, such as understanding how microbes help (or hinder) the growth of crops that could serve as bioenergy feedstocks. As reported September 22, 2017, in *Plos Biology*, they used RB Tn-Seq to identify 115 genes that, when mutated, cause reduced root colonization capabilities. These genes are involved in functions such as sugar metabolism, cell wall synthesis, and motility. The team also identified 243 genes that, when mutated, positively alter root colonization capabilities, many of which are likely involved in amino acid transport and metabolism. From these nearly 360 genes, the team identified a subset of 43 genes to which very little or no functional information could be assigned. The researchers suggested that these genes may represent novel functions or pathways yet to be characterized. The work shows that RB-TnSeq can be applied to assess *in vivo* bacterial plant root colonization.

Among the contributors to this project was Sabah Ul-Hasan, a 2015 intern through the JGI/University of California (UC), Merced Genomics Distinguished Graduate Internship Program. The program offers UC Merced graduate students hands-on experience in cutting-edge genome research as part of the JGI's commitment to training the next generation of scientific talent.

A bioluminescent assay helped researchers visually quantify the colonization ability of *P. simiae* mutant strains identified by the RB-TnSeq screen. (Benjamin Cole)

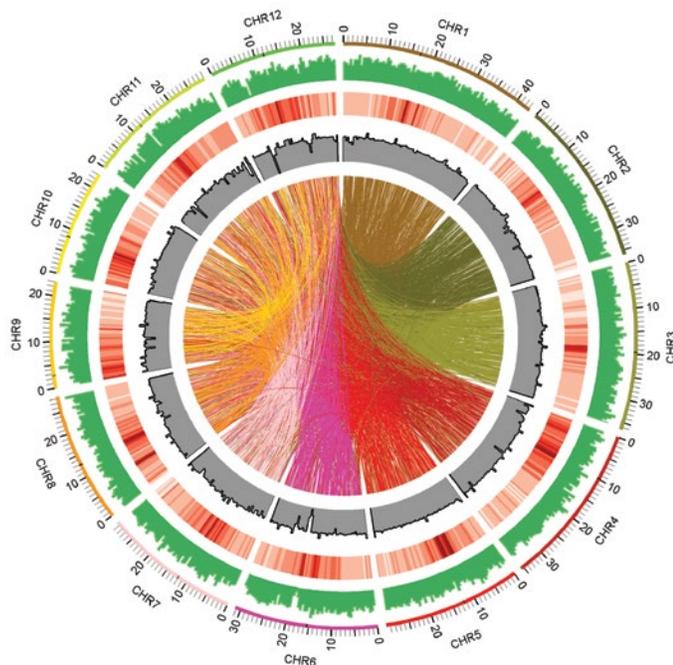


## Mutant Rice Database for Bioenergy Research

For more than half of the world's population, rice is the primary staple crop. As a grass, it is a close relative of the candidate bioenergy feedstock switchgrass. Boosting yields of bioenergy feedstock crops such as grasses requires a better understanding of how enzymes and proteins synthesize plant cell walls to modify the processes and the composition. Until now, mutant collections for grass models have lagged behind those available for the *Arabidopsis* model system. In fast-neutron irradiation, exposure to high-energy neutrons induces a wide variety of mutations by making changes in DNA. Using this approach, rice researchers created the first major, large-scale collection of mutations for grass models.

As reported June 2017 in *The Plant Cell*, a team led by Pam Ronald of UC Davis, and including researchers at the JGI and the Joint BioEnergy Institute (JBEI), a DOE Bioenergy Research Center, have assembled the first major large-scale collection of mutations for grass models. Their goal is to have a functional genomics resource for grass models involved in plant cell wall biosynthesis studies.

Genome-wide distribution of fast neutron-induced mutations in the Kitaake rice mutant population. (Guotian Li and Rashmi Jain)



The team used the model rice cultivar Kitaake (*Oryza sativa* L. ssp. *japonica*), and compared the genes against the reference rice genome of another *japonica* subspecies called Nipponbare available on the JGI Plant Portal Phytozome. Through fast-neutron irradiation, the time-consuming procedures involving plant transformation or tissue culture were bypassed, allowing for faster development of rice mutant collections. Resequencing the 1,504 mutants has allowed researchers to identify structural variants and mutations, providing an invaluable resource for grass models being used to improve candidate bioenergy feedstock crops such as switchgrass.

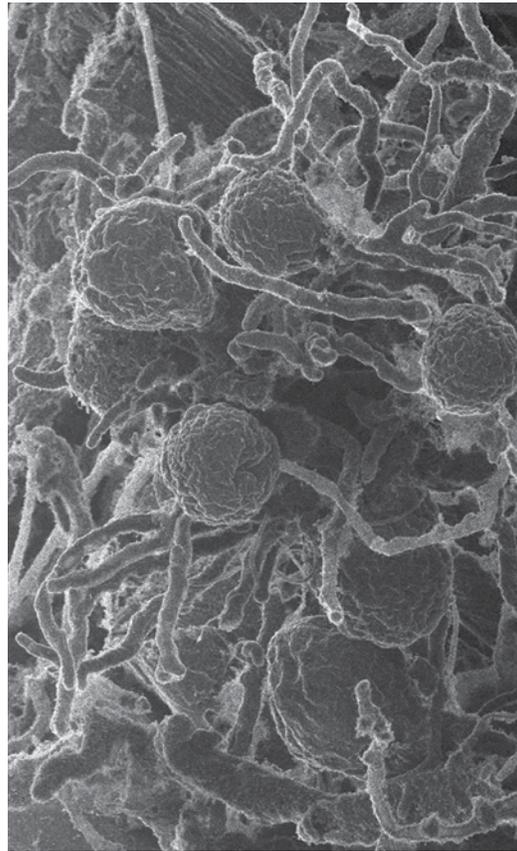
Information on this new, large-scale collection of more than 90,000 mutations affecting nearly 60% of all rice genes is available on the publicly accessible database KitBase.

## Fungal Enzyme Clusters Better at Attacking Biomass

One of the biggest barriers to commercial production of sustainable biofuels is cost-effectively breaking down the bioenergy crops into sugars that can then be converted into fuel. To reduce this barrier, bioenergy researchers are looking to nature and the estimated 1.5 million species of fungi that, collectively, can break down almost any substance on earth, including plant biomass.

As reported May 26, 2017, in *Nature Microbiology*, a team led by researchers at UC Santa Barbara, found for the first time that early lineages of fungi can form complexes of enzymes capable of degrading plant biomass. Consolidating these enzymes, in effect into protein assembly lines, makes them team up to work more efficiently than they would as individuals.

Scanning electron micrograph (SEM) of *Neocallimastix californiae*. (Chuck Smallwood, PNNL)



“There are protein complexes in bacteria called cellulosomes that pack together the enzymes to break down plant biomass,” said study senior author Michelle O’Malley of UC Santa Barbara. “The idea is that these clusters are better at attacking biomass because they are keeping the different enzymes in place with plugs called dockerins, so they work more efficiently. This has been detailed in bacteria for more than 20 years, but now seen for the first time in fungi!”

The work was enabled by harnessing the capabilities of two DOE Office of Science User Facilities: the JGI at Berkeley Lab, and the Environmental Molecular Sciences Laboratory (EMSL) at Pacific Northwest National Laboratory (PNNL). With help from both user facilities through the Facilities Integrating Collaborations for User Science (FICUS) initiative, the team has now found protein complexes in anaerobic gut fungi that O’Malley said in principle do the same thing: attack plant biomass as a cluster of enzymes. While the team found that many of the enzymes in these complexes resulted from horizontal gene transfers with gut bacteria, it also noted differences in the composition compared with the bacterial cellulosomes. For one thing, dockerins and scaffoldin are not similar between fungi and bacteria. Also, the bacterial cellulosomes are species-specific.

The study involved a comparative genomics analyses of five fungi that belong to the Neocallimastigomycetes, a clade of the early-diverging lineages that are not well-studied. Three of the fungi were isolated from animal gut samples collected by the UC Santa Barbara team and sequenced and annotated by the JGI. Study co-senior author and JGI Fungal Program head Igor Grigoriev noted that a multi-omics approach that harnessed the genomics and molecular characterization capabilities through the FICUS collaborative science initiative was critical for the research.

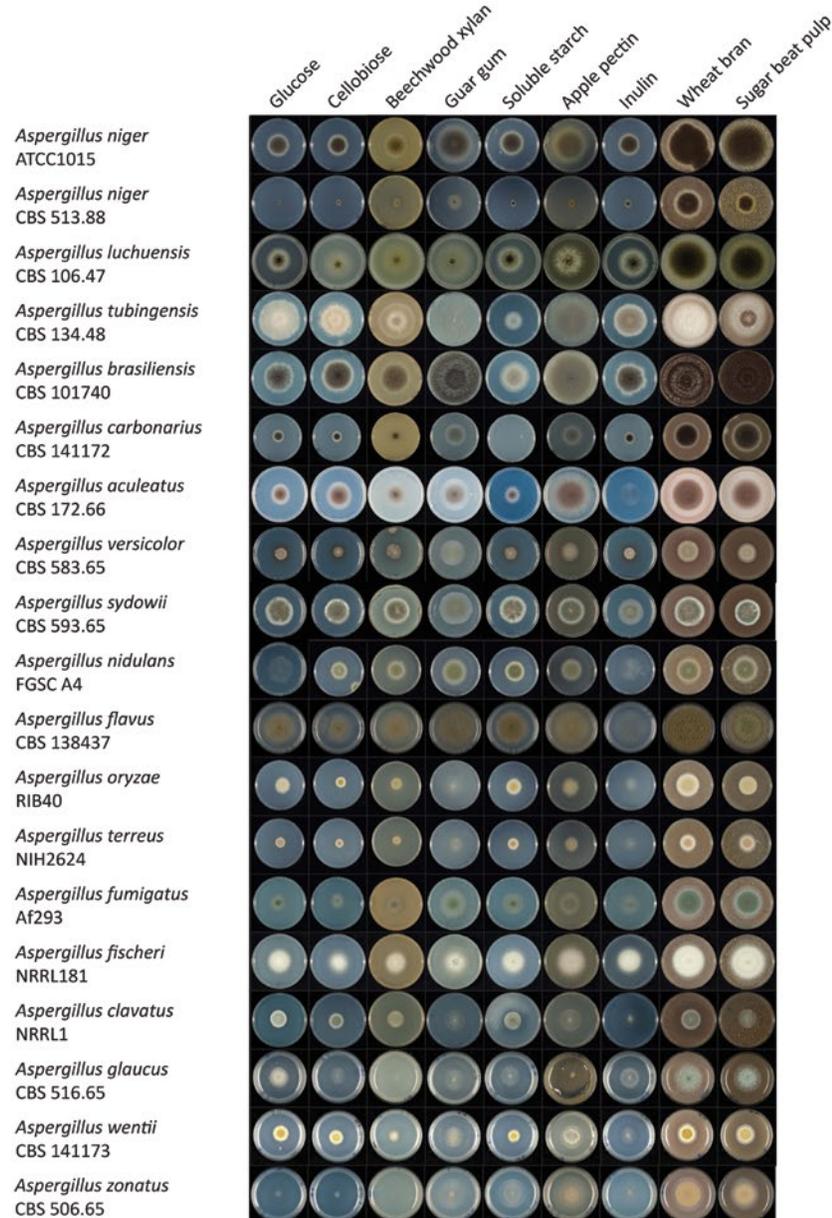
“It’s the first time we’ve seen parts of the fungal cellulosome,” Grigoriev said. “Through the JGI-EMSL FICUS initiative, proteomics allowed us to find the first of these really large ~700 kiloDaltons (kDa) fungal proteins that hold all enzymes together, compared to the molecular weight of 34 kDa of an average protein. Then the high quality of genome assemblies enabled identification of multiple copies of this protein in each of the gut fungi genomes. Just having proteomics or sequencing tools isn’t enough, since these proteins are not similar to anything else outside of Neocallimastigomycetes. Though the fungal cellulosome was discovered through proteomics, we needed genomics and transcriptomics to decode all its parts.”

The work is an extension of O’Malley’s studies of anaerobic gut fungi, which appeared in *Science*. “It’s a lot of the same players, but we’re digging deeper now because we have high-resolution genomes, and we didn’t have them then,” she said. “We’re able to conduct more comparative genetics and now we’re trying to figure out the ecological roles in their microbiome.” Watch her talk on this topic from the 2016 Genomics of Energy and Environment Meeting at [bit.ly/JGI2016OMalley](http://bit.ly/JGI2016OMalley).

## A Genus-Wide View of *Aspergillus* Diversity

*Aspergillus niger* has been used for decades to produce citric acid — a compound frequently added to foods and pharmaceuticals — through fermentation at an industrial scale. Other *Aspergillus* species play critical roles in biofuel production and plant and human health. Since the majority of *Aspergillus* fungi's 350 species have yet to be sequenced and analyzed, researchers are still at the tip of the iceberg when it comes to understanding the full spectrum of useful compounds the fungi may generate.

As reported February 14, 2017, in *Genome Biology*, an international team including JGI researchers sequenced the genomes of 10 novel *Aspergillus* species, more than doubling the number of *Aspergillus* species sequenced to date. The newly sequenced genomes were compared with the eight other sequenced *Aspergillus* species. With this first ever genus-wide view, the international consortium found that *Aspergillus* has a greater genomic and functional diversity than previously understood, broadening the range of potential applications for the fungi considered one of the most important workhorses in biotechnology.



Comparative growth of *Aspergillus* spp. (Ad Wiebenga & Ronald de Vries, Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands)

Conducted through the JGI's CSP, the study demonstrates the importance of evaluating biodiversity within a genus to understand how fungi can be harnessed to solve a variety of problems. Sequencing a diverse set of *Aspergillus* genomes allows researchers to build a more comprehensive catalog of enzymes for biotechnological applications, added JGI Fungal Program Head Igor Grigoriev, senior author of the paper. Those applications include harnessing *Aspergillus* to help protect crops and ward off agents that can cause diseases in plants. Comparing the newly sequenced genomes to those already available, researchers found a huge variety of carbohydrate-active enzymes (CAZymes) among the *Aspergillus* species, suggesting distinct strategies to break down plant biomass. CAZymes are responsible for breaking down plant cell walls, which is useful for industrially processing plants the DOE considers candidate bioenergy crops. The sugars that are part of these cell walls can't be accessed and fermented to make biofuels unless the walls are broken down by agents like CAZymes.

"Several *Aspergillus* species have already established status as cell factories for enzymes and metabolites," said study lead author Ronald de Vries of the Westerdijk Fungal Biodiversity Institute in the Netherlands. But despite the new insights gained, he emphasized that much still remains unknown about the full spectrum of what *Aspergillus* can do. "There is [still] much to learn and get from a better study [of *Aspergillus*]," he said. "The potential for applications within the genus has barely been touched."

## Nitrogen Uptake Between Fungi and Orchids



Long-lipped serapias.  
(Ziegler175, CC BY-SA 3.0)

Orchids, like the majority of terrestrial plants, form symbiotic relationships between their plant roots and soil fungi, known as mycorrhizal association. However, unlike other terrestrial plants, orchids rely on their mycorrhizal fungal partners for nutrient supply during the seed germination and development stages. Following this stage, most orchid species develop leaves and are capable of self-nourishment, whereas some species continue to rely on their fungal partners for an organic carbon supply. As reported in the January 2017 issue of *New Phytologist*, a team led by University of Turin researchers investigated the orchid mycorrhizal fungus *Tulasnella calospora* as both a free-living mycelium and in symbiosis with the photosynthetic orchid long-lipped serapias, or *Serapias vomeracea*.

For the first time, researchers looked at the fungal genes that may have been involved in both the uptake and transfer of nitrogen to the host plant. RNA sequencing for the project was performed at the JGI as part of the 2013 CSP portfolio. The team also used the JGI fungal genome database MycoCosm to identify fungal genes coding for proteins that were involved in nitrogen uptake and transfer. They found that the *T. calospora* genome has two genes coding for ammonium transporters and several genes coding for amino acid transporters, proteins that play roles in the nitrogen nutrient pathway.

This study provided a model system amenable to experimental manipulation for plant-fungi nutrient exchanges on a symbiotic level and offers insights into how host plants benefit from the mutualistic relationships formed with soil fungi, which can expand their habitat range. Understanding these vital relationships may shed light on microbial symbioses applicable to growing bioenergy feedstock plants. Overall, the orchid mycorrhizal fungi's use of nitrogen may broaden the habitat ranges of orchids, allowing them to grow in a variety of soil types.

## Determining a Driving Force in Speciation

In a hybrid zone on the Northern Rocky Mountains, the flowering mustard plant *Boechea stricta* is evolving into a fitter species better adapted to its local environment. As reported April 3, 2017, in *Nature Ecology & Evolution*, a team including JGI researchers analyzed the mechanisms by which the *B. stricta* plants experienced positive directional selection.

As part of both the JGI's CSP and Emerging Technologies Opportunity Program (ETOP), the researchers sequenced and analyzed the genome of *B. stricta*, a relative of the model plant *Arabidopsis*. With the genome in hand, they used techniques including gene mapping and chromosome painting methods to identify a major chromosomal inversion that controls ecologically important traits in the plant. They tested for QTLs, genome regions on chromosomes to which genetic traits can be mapped, in *B. stricta*'s chromosomal inversion. The researchers found several linked QTLs that changed ecologically important characteristics of the plant, such as flowering time and plant size, enabling it to adapt to its local environment, which in turn increased its fitness. The study shows that inversions that capture beneficial alleles such as QTLs can accelerate through the population and produce speciation.

This study offers the first direct evidence showing that QTLs are a driving force behind speciation. The knowledge gained from this study, said senior author and JGI collaborator Thomas Mitchell-Olds of Duke University, "gives evolutionary biologists experimental evidence showing how chromosomal changes contribute to adaptation and speciation. Furthermore, the genome sequence will help us understand how *Boechea* species are able to reproduce asexually by seeds, a process that can be used by farmers to speed up crop improvement practices."

Researchers used snow-filled garbage cans to water a test population planted in a mountain-top meadow near the Lost Trail Pass ski resort in the mountains of Montana throughout the summer. (Tom Mitchell-Olds)



## The Secrets Behind Grasses' Stomatal Success



John Vogel with *Brachypodium* plants. (Roy Kaltschmidt, Berkeley Lab)

Sequenced and annotated as part of the JGI's CSP, *Brachypodium distachyon* is a model grass for candidate bioenergy grasses such as *Miscanthus* and switchgrass. To further accelerate research in the development of biofuel feedstocks, a project to sequence thousands of *B. distachyon* mutants was selected for the 2015 CSP portfolio. This library of sequenced mutants will aid researchers to study and rapidly identify and order plants with mutations in any gene in their genomes.

Plants have evolved two kidney-shaped guard cells that swell to create the stomate. In grasses, however, the stomate have further evolved with the addition of two subsidiary cells flanking the guard cells, which may be linked to improved stomatal physiology. Unique to grasses, subsidiary cells have been associated with an improved physiological performance by allowing a greater range of pore size and quicker stomatal responsiveness. A plant's ability to better control water loss and increase carbon assimilation could affect how it handles stressors such as drought, and play a role in the health and yields of candidate bioenergy feedstocks. Understanding the bases for water management could aid the identification and selection of individuals better suited for growing in otherwise marginal soils.

Using a genetic screen, a Stanford University team led by Dominique Bergmann identified a *B. distachyon* mutant unable to produce subsidiary cells. As reported March 17, 2017, in *Science*, by comparing the whole genome sequence of *B. distachyon* with this *sid* mutant, a 5-base pair deletion that encodes for the transcription factor BdMUTE was discovered. Further, BdMUTE was identified as the mobile transcription factor responsible for coordinating the development of subsidiary and guard cell complexes. The unique subsidiary cells in grasses may allow for an enhanced performance when stressors such as increased temperature or drought are placed on the plant. Though his contribution to the work predates his time at the JGI (he was then at the U.S. Department of Agriculture) Plant Functional Genomics lead and study co-author John Vogel provided the team with the mutant population and showed them how to manipulate the plant for their studies.



JGI researchers sequenced and analyzed the microbial communities in restored wetlands in the Sacramento-San Joaquin Delta ([learn more on page 48](#)).



# Biogeochemistry

## Toward More Effective Carbon Fixation

Unlike DNA sequencing, where the language of life is read from the genome of an organism, DNA synthesis entails first the identification of a particular genetic element — such as an enzyme for fixing carbon from the atmosphere — and writing and expressing that code in a new system.

By tapping the DNA synthesis expertise of the JGI, a team from the Max Planck Institute (MPI) for Terrestrial Microbiology has reverse engineered a biosynthetic pathway for more effective carbon fixation. As reported November 18, 2016, in *Science*, this novel pathway is based on a new CO<sub>2</sub>-fixing enzyme that is nearly 20 times faster than the most prevalent enzyme in nature responsible for capturing CO<sub>2</sub> in plants by using sunlight as energy.

Photosynthesis harnesses sunlight to turn carbon dioxide into sugars that cells can use as energy along with other natural processes on the planet. It accounts for the transformation of some 350 billion tons of CO<sub>2</sub> annually. Researchers are concerned about how to capture the excess CO<sub>2</sub>, remove it from the atmosphere, and render it into energy and natural products for the economy. In support of the MPI team's efforts, the JGI synthesized hundreds of Enoyl-CoA Carboxylase/Reductase (ECR) enzyme variants through its CSP. This enabled the MPI team to zero in on the ECR with the highest CO<sub>2</sub>-fixation activity to successfully build a more efficient artificial CO<sub>2</sub> fixation pathway in a test tube. Through sequencing and synthesis, they sourced 17 different enzymes from nine different organisms across the three kingdoms of life, and orchestrated these parts to achieve a proof of principle CO<sub>2</sub>-fixation pathway performance exceeding that found in nature. Erb dubbed it the "CETCH cycle," because it "catches" CO<sub>2</sub> more efficiently from the atmosphere.

Emboldened by the successful reconstitution of a synthetic enzymatic network in a test tube for the conversion of CO<sub>2</sub> into organic products that is superior to chemical processes and competes favorably with those in nature, Erb said this important discovery opens the door for other future applications. "These could include the introduction of synthetic CO<sub>2</sub>-fixation cycles into organisms to bolster natural photosynthesis, or say, in combination with photovoltaics, lead the way to artificial photosynthesis. This might at the end jumpstart the design of self-sustaining, completely synthetic carbon metabolism in bacterial and algal systems."

Yasuo Yoshikuni, the head of the JGI's DNA Synthesis Science group, looks to a future where DNA sequencing and biological functions further converge and leverage DNA synthesis. "Through JGI's high-throughput sequencing capabilities coupled with the rapidly decreasing price of DNA synthesis, we continue to enable our user community in bringing to light the physiological potential of microorganisms and microbial communities. In the longer term, we hope to see these test-tube results yield a new generation of real bioproducts delivered to address critical energy and environmental challenges."

## Antarctic Adaptations in Diatoms

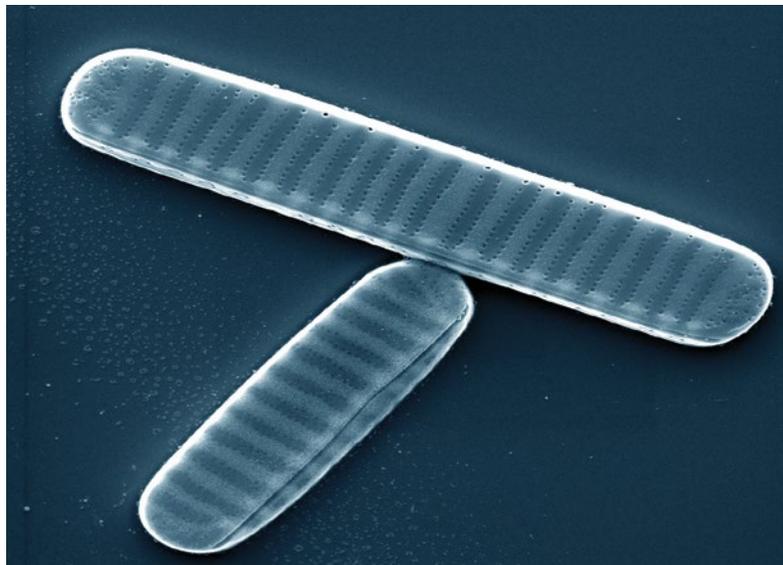
In the Antarctic Ocean, large populations of the diatom *Fragilariopsis cylindrus* dominate phytoplankton communities. Diatoms are a common type of photosynthetic microorganism, found in many environments from marine to soil. In the oceans, they are responsible for more than a third of the global ocean carbon captured during photosynthesis. This leads to a significant amount of sequestered carbon ending up in the sediments at the bottom of the ocean.

To learn more about how *F. cylindrus* adapted to its extremely cold environment, a team led by University of East Anglia (UEA) scientists conducted a comparative genomic analysis involving three diatoms by tapping the JGI's sequencing and annotation expertise. Said UEA's Thomas Mock, senior author of the work reported January 16, 2017, in *Nature*, "Our data provide first insights into how these key organisms underpinning one of the largest and unique marine ecosystems on Earth have evolved."

The genome of *F. cylindrus* was sequenced as part of the JGI's 2007 CSP portfolio. For the comparative analysis, its genome was compared against those of the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*, both found in temperate oceans with higher concentrations of dissolved iron, and both genomes previously reported by the JGI. *F. cylindrus* is diploid (has two copies of each chromosome, and thus two versions of each gene) and can selectively express the variant best suited to helping it deal with its environment. This provides additional genome-rooted resilience to the organism as its environment changes.

"Finding that the *F. cylindrus* population maintains and supports extensive variation to provide the adaptive ability of the population under harsh environmental conditions has broad implications for our understanding of how natural populations respond to changing environmental conditions," said Jeremy Schmutz, head of the JGI's Plant Program and a study co-author. "This will likely change the way the genomic techniques and assays are applied by the community to ocean-dwelling eukaryotic species."

Thomas Mock spoke at the 12th Annual Genomics of Energy and Environment Meeting. Watch his talk at <http://bit.ly/JGI2017Mock>.



Scanning electron micrograph of two cells of *Fragilariopsis cylindrus*. Shown are two silica shells (Frustules) in valve view. Magnification: 15,000X; scale bar: 5  $\mu\text{m}$ . (Gerhard S. Dieckmann)

## Algal Secrets to Thriving in the Intertidal Zone

The intertidal zone is the area between land and sea sometimes concealed by high tide or revealed by low tide. As this ecosystem is in constant flux, the organisms that inhabit the area have adapted to thrive under a range of constantly changing environmental conditions. *Porphyra* and other genera of bangiophyte red algae thrive in the intertidal zones of the northern and southern hemispheres.

As reported July 17, 2017, in the *Proceedings of the National Academy of Sciences*, the JGI sequenced, assembled, and annotated the genome of the red alga *Porphyra umbilicalis* to better understand how it harvests light and nutrients, and how warming oceans might affect its ability to fix carbon. Though red algae are one of the oldest multicellular lineages, only a few have had their genomes sequenced. *P. umbilicalis* is found in the ocean's intertidal zone, and is subject to constantly changing environmental conditions including temperature, light, and desiccation levels. As diatoms and other photosynthesizing microorganisms evolved from red algae, red algae metabolism had a significant impact on the planet's carbon cycle.

Green algae and red algae are both groups of plants that carry out photosynthesis using light-harnessing organelles called chloroplasts, which evolved from cyanobacteria that were engulfed by the ancestral eukaryotic algae. Later, other environmentally important algae, such as diatoms, dinoflagellates, and haptophytes, evolved when other non-photosynthetic eukaryotes captured red algae and integrated the red algal chloroplast and red algal nuclear genes into their genomes. These processes greatly diversified the organisms capable of conducting photosynthesis, and the red algal imprint on global productivity, aquatic food webs, and oxygen production is significant.

The team led by University of Maine researchers found that *P. umbilicalis* has previously unrecognized means of tolerating its physically stressful intertidal habitat. For example, the red alga has multiple strategies to protect cells from being damaged by high light levels, including expanded families of proteins that protect the photosynthetic apparatus from high light and unusual genomic arrangements of the genes that synthesize the mycosporine-like amino acids protecting against ultraviolet light. The team also found strong cytoskeletal limitations in *Porphyra* and most other red algae with sequenced genomes, offering a possible explanation for why red algae tend to be small compared with other multicellular eukaryotes.

*Porphyra umbilicalis* (laver) attains high biomass despite the high levels of stress in its habitat in the upper intertidal zone of the North Atlantic, as shown here at low tide at Sand Beach, Acadia National Park, Maine. (Susan Brawley)

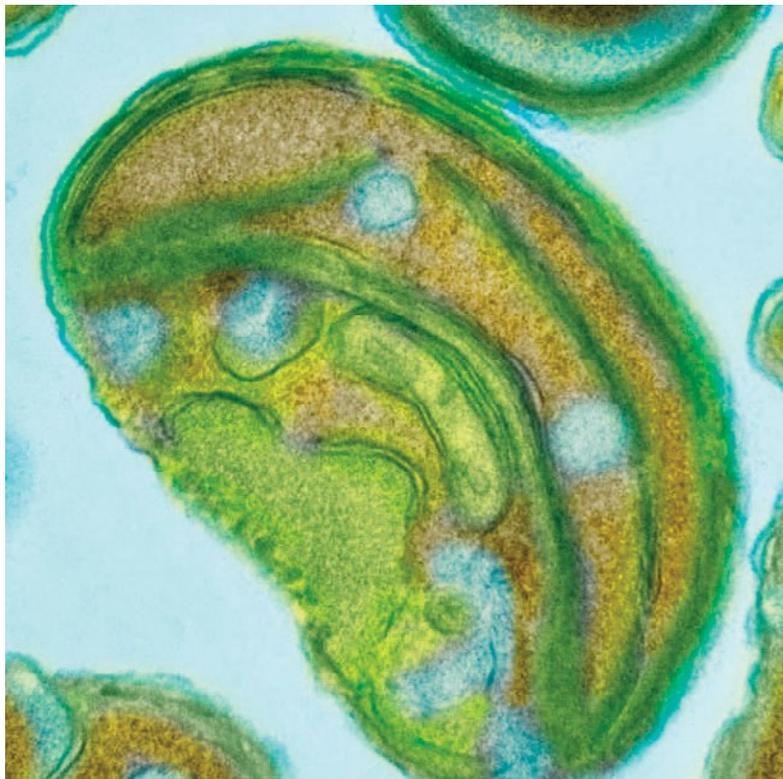


## Tiny Algae, Large Genomic Variation

As their name suggests, picophytoplankton such as *Ostreococcus* are invisible to the naked eye. Despite their size, their global abundance means they are a widespread primary producer and form the bases of several marine food webs. In coastal areas, they account for as much as 80% of the available biomass.

*Ostreococcus* is considered a model species to study marine picophytoplankton. A decade ago, JGI sequenced one of the *Ostreococcus* strains. That genome, along with other genome sequences from three *Ostreococcus* groups, revealed the tiny algae's diversity and adaptation to different ecological niches around the world. As reported July 5, 2017, in *Science Advances*, a team led by researchers at the Oceanological Observatory of Banyuls, France, and including JGI scientists, has resequenced and analyzed 13 members of a natural population of *Ostreococcus tauri* from the northwest Mediterranean Sea.

The analysis, enabled in part by the JGI's CSP, revealed that the *O. tauri* population is larger than anticipated, with high genetic and phenotypic diversity influenced by the algae's natural resistance to ocean viruses. The team identified two large candidate mating type loci, consistent with the pervasive evidence of recombination and thus sexual reproduction within the population. Understanding the genetic variability of various *Ostreococcus* strains will help researchers understand how environmental changes affect their abundance and ability to photosynthesize.



*Ostreococcus* micrograph.  
(courtesy Wenche Eikrem  
and Jahn Throndsen,  
University of Oslo)

Study co-author Rhonzhong Ye and graduate student Jennifer Morris collecting greenhouse gas fluxes from the rice fields studied on Twitchell Island, Calif. (Wyatt Hartman)



## Nutrient Availability in Model Wetlands

Adjacent to the restored wetlands of Twitchell Island in California's Sacramento-San Joaquin River Delta are rice fields with soil carbon contents that can vary between 2.5% and 25%, covering much of the global range of carbon found in soils. Conducting studies across a nutrient gradient allows researchers to better understand the relationship between carbon cycling, nutrient availability, and microbial communities in soil.

JGI researchers studied the ecosystems of Twitchell Island in the Sacramento-San Joaquin Delta, where the U.S. Geological Survey had a pilot study on restored wetlands. Rice fields are model wetland systems that allow researchers to focus on chosen biogeochemical variables, while factors such as water and vegetation are controlled. The DOE is interested in understanding the roles of microbial communities in long-term impacts on carbon emissions and carbon sequestration. Wetland ecosystems can trap as much as 30% of global soil carbon but contribute nearly 40% of global methane emissions, providing an opportunity to understand their roles as both carbon sinks and carbon sources. As reported July 21, 2017, in *The ISME Journal*, the team's findings suggest that the microbial metabolic rates align with biological stoichiometry theory, a metabolic theory of ecology that suggests organisms with faster growth rates require more phosphorus to increase nitrogen-rich protein synthesis. Until now, this theory had not been applied to soil microbes *in situ* due to methodological limitations, which the scientists addressed using a novel genomic approach.

By studying the microbial communities in these soils, the researchers found that the rate at which microbes break down organic matter is coupled to the availability of carbon, nitrogen, and phosphorus in the soils. Specifically, the availability of phosphorus is a key factor in determining soil carbon cycling rates. An abundance of phosphorus increases microbial activity and metabolic rates, which in turn means higher carbon turnover. Lower phosphorus in high-carbon soils may help stabilize accumulated carbon, while high-phosphorus soils may more rapidly lose carbon stores. These associations at the ecosystem scale were also reflected in genomic data from the soil microbes that drive soil element cycling. In showing how microbial metabolism is regulated by coupled nutrient cycling and soil carbon availability, the researchers demonstrate how genomics studies of microbial communities can be scaled up to the ecosystems level, which will contribute to a deeper understanding of ecological processes and aid the development of better global carbon cycling models.

## Microbial Responses to Drilling for Oil

Despite a shift toward more sustainable, alternative energy sources, people still rely on fossil fuels for energy and transportation. Owned by Denmark, the Halfdan North Sea oil field is an offshore petroleum reservoir managed by a joint venture that includes Shell, one of the world's largest oil companies. Over the past 15 years, 32 oil wells have been drilled, reaching depths over 2 kilometers below the seafloor to extract petroleum from this deep subsurface reservoir. Recovering gas and oil for energy and transportation fuels affects the previously isolated microbial communities in the deep subsurface. For example, water injections introduce new microbes to the existing indigenous populations, which can lead to the production of the highly toxic and corrosive hydrogen sulfide, altering oil well productivity and oil quality.

Understanding how microbial communities in subsurface petroleum reservoirs are affected by human activity, and how their responses can sour oil production, can influence oil industry practices. To learn more about how some of these populations respond to disruptions in their environment, researchers at Shell collaborated with the JGI and Newcastle University to analyze samples from the field's 32 oil wells. As reported May 19, 2017, in *The ISME Journal*, they conducted a comparative genetic analysis of the microbial communities in multiple oil wells within an offshore oil field.

They tracked the succession of microbial populations in oil wells drilled at different times, community compositions, and how these compositions change over time the longer an oil well stays in production mode. Tools developed at the JGI and made available through the IMG/M system were utilized during the analyses. The data generated provide insights into just how deep subsurface microbial communities are perturbed by active oil wells injecting foreign substances into these previously isolated populations, and help industry researchers develop new techniques for managing microbiological problems.

"For the first time, we are able to shed light into this remote and largely unexplored ecosystem," said study senior author Nicolas Tsesmetzis of Shell International Exploration and Production Inc. "The largely heterogeneous and highly diverse microbial communities recovered from the different oil wells of the same oil field were rather unexpected and bring about a paradigm shift in our standard microbial monitoring practices of petroleum systems."



Shell researchers collected samples from oil wells in a North Sea oil field like this one. (Berardo62, Flickr, CC BY-SA 2.0)

The oil slick in the Gulf of Mexico from the Deepwater Horizon oil spill as seen on April 29, 2010. (NASA Earth Observatory image created by Jesse Allen, using data provided courtesy of the University of Wisconsin's Space Science and Engineering Center MODIS Direct Broadcast system)



## Lessons from Simulating a Deep Ocean Oil Spill

In the three months between the time the Deepwater Horizon Macondo well exploded on April 22, 2010, until it was capped in mid-July, thousands of barrels of oil and large quantities of natural gases flowed into the Gulf of Mexico daily. It was the first major release of oil and natural gases into the deep ocean (1,500 meters). Due to the depth of the spill, vast plumes of small oil droplets remained trapped deep in the ocean (900–1,300 meters) where they underwent biodegradation by the local microbial community.

A team led by Berkeley Lab researchers reproduced the dispersal to replicate the successive populations of diverse microbes over 64 days and recover high-quality draft genomes to determine metabolic factors driving microbial community shifts. They discovered that the microbial community transformed with chemical changes in residual oil. Additionally, their lab-based method allowed for the first time the successful resolution of high-quality genomes and the characterization of functional capabilities for all the key microbes. The expertise and resources used to reconstruct the microbial genomes demonstrates how new technology development at the JGI is enabling energy and environment research. Identifying the microbes involved in degrading hydrocarbons (the chief components of petroleum and natural gas), as well as the drivers that trigger successive waves of microbial responders, allows researchers to better understand how the microbial community adapted to the events of seven years ago.

The researchers presented the first complete picture of how successive waves of microbial populations degraded the released oil on June 26, 2017, in *Proceedings of the National Academy of Sciences*. Samples from three time points (days six, 18, and 64) were chosen for metagenomic sequencing and analysis. Assembled genome fragments were binned or assigned to draft genomes of origin using resources provided by the JGI's Emerging Technologies Opportunity Program (ETOP). The researchers were also able to recover high-quality genomes of the key microbial players, and determine the metabolic factors driving the shifts between microbial communities. This research demonstrates the rapid, specialized biodegradation by the deep ocean microbial community was possibly driven by crude oil alone.

**JGI-owned Compute Resources**  
170+ million NERSC hours

**JGI BER Allocation**  
25 million NERSC hours

**JAMO**  
5.5 million records

**JAMO Archived Data Footprint**  
5.9 Petabytes

**JGI File Systems**  
7.1 Petabytes of data





# Computational Infrastructure

The Joint Genome Institute's (JGI) capacity as a next-generation genomics user facility has generated petabytes of high-quality sequence data and analysis. To support this workload, the JGI has invested significant resources in its high-performance compute cluster, Genepool, as well as its storage and web infrastructure, adding 192 nodes of capacity in the Cori supercomputer.

The success of these computing projects is in part due to the JGI's ongoing partnership with the National Energy Research Scientific Computing Center (NERSC), one of the nation's foremost centers for high-performance computing. Both sides have learned a great deal since 2010, when all of the JGI's computational resources were moved to NERSC. The infrastructure advancements to Genepool and other JGI portals mean rapid and smooth access for users across the globe. The JGI's partnership with NERSC enables its researchers and users to devote more of their time to cutting-edge genomics research.

In 2016, NERSC acquired Cori, a \$70-million investment in data-intensive and high-performance computing infrastructure. The JGI used more than 10 million central processing unit hours on this petascale supercomputer. Many of these calculations could not have been completed on the Genepool cluster because of the computing scale required. One of the core features of Cori is 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 software called Shifter to allow Docker containers to run on the Cray supercomputers. This work grew out of collaboration with the JGI staff and has reduced the hurdles required to move bioinformatics, light source, and astronomy workloads to Cori. In 2017, in collaboration with NERSC consultants Daniel Udvary and Anthony Wildish, JGI staff undertook a major effort to move pipelines and software to Docker containers that enable JGI software to run on Cori, Genepool, the DOE Knowledgebase (KBase), and the cloud. The JGI also leverages Bioboxes where possible, to standardize the Docker container interfaces for bioinformatics tools. Bioboxes is an effort in the community that has been led in part by Michael Barton of the Research and Development Group.

The KBase and JGI released a new feature for users that enables scientists on the KBase platform to search and retrieve any public JGI data set. This search interface relies on an API to an Elastic Search instance built on top of the JGI Archive and Metadata Organizer (JAMO) developed by Harika Tandra and Chris Beecroft of the Sequence Data Management Group. In 2017, BBtools developed by Brian Bushnell and some of the DOE QC pipelines were made available on the KBase platform. The JGI is currently working with KBase to ensure JGI software and tools are available on the KBase platform via Docker and the KBase SDK.

The JGI and NERSC co-organized a workshop at the 12th Annual Genomics of Energy and Environment Meeting. In this workshop, users learned about Docker technology and got hands-on experience running containers in the cloud. In 2018, the computational workshops at the Genomics of Energy and Environment Meeting will focus on the KBase platform, and users can learn how to access NERSC resources via the KBase graphical user interface. In 2017, the JGI and NERSC accepted six proposals (see Appendix C) from the Microbiome Data Science call under the aegis of the Facilities Integrating Collaborations for User Science (FICUS) program. This joint call, led by Emilee Elloe-Fadrosh and Nikos Kyrpides, was issued by the two DOE user facilities and helps scientists perform large-scale computational analyses of sequence data.

The page features a large, vertical green abstract graphic on the left side. This graphic is composed of several overlapping, semi-transparent shapes in various shades of green, including circles, triangles, and irregular polygons. The text 'Appendices' is centered within this green area in a white, sans-serif font. To the right of the green graphic, the page is white and contains two thin, horizontal grey lines that span the width of the page, one near the top and one near the bottom.

# Appendices

## Appendix A

### Acronyms at a Glance

<b>AF</b>	Alignment Fraction	<b>ETOP</b>	Emerging Technologies Opportunity Program
<b>ANI</b>	Average Nucleotide Identity	<b>FACS</b>	Fluorescence Activated Cell Sorting
<b>BER</b>	DOE Office of Biological and Environmental Research	<b>FICUS</b>	Facilities Integrating Collaborations for User Science
<b>BERAC</b>	Biological and Environmental Research Advisory Committee	<b>FISH</b>	Fluorescent <i>In Situ</i> Hybridization
<b>BESC</b>	BioEnergy Science Center (at ORNL)	<b>FY</b>	Fiscal Year
<b>BLAST</b>	Basic Local Alignment Search Tool	<b>GEBA</b>	Genomic Encyclopedia of Bacteria and Archaea
<b>BOOST</b>	Build Optimization Software Tools for DNA Synthesis	<b>GLBRC</b>	Great Lakes Bioenergy Research Center (at the University of Wisconsin-Madison)
<b>BRC</b>	Bioenergy Research Center (i.e., BESC, CABBI, CBI, GLBRC, JBEI)	<b>GOLD</b>	Genomes OnLine Database
<b>CABBI</b>	Center for Advanced Bioenergy and Bioproducts Innovation (at the University of Illinois)	<b>GSC</b>	Genomic Standards Consortium
<b>CAMI</b>	Critical Assessment of Metagenome Interpretation	<b>HPC</b>	High-Performance Computing
<b>CBI</b>	Center for Bioenergy Innovation (at ORNL)	<b>HPPS</b>	High-Performance Storage System
<b>CBP</b>	Consolidated Bioprocessing	<b>IEP</b>	Industry Engagement Program
<b>CRISPR</b>	Clustered Regularly Interspaced Short Palindromic Repeats	<b>IGB</b>	Integrative Genomics Building
<b>CRADA</b>	Cooperative Research & Development Agreement	<b>IMG</b>	IMG/M Integrated Microbial Genomes & Microbiomes System
<b>CRT</b>	Computational Research and Theory Building (at Berkeley Lab)	<b>ISM</b>	Integrated Safety Management
<b>CSP</b>	Community Science Program	<b>ITS</b>	Integrated Tracking System
<b>DOE</b>	Department of Energy	<b>JAMO</b>	JGI Archive and Metadata Organizer
<b>ECP</b>	Exascale Computing Project	<b>JBEI</b>	Joint BioEnergy Institute (at Berkeley Lab)
<b>EMSL</b>	Environmental Molecular Sciences Laboratory (at PNNL)	<b>JLT</b>	JGI Leadership Team
		<b>KBBase</b>	DOE Systems Biology Knowledgebase
		<b>LANL</b>	Los Alamos National Laboratory
		<b>LBNL</b>	Lawrence Berkeley National Laboratory (Berkeley Lab)

<b>LLNL</b>	Lawrence Livermore National Laboratory
<b>MAG</b>	Metagenome-Assembled Genome
<b>MIMAG</b>	Minimum Information about a Metagenome-Assembled Genome
<b>MISAG</b>	Minimum Information About a Single Amplified Genome
<b>MGM</b>	Microbial Genomics and Metagenomics
<b>NERSC</b>	National Energy Research Scientific Computing Center
<b>NGEE</b>	Next-Generation Ecosystem Experiments
<b>NREL</b>	National Renewable Energy Laboratory
<b>ORNL</b>	Oak Ridge National Laboratory
<b>PMO</b>	Project Management Office
<b>PNN</b>	Pacific Northwestern National Laboratory
<b>QTL</b>	Quantitative Trait Loci
<b>SAC</b>	Scientific Advisory Committee
<b>SAG</b>	Single Amplified Genome
<b>SFA</b>	Scientific Focus Area
<b>UEA</b>	University of East Anglia
<b>USDA-ARS</b>	United States Department of Agriculture- Agricultural Research Service
<b>WIP</b>	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 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 constitutes the blueprint of life.

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

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

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

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

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

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

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

**Carbon cycle:** The biogeochemical process by which carbon is exchanged 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.

**DAP-seq:** A high-throughput method for identifying protein binding sites in DNA.

**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 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.

**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 science of information and computer information systems. At JGI, it is the science of managing and interpreting genomic information with computational tools.

**Library:** A collection of DNA fragments.

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

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

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

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

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

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

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

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

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

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

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

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

---

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

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

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

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

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

**Quantitative trait loci (QTL):** Genome regions on chromosomes to which genetic traits can be mapped.

**RB-TnSeq:** Randomly-Barcoded Transposon Sequencing is a technique to generate sequence-tagged insertion mutant strains of a single-celled organism (typically bacteria) to simultaneously assess the functions of every gene in the genome in a variety of conditions by sequencing and counting the abundance of each tag.

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

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

**Sequence:** Order of nucleotides (base sequence) in a nucleic acid molecule. In the case of DNA sequence, it is the precise ordering of the bases (A, T, G, and C) from which the DNA is composed. Also used as a verb to describe the process of determining the nucleotide order.

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

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

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

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

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

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

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

## Appendix C

### 2017 User Program Supported Proposals

#### Community Science Program (CSP)

Investigator	Affiliation	Description
Aitken, Karen	CSIRO (Australia)	Understanding polyploidy through the generation of the first sugarcane genome sequence
Attwood, Graeme	AgResearch Ltd (New Zealand)	Defining gene function in rumen microbes
Bohlmann, Joerg	University of British Columbia (Canada)	Exploring the G3 "Gymnosperm Giga-Genomes" for carbon sequestration, biofuels, and bioproducts
Bonito, Gregory	Michigan State University	Functional gene network regulation in tripartite plant-fungal-bacterial mutualisms of bioenergy plants and algae
Bruns, Thomas	University of California, Berkeley	Functional genomics of pyrophilous fungi determining the fate of pyrolyzed carbon in post-fire soils
Buchan, Alison	University of Tennessee Knoxville	Geo-metabolomics of a saltmarsh: Combining <i>in situ</i> , bulk, genomic, transcriptomic, and DOMEomic data streams
Coleman-Derr, Devin	USDA-ARS	Exploring the role of drought-induced plant associated microbes in promoting plant fitness in <i>Sorghum bicolor</i> and <i>Oryza sativa</i>
Cox, Michael	University of Wisconsin-Madison	Defining the molecular basis of extreme resistance to ionizing radiation
Crowe, Sean	University of British Columbia (Canada)	Mapping the global methanome
Cullen, Daniel	Forest Products Laboratory	Metatranscriptome analysis of fungal decay of <i>Pinus contorta</i>
Cushman, John	University of Nevada	Ice plant gene atlas resource development for <i>Mesembryanthemum crystallinum</i> , a facultative crassulacean acid metabolism (CAM) model for improved water-use efficiency of bioenergy feedstocks
Des Marais, David	Harvard University	Perenniality, abiotic stress tolerance, and biomass allocation in <i>Brachypodium</i> , a model grass genus for bioenergy
Dijkstra, Paul	Northern Arizona University	Stress in microbial communities in response to changes in carbon and nitrogen availability

## Community Science Program (CSP) (continued)

Investigator	Affiliation	Description
Dopson, Mark	Linnaeus University (Sweden)	Exploring deep biosphere microbial communities by single-cell DNA sequencing
Doty, Sharon	University of Washington	Functional genomics of poplar endophytes for elucidation of mechanisms of improved plant growth under challenging conditions
Dudycha, Jeffry	University of South Carolina	Unlocking the photosynthetic diversity of cryptophyte algae through whole-genome sequencing
Francis, Christopher	Stanford University	Spatiotemporal characterization of microbial communities controlling estuarine nitrogen and carbon cycling in the San Francisco Bay-Delta
Göker, Markus	DSMZ (Germany)	The One Thousand Microbial Genomes Phase 4 Project (KMG-4) – sequencing the most valuable type-strain genomes for metagenomic binning, comparative biology and taxonomic classification
Harris, Steven	University of Nebraska – Lincoln	Fungal interaction networks in biological soil crusts
Heyduk, Karolina	University of Georgia	Genome sequencing of C3 and CAM Yucca species
Johnson, Matthew	Woods Hole Oceanographic Institute	The role of acquired phototrophy in phytoplankton blooms: Insights from the <i>Mesodinium rubrum</i> genome
Kellogg, Elizabeth	Donald Danforth Plant Science Center	Pan-genomics of big bluestem, a broadly adapted dominant grass
LeBoldus, Jared	Oregon State University	RNAseq enabled metabolic modeling of disease resistance to <i>Septoria</i> canker in the DOE flagship <i>P. trichocarpa</i>
Lopez Peredo, Elena	Marine Biological Laboratory	Protecting photosynthesis during desiccation: do the genomes of desert-derived and aquatic <i>Scenedesmus</i> species hold the key to understanding extreme desiccation tolerance among green algae?
MacGregor, Barbara	University of North Carolina	Single-cell (meta-)genomics of uncultivable large sulfur bacteria and their epibionts: investigating host-microbe mediation of biogeochemical cycling
Mockler, Todd	Danforth Center	A complete-sequence population for pan-genome analysis of sorghum
Moran, Mary Ann	University of Georgia	Dynamics of bacterial carbon and sulfur cycling in a coastal environment

## Community Science Program (CSP) (continued)

Investigator	Affiliation	Description
Nguyen, Nhu	University of Hawai'i at Manoa	A genome atlas of the ectomycorrhizal genus <i>Suillus</i> : Phylogenetic diversity and population genomics of a keystone guild of symbiotic forest fungi
Nicholson, Wayne	University of Florida	Transcriptomic and methylomic responses of <i>Carnobacterium</i> species to extreme low pressure
Pett-Ridge, Jennifer	Lawrence Livermore National Laboratory	Microbial carbon transformations in wet tropical soils: Effects of redox fluctuation
Raff, Jonathan	Indiana University	Combined flux chamber and genomics approach to understanding soil emissions of reactive nitrogen oxides in a forested environment
Shade, Ashley	Michigan State University	Greater than the sum of its parts? A synthetic microbial community approach to untangle member interactions and exometabolite production
Smart, Christine	Cornell University	Genetic diversity of shrub willow pathogen <i>Melampsora americana</i> aided by genome sequence
Stepanauskas, Ramunas	Bigelow Laboratory for Ocean Sciences	Expanding the dark matter reference catalog by targeting taxonomic blind spots
Whitman, William	University of Georgia	Core and pangenomes of soil and plant-associated prokaryotes
Yang, Xiaohan	Oak Ridge National Laboratory	Gene atlas for <i>Kalanchoe laxiflora</i> , a obligate crassulacean acid metabolism (CAM) model for genetic improvement of water-use efficiency in bioenergy feedstocks
Zhang, Chi	University of Nebraska – Lincoln	Genome sequencing of Zygnematales, the closest algal lineage to land plants, as a foundation for comparative genomic, transcriptomic, epigenetic, evolutionary, and biochemical studies



*Kalanchoe fedtschenkoi*  
(Forest and Kim Starr,  
Flickr CC BY 2.0)

## Small-Scale Proposals

INVESTIGATOR	AFFILIATION	DESCRIPTION
Alvarez-Cohen, Lisa	University of California, Berkeley	Unraveling functional dynamics and regulation crucial for the stability of an anaerobic ammonium oxidizing (anammox) community via community metatranscriptomics and 16S rRNA sequencing
Averill, Colin	Boston University	Molecular mechanisms of ectomycorrhizal interactions that stabilize soil carbon
Andras, Jason	Mount Holyoke College	The effect of ecological restoration on the structure and function of soil microbial communities in coastal wetlands
Bell, Terrence	Pennsylvania State University	How interactions with soil microbiomes impact the survival and activity of a nitrogen-fixing soil surface consortium
Bowen, Jennifer	Northeastern University	Building a foundation for understanding carbon and nitrogen cycling in microbially diverse salt marsh sediments
Chistoserdova, Ludmila	University of Washington	Probing the role of electron transfer redox mediators in methane-oxidizing communities
Coleman, Maureen	University of Chicago	Teasing apart coexisting picocyanobacteria and their contributions to biogeochemistry
Cotner, James	University of Minnesota-Duluth	Translating stoichiometric diversity into genomic diversity: What elements are responsible for variability in bacterial biomass stoichiometry?
Craft, Chris	Indiana University	Microbial activity and community composition in a tidal freshwater marsh in response to sea level rise and saltwater intrusion: A field manipulation
D'Agostino, Paul	Technische Universität München (Germany)	Genome mining and synthetic biology of underrepresented organisms in the search for novel (bio-)chemistry
Dove, Nicholas	University of California, Merced	Immediate effects of prescribed fire on microbial communities, decomposition, and nitrification
Dynarski, Katherine	University of California, Davis	Microbial community composition and function throughout decomposition in temperate forests across a bedrock nitrogen gradient
Freedman, Zachary	West Virginia University	Assessing the recovery of microbial traits in bioenergy crop agroecosystems on reclaimed surface mines

### Small-Scale Proposals *(continued)*

INVESTIGATOR	AFFILIATION	DESCRIPTION
Gutierrez, Tony	Heriot-Watt University (UK)	Genomic reconstruction to resolve the functional potential of oil-degrading bacteria in deep and surface waters of the Faroe-Shetland Channel
Hatzenpichler, Roland	Montana State University	Genomic characterization of cosmopolitan sediment-dwelling archaea hypothesized to be involved in anaerobic carbon cycling
Hatzinikolaou, Dimitris	University of Athens (Greece)	Genome sequencing of lignin and cellulose degrading isolates from Greek habitats
Hunt, Dana	Duke University	Seasonal and disturbance-related alterations in the biogeochemical cycling of estuarine carbon
Jensen, Paul	University of California, San Diego	Extracellular electron shuttles: A mechanism for obligate aerobes to survive anaerobic conditions
Kelly, Charlene	West Virginia University	Quantifying ecosystem changes following American chestnut restoration: From microbes to ecosystem function
Konstantinidis, Konstantinos	Georgia Institute of Technology	Real-time identification of horizontal gene transfer events using series metagenomes
Lebeis, Sarah	University of Tennessee	Characterizing functional chemotaxis receptors in different root zones
Lindemann, Steve	Purdue University	Identifying genome properties and environmental conditions governing the assembly of stable cyanobacterial-heterotroph consortia
Lindow, Steven	University of California, Berkeley	Trait-driven assembly of the leaf surface microbiome in an agricultural context
Maresca, Julia	University of Delaware	Light-induced gene expression in cosmopolitan freshwater Actinobacteria
Martens-Habbena, Willm	Fort Lauderdale Research and Education Center	Biogeochemistry of carbon and nitrogen cycling in subsiding subtropical soils
Mavrodi, Dmitri	University of Southern Mississippi	Transcriptomic responses of beneficial <i>Pseudomonas</i> rhizobacteria to <i>Brachypodium</i> root exudates
McMahon, Katherine	University of Wisconsin-Madison	Reference Genomes for Freshwater Bacteria – Alpha- and Beta-Proteobacteria

### Small-Scale Proposals (continued)

INVESTIGATOR	AFFILIATION	DESCRIPTION
Müller, Henry	Graz University of Technology (Austria)	Elucidating the relevance of DNA methylation of strain-specific plant-microbe interactions
Myroid, David	Oregon State University	The metagenome of DIRT: How 20 years of altered carbon inputs have altered the functional potential of soil microbial communities
Oliveira, Rafael	State University of Campinas (Brazil)	Unravelling microbial communities associated with native plant species from P-impooverished soils of a global biodiversity hotspot
Ottesen, Elizabeth	University of Georgia	The genetic drivers and metabolic consequences of pelagic stream microbial community assembly
Pan, Chongle	Oak Ridge National Laboratory	Multi-cycle selection of sorghum microbiomes for biological nitrogen fixation
Richardson, Ruth	Cornell University	Metatranscriptomic responses of a butyrate-to-methane enrichment culture to sulfate and iron availability: elucidating the identities and strategies of metabolically versatile sulfate and iron reducing populations
Samuel, Buck	Baylor College of Medicine	Genomic analyses of the natural microbiome of <i>Caenorhabditis elegans</i>
Schmer, Marty	USDA-ARS	Nitrogen effects on soil microbial communities and soil organic carbon in a resilient bioenergy cropping system
Seyfferth, Angelia	University of Delaware	Impacts of rice cultivation practices on carbon and iron cycling: Unraveling functional microbial diversity in rice paddy rhizocompartments subject to silica-rich rice residue amendment
Valentine, David	University of California, Santa Barbara	Metagenomic reconstruction of novel archaeal genomes from energy-limited hypersaline sediments.
Veley, Kira	Donald Danforth Plant Science Center	Epigenetic and transcriptional changes that occur in populations of bacteria colonizing plant hosts
Venturi, Vittorio	International Centre of Genetic Engineering and Biotechnology (ICGEB) (Italy)	Sequencing of a set of identified and characterized rice bacterial endophytes

## Synthesis Proposals

INVESTIGATOR	AFFILIATION	DESCRIPTION
Baker, David	University of Washington	Large scale <i>de novo</i> enzyme design for next generation metabolic engineering
Beisel, Chase	North Carolina State University	Profiling the functional diversity of CRISPR-Cas systems using cell-free transcription-translation systems
Bokinsky, Gregory	Delft University of Technology (Netherlands)	Plug adapters for biology: Activating heterologous iron-sulfur enzymes to fully exploit Nature's catalytic potential
Chistoserdova, Ludmila	University of Washington	Establishing lanthanides as new life metals and understanding redox properties of lanthanide enzymes in metabolism of methane
Dueber, John	University of California, Berkeley	Identification of novel D-altronate dehydratases in the enolase superfamily enabling pectin utilization in <i>S. cerevisiae</i>
Huang, Possu	Stanford University	New protein platform for secondary metabolite detection
Jewett, Michael	Northwestern University	Reframing combinatorial assembly and rapid prototyping of biosynthetic pathways with cell-free systems
Juminaga, Alex	LanzaTech, Inc.	Advancing understanding of acetogenic CO <sub>2</sub> fixation by generating a gene knock-out library to accelerate design of commercial strains for autotrophic production of fuels and bioproducts
Mead, David	Varigen Biosciences	Metagenomic mining for next generation DNA polymerases
Philmus, Benjamin	Oregon State University	Linking cyanobacterial orphan biosynthetic gene clusters to secondary metabolites
Redding, Kevin	Arizona State University	Chloroplastic CO <sub>2</sub> reduction to formate supporting synthetic carbon fixation
Schmeing, Thomas	McGill University	Synthetic DNA to facilitate structural and functional understanding of nonribosomal peptide synthetase production of secondary metabolites
Scott, Kathleen	University of South Florida	Creation of constructs for functional expression of recently described inorganic carbon transporters widespread among sulfur- and iron-oxidizing chemolithoautotrophic microorganisms
Subramanian, Venkataraman	National Renewable Energy Laboratory	Molecular engineering of <i>Trichoderma reesei</i> for improved cellulase production
Trudeau, Devin	Weizmann Institute of Science (Israel)	Genome mining to find a novel ribulose-1-phosphate kinase
Welander, Paula	Stanford University	Expression of novel triterpenoid biosynthesis proteins from environmental metagenomes

## Facilities Integrating Collaboration for User Science (FICUS) Proposals

Through the joint Facilities Integrating Collaborations for User Science (FICUS) initiative, the JGI has partnered with other national user facilities and called for Collaborative Science Initiative proposals. The accepted proposals began on October 1, 2017, providing the researchers with access to the capabilities of both user facilities and datasets beyond what could be generated by either facility alone.

### JGI-EMSL Collaborative FICUS proposals:

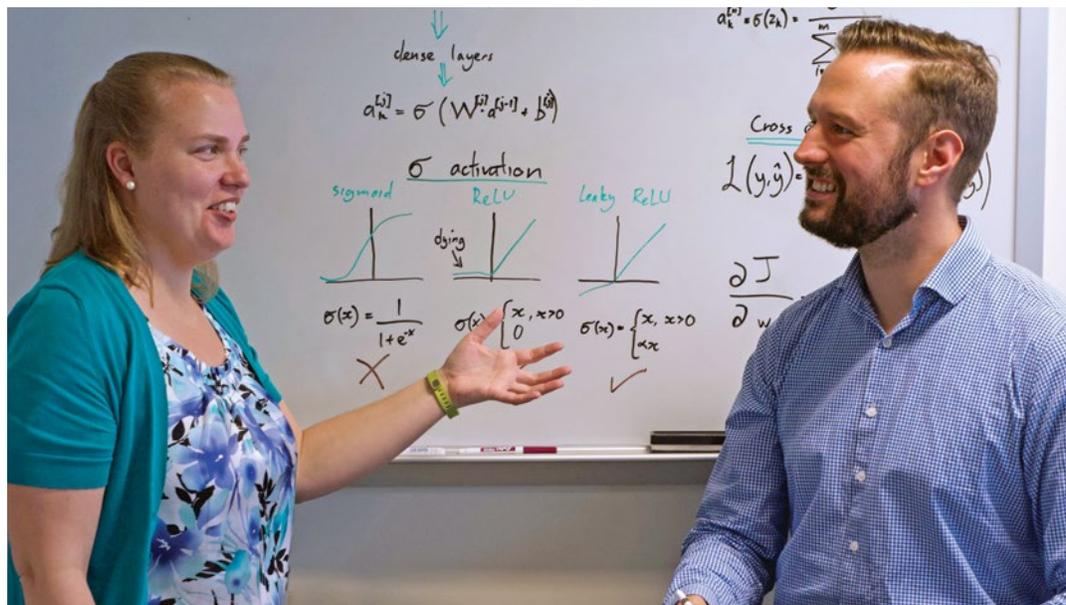
The FICUS JGI-EMSL call represents a unique opportunity for researchers to harness the combined power of genomics and molecular characterization in one research project to help advance the missions of the DOE's Office of Biological and Environmental Research (BER).

INVESTIGATOR	AFFILIATION	DESCRIPTION
Baldrian, Petr	Institute of Microbiology ASCR (Czech Republic)	The impacts of nitrogen availability and seasonal dynamics on plant-microbial interactions affecting C and N cycling in coniferous forest soils
Bartley, Laura	University of Oklahoma	Systems analysis of grass secondary cell wall development and regulation for biofuel production
Bianchi, Thomas	University of Florida	The role of priming effects on the conversion of blue carbon to CO <sub>2</sub> in the coastal zone
Blanchard, Jeffrey	University of Massachusetts Amherst	Molecular mechanisms underlying changes in the temperature sensitive respiration response of forest soils to long-term experimental warming
Cattolico, Rose Ann	University of Washington	Global warming induced salinity shifts: metabolic responses by algal-bacterial consortia
Fendorf, Scott	Stanford University	Metabolic constraints on organic matter decomposition and metal cycling in sediment deposits
Liao, Hui-Ling	Duke University	Combined 'omics approaches for the study of ectomycorrhizal symbiosis between <i>Suillus</i> and <i>Pinaceae</i> , with emphasis on their role in nutrient cycling
Rich, Virginia	The Ohio State University	Something old, something new: systems-level insights into plant-microbial-permafrost carbon dynamics by parallel high-resolution organic matter and microbial meta-omics
Skerker, Jeffrey	University of California, Berkeley	Understanding conversion of biomass-derived carbon into lipids and terpenoids in the oleaginous yeast <i>Rhodospiridium toruloides</i>
Wrighton, Kelly	The Ohio State University	Deciphering controls on plant decomposition in Arctic ecosystems: Identifying unknown microbial condensed tannin degradation pathways

### JGI-NERSC Collaborative FICUS proposals:

The expertise and capabilities available at the JGI and the National Energy Research Scientific Computing Center (NERSC) were made available through the FICUS JGI-NERSC call. Researchers can explore the wealth of genomic and metagenomic data generated worldwide through access to supercomputing resources and computational science experts to accelerate discoveries.

INVESTIGATOR	AFFILIATION	DESCRIPTION
Babbitt, Patricia	University of California, San Francisco	Environmental profiling of enzyme superfamilies for function prediction
Baker, David	University of Washington	Eukaryotic protein structure determination using metagenome and metatranscriptome sequence data
Brooks, Phillip	University of California, Davis	Advancing metagenome classification and comparison by MinHash fingerprinting of IMG/M data sets
DeLong, Edward	University of Hawaii at Manoa	Quantitative analyses of naturally occurring small RNAs in global metagenomic and metatranscriptomic datasets
Hallam, Steven	University of British Columbia (Canada)	Charting global biogeochemical cycles using fast phylogenetic mapping of functional anchor genes
Konstantinidis, Konstantinos	Georgia Institute of Technology	Assessing microbiomes at the individual population level: Tool development and applications to soil carbon cycling



JGI Genome Assembly group lead Alicia Clum talks strategy with Michael Barton of the Genome Analysis group

## Appendix D

### Advisory and Review Committee Members

#### The Scientific Advisory Committee (SAC)

The SAC is a board convened by the JGI Director to provide a scientific and technical overview of the JGI. Responsibilities of this board include providing technical input on large-scale production sequencing, new sequencing technologies, and related informatics; an overview of the scientific programs at the JGI; and an 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 DOE BER concurrence, set the final sequence allocation for this program.

##### Members

Mark Adams, *J. Craig Venter Institute*

Carol Bult, *The Jackson Laboratory*

Steve Briggs, *University of California, San Diego*

Jeff Dangl, *University of North Carolina*

Paul Flicek, *European Bioinformatics Institute (ad hoc)*

Claire M. Fraser, *University of Maryland*

N. Louise Glass, *Lawrence Berkeley National Laboratory*

Glenn Kubiak, *Lawrence Berkeley National Laboratory*

Mary Ann Moran, *University of Georgia*

Trina McMahon, *University of Wisconsin-Madison*

Deirdre Meldrum, *Arizona State University*

Juan Meza, *University of California, Merced*

Sue Wessler, *University of California, Riverside (ad hoc)*

#### Informatics Advisory Committee

##### 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, *University of California, San Diego (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, *ADM*

Ronald de Vries, *Westerdijk Fungal Biodiversity Institute (Netherlands)*

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

N. Louise Glass, *University of California, Berkeley; Lawrence Berkeley National Laboratory*

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

Siobhan Brady, *University of California, Davis*

Gloria Coruzzi, *New York University*

Jeff Dangl, *University of North Carolina, Chapel Hill*

Joe Ecker, *The Salk Institute for Biological Studies*

Samuel Hazen, *University of Massachusetts-Amherst*

Tom Juenger, *University of Texas, Austin*

Toby Kellogg, *Donald Danforth Plant Science Center*

Sabeeha Merchant, *University of California, Los Angeles*

Stephen Moose, *University of Illinois*

Sue Rhee, *Carnegie Institution for Science, Stanford*

Bob Schmitz, *University of Georgia*

Gary Stacey, *University of Missouri*

### Prokaryotic Super Program Advisory Committee

#### Members

Jill Banfield, *University of California, Berkeley*

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)*

Janet Jansson, *Pacific Northwest National Laboratory*

Kostas Konstantinidis, *Georgia Institute of Technology*

Monica Medina, *Penn State University*

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

Mary Ann Moran, *University of Georgia*

Nancy Moran, *University of Texas at Austin*

Jennifer Pett-Ridge, *Lawrence Livermore National Laboratory*

Rich Roberts, *New England Biolabs*

Ramunas Stepanauskas, *Bigelow Laboratory for Ocean Sciences*

Matt Sullivan, *The Ohio State University*

### DNA Synthesis Science User 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

### 2017 Genomics of Energy and Environment Meeting

With more than attendees at the Walnut Creek Marriott, the 12th Annual Genomics of Energy and Environment Meeting took place March 20–23, 2017.

#### Keynote Speakers



*“Our lab has started to ask different questions. What if we were able to find or design new enzymes for CO<sub>2</sub> fixation? And what if we were able to actually design new pathways that would give us faster access to biomass productivity and convert CO<sub>2</sub> to useful compounds?”*

— Tobias Erb

**Tobias Erb** of the Max Planck Institute for Terrestrial Microbiology in Marburg, Germany opened the meeting. A microbial synthetic biologist, he spoke about the importance of understanding how biochemical pathways are made in nature to construct more efficient synthetic approaches. Case in point, his research team has been working to engineer an approach to fixing carbon that would be as good or better than the ones that have evolved from nature (see page 41).

Watch his talk on the JGI YouTube channel at [bit.ly/JGI2017Erb](http://bit.ly/JGI2017Erb)

Read about the meeting at <http://bit.ly/JGIPrimerSpring17>

Learn more about the meeting talks at <https://usermeeting.jgi.doe.gov/past-meetings/2017-agenda/>

Videos of the talks are available on JGI's YouTube channel at [http://bit.ly/JGI2017\\_Videos](http://bit.ly/JGI2017_Videos)

*“JGI’s not stopping its sequencing [of] data, so we’re going to have all the data we want, and it’s going to be our job to make sense of it. What can we do as a community to better take advantage of it?”*

— C. Titus Brown



In his closing keynote, **C. Titus Brown** (2nd from the left above) from the University of California, Davis, spoke about the issues he’s faced wrestling with the unique challenges posed by genomic sequence analysis. As the datasets continue to pile on, he said, researchers must think about what they need to get out of the data, analyze that, and discard the rest. He also suggested that researchers take both past and future analyses of large-scale datasets to a higher level, and closed with a few recommendations: integrate multiple data types, develop better metadata exploration tools, and boost computational training.

Watch his talk on the JGI YouTube channel at [bit.ly/JGI2017Brown](http://bit.ly/JGI2017Brown).

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

Jan Leach, <i>Colorado State University</i>	Philipp Zerbe, <i>UC Davis</i>	Thomas Mock, <i>University of East Anglia</i>
Cat Adams, <i>UC Berkeley</i>	Mikael Andersen, <i>Technical University of Denmark</i>	Alex Greenspan, <i>University of California, Davis</i>
Colleen Hansel, <i>WHOI</i>	Stephen Mondo, <i>JGI</i>	Andrew Roger, <i>Dalhousie University</i>
Sunny Liao, <i>University of Florida</i>	Gloria Coruzzi, <i>New York University</i>	Teresa Pawlowska, <i>Cornell University</i>
Jonathan Lynch, <i>Penn State University</i>	Jay Chen, <i>Oak Ridge National Laboratory</i>	Abhishek Biswas, <i>University of Tennessee</i>
Renee Wegrzyn, <i>Defense Advanced Research Projects Agency</i>	Frederik Schulz, <i>JGI</i>	Mary Lipton, <i>Environmental Molecular Sciences Laboratory</i>
Lauren Alteio, <i>University of Massachusetts, Amherst</i>	Katherine (Trina) McMahon, <i>University of Wisconsin – Madison</i>	Harry Beller, <i>JBEI</i>
Patsy Babbitt, <i>UC San Francisco</i>	J. Chris Pires, <i>University of Missouri</i>	

## Appendix F

### 2017 Publications

- Ahn AC et al. Genomic diversity within the haloalkaliphilic genus *Thioalkalivibrio*. *PLoS One*. 2017 Mar 10;12(3):e0173517. doi: 10.1371/journal.pone.0173517.
- Andreopoulos W. Clustering categorical data. Wiley StatsRef: Statistics Reference Online. 2017 Feb 15; 1–12. doi: 10.1002/9781118445112.stat07907.
- Aserse AA et al. Draft genome sequence of type strain HBR26(T) and description of *Rhizobium aethiopicum* sp. nov. *Stand Genomic Sci*. 2017 Jan 26;12:14. doi: 10.1186/s40793-017-0220-z.
- Baltrus DA et al. Absence of genome reduction in diverse, facultative endohyphal bacteria. *Microb Genom*. 2017 Feb 28;3(2):e000101. doi: 10.1099/mgen.0.000101. eCollection 2017 Feb.
- Beukes CW et al. Genome data provides high support for generic boundaries in *Burkholderia Sensu Lato*. *Front Microbiol*. 2017 Jun 26;8:1154. doi: 10.3389/fmicb.2017.01154. eCollection 2017.
- Blanc-Mathieu R et al. Population genomics of picophytoplankton unveils novel chromosome hypervariability. *Sci Adv*. 2017 Jul 5;3(7):e1700239. doi: 10.1126/sciadv.1700239. eCollection 2017 Jul.
- Booker AE et al. Sulfide generation by dominant Halanaerobium microorganisms in hydraulically fractured shales. *mSphere*. 2017 Jul 5;2(4). pii: e00257-17. doi: 10.1128/mSphereDirect.00257-17. eCollection 2017 Jul–Aug.
- Bowers RM et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol*. 2017 Aug 8;35(8):725-731. doi: 10.1038/nbt.3893.
- Brawley SH et al. Insights into the red algae and eukaryotic evolution from the genome of *Porphyra umbilicalis* (Bangioophyceae, Rhodophyta). *Proc Natl Acad Sci USA*. 2017 Aug 1;114(31):E6361-E6370. doi: 10.1073/pnas.1703088114. Epub 2017 Jul 17.
- Brown CT et al. Measurement of bacterial replication rates in microbial communities. *Nat Biotechnol*. 2016 Dec;34(12):1256-1263. doi: 10.1038/nbt.3704.
- Browne DR et al. Draft nuclear genome sequence of the liquid hydrocarbon-accumulating green microalga *Botryococcus braunii* Race B (Showa). *Genome Announc*. 2017 Apr 20;5(16). pii: e00215-17. doi: 10.1128/genomeA.00215-17.
- Browne P et al. Genomic composition and dynamics among Methanomicrobiales predict adaptation to contrasting environments. *ISME J*. 2017 Jan;11(1):87-99. doi: 10.1038/ismej.2016.104.
- Butterfield CN et al. Proteogenomic analyses indicate bacterial methylophony and archaeal heterotrophy are prevalent below the grass root zone. *PeerJ*. 2016 Nov 8;4:e2687.
- Chen IA et al. IMG/M: Integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res*. 2017 Jan 4;45(D1):D507-D516. doi: 10.1093/nar/gkw929. Epub 2016 Oct 13.
- Chin CS et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods*. 2016 Oct 17. doi: 10.1038/nmeth.4035.
- Cole BJ et al. Genome-wide identification of bacterial plant colonization genes. *PLoS Biol*. 2017 Sep 22;15(9):e2002860. doi: 10.1371/journal.pbio.2002860. eCollection 2017 Sep.
- Contreras-Moreira B et al. Analysis of plant pan-genomes and transcriptomes with GET\_HOMOLOGUES-EST, a clustering solution for sequences of the same species. *Front Plant Sci*. 2017 Feb 14;8:184. doi: 10.3389/fpls.2017.00184.
- Daly P et al. Expression of *Aspergillus niger* CAZymes is determined by compositional changes in wheat straw generated by hydrothermal or ionic liquid pretreatments. *Biotechnol Biofuels*. 2017 Feb 7;10:35. doi: 10.1186/s13068-017-0700-9.
- de Raad M et al. OpenMSI arrayed analysis toolkit: Analyzing spatially defined samples using mass spectrometry imaging. *Anal Chem*. 2017 Jun 6;89(11):5818-5823. doi: 10.1021/acs.analchem.6b05004. Epub 2017 May 15.

- de Rond T et al. Oxidative cyclization of prodigiosin by an alkylglycerol monooxygenase-like enzyme. *Nat Chem Biol*. 2017 Nov;13(11):1155-1157. doi: 10.1038/nchembio.2471. Epub 2017 Sep 11.
- de Vries RP et al. Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*. *Genome Biol*. 2017 Feb 14;18(1):28. doi: 10.1186/s13059-017-1151-0.
- Díaz-Cárdenas C et al. Draft genome and description of *Consotaella salsifontis* gen. nov. sp. nov., a halophilic, free-living, nitrogen-fixing alphaproteobacterium isolated from an ancient terrestrial saline spring. *Int J Syst Evol Microbiol*. 2017 Oct;67(10):3744-3751. doi: 10.1099/ijsem.0.002185. Epub 2017 Sep 7.
- Dinh Thi VH et al. Recreating stable *Brachypodium hybridum* allotetraploids by uniting the divergent genomes of *B. distachyon* and *B. stacei*. *PLoS One*. 2016 Dec 9;11(12):e0167171. doi: 10.1371/journal.pone.0167171.
- Doré J et al. The ectomycorrhizal basidiomycete *Hebeloma cylindrosporum* undergoes early waves of transcriptional reprogramming prior to symbiotic structures differentiation. *Environ Microbiol*. 2017 Mar;19(3):1338-1354. doi: 10.1111/1462-2920.13670. Epub 2017 Feb 6.
- Doud DFR et al. Novel approaches in function-driven single-cell genomics. *FEMS Microbiol Rev*. 2017 Jul 1;41(4):538-548. 2017 Jun 7. doi: 10.1093/femsre/ful009.
- D'Souza M et al. Strategic integration of multiple bioinformatics resources for system level analysis of biological networks. *Methods Mol Biol*. 2017;1613:85-99. doi: 10.1007/978-1-4939-7027-8\_5.
- Eudes A et al. Exploiting members of the BAHD acyltransferase family to synthesize multiple hydroxycinnamate and benzoate conjugates in yeast. *Microb Cell Fact*. 2016 Nov 21;15(1):198.
- Fernández-González AJ et al. The rhizosphere microbiome of burned holm-oak: Potential role of the genus *Arthrobacter* in the recovery of burned soils. *Sci Rep*. 2017 Jul 20;7(1):6008. doi: 10.1038/s41598-017-06112-3.
- Fochi V et al. Fungal and plant gene expression in the *Tulasnella calospora*-*Serapias vomeracea* symbiosis provides clues about nitrogen pathways in orchid mycorrhizas. *New Phytol*. 2017 Jan;213(1):365-379. doi: 10.1111/nph.14279. Epub 2016 Nov 11.
- Frazier TP et al. Identification, characterization, and gene expression analysis of nucleotide binding site (NB)-type resistance gene homologues in switchgrass. *BMC Genomics*. 2016 Nov 8;17(1):892.
- Frindte K et al. Draft genome sequences of two gammaproteobacterial methanotrophs isolated from rice ecosystems. *Genome Announc*. 2017 Aug 17;5(33). pii: e00526-17. doi: 10.1128/genomeA.00526-17.
- Gach PC et al. Droplet microfluidics for synthetic biology. *Lab Chip*. 2017 Oct 11;17(20):3388-3400. doi: 10.1039/c7lc00576h. Epub 2017 Aug 18.
- Ganji R et al. High-quality draft genome sequence of *Thermocrinis jamiesonii* GBS1T isolated from Great Boiling Spring, Nevada. *Genome Announc*. 2016 Oct 20;4(5). pii: e01112-16. doi: 10.1128/genomeA.01112-16.
- Gaskell J et al. Draft genome sequence of a monokaryotic model brown-rot fungus *Postia (Rhodonina) placenta* SB12. *Genom Data*. 2017 Aug 10;14:21-23. doi: 10.1016/j.gdata.2017.08.003. eCollection 2017 Dec.
- Gontang EA et al. Major changes in microbial diversity and community composition across gut sections of a juvenile *Panochlora* cockroach. *PLoS One*. 2017 May 18;12(5):e0177189. doi: 10.1371/journal.pone.0177189. eCollection 2017.
- Gregory AC et al. Genomic differentiation among wild cyanophages despite widespread horizontal gene transfer. *BMC Genomics*. 2016 Nov 16;17(1):930.
- Gupta VK et al. Biotechnological innovations through fungi. *Mycosphere*. 2016. 7 (10), 1490, doi: 10.5943/mycosphere/si/3b/11.
- Gutierrez T et al. Genome sequence of *Oceanicola* sp. strain MCTG156(1a), isolated from a Scottish coastal phytoplankton net sample. *Genome Announc*. 2017 Aug 10;5(32). pii: e00796-17. doi: 10.1128/genomeA.00796-17.

- Gutierrez T et al. Genome sequence of *Roseovarius* sp. strain MCTG156(2b) isolated from a phytoplankton net trawl on the Scottish west coast. *Genome Announc.* 2017 Aug 31;5(35). pii: e00837-17. doi: 10.1128/genomeA.00837-17.
- Hadjithomas M et al. IMG-ABC: New features for bacterial secondary metabolism analysis and targeted biosynthetic gene cluster discovery in thousands of microbial genomes. *Nucleic Acids Res.* 2017 Jan 4;45(D1):D560-D565. doi: 10.1093/nar/gkw1103. Epub 2016 Nov 29.
- Hahnke RL et al. Genome-based taxonomic classification of *Bacteroidetes*. *Front Microbiol.* 2016 Dec 20;7:2003. doi: 10.3389/fmicb.2016.02003.
- Haitjema CH et al. A parts list for fungal cellulosomes revealed by comparative genomics. *Nat Microbiol.* 2017 May 30;2:17087. doi: 10.1038/nmicrobiol.2017.87.
- Hamilton JJ et al. Metabolic network analysis and metatranscriptomics reveal auxotrophies and nutrient sources of the cosmopolitan freshwater microbial lineage acl. *mSystems.* 2017 Aug 29;2(4). pii: e00091-17. doi: 10.1128/mSystems.00091-17. eCollection 2017 Jul–Aug.
- Hartman WH et al. A genomic perspective on stoichiometric regulation of soil carbon cycling. *ISME J.* 2017 Dec;11(12):2652-2665. doi: 10.1038/ismej.2017.115. Epub 2017 Jul 21.
- Hausmann B et al. Consortia of low-abundance bacteria drive sulfate reduction-dependent degradation of fermentation products in peat soil microcosms. *ISME J.* 2016 Oct;10(10):2365-75. doi: 10.1038/ismej.2016.42.
- He S et al. Ecophysiology of freshwater *Verrucomicrobia* inferred from metagenome-assembled genomes. *mSphere.* 2017 Sep 27;2(5). pii: e00277-17. doi: 10.1128/mSphere.00277-17. eCollection 2017 Sep–Oct.
- Holmfeldt K et al. Large-scale maps of variable infection efficiencies in aquatic *Bacteroidetes* phage-host model systems. *Environ Microbiol.* 2016 Nov;18(11):3949–3961. doi: 10.1111/1462-2920.13392.
- Hossain MS et al. Divergent cytosine DNA methylation patterns in single-cell, soybean root hairs. *New Phytol.* 2017 Apr;214(2):808-819. doi: 10.1111/nph.14421. Epub 2017 Jan 20.
- Hsia MM et al. Sequencing and functional validation of the JGI *Brachypodium distachyon* T-DNA collection. *Plant J.* 2017 Aug;91(3):361–370. doi: 10.1111/tpj.13582. Epub 2017 Jun 21.
- Hu P et al. Simulation of Deepwater Horizon oil plume reveals substrate specialization within a complex community of hydrocarbon degraders. *Proc Natl Acad Sci U S A.* 2017 Jul 11;114(28):7432-7437. doi: 10.1073/pnas.1703424114. Epub 2017 Jun 26.
- Huang P et al. Sparse panicle1 is required for inflorescence development in *Setaria viridis* and maize. *Nat Plants.* 2017 Apr 18;3:17054. doi:10.1038/nplants.2017.54.
- Hutt LP et al. Permanent draft genome of *Thiobacillus thioparus* DSM 505(T), an obligately chemolithoautotrophic member of the Betaproteobacteria. *Stand Genomic Sci.* 2017 Jan 19;12:10. doi: 10.1186/s40793-017-0229-3.
- Hwang C et al. Complete genome sequence of *Alkaliphilus metalliredigens* strain QYMF, an alkaliphilic and metal-reducing bacterium isolated from borax-contaminated leachate ponds. *Genome Announc.* 2016 Nov 3;4(6). pii: e01226-16. doi: 10.1128/genomeA.01226-16.
- Jay ZJ et al. The distribution, diversity, and function of predominant thermoproteales in high-temperature environments of Yellowstone National Park. *Environ Microbiol.* 2016 Dec;18(12):4755-4769. doi: 10.1111/1462-2920.13366.
- Jennings RM et al. Integration of metagenomic and stable carbon isotope evidence reveals the extent and mechanisms of carbon dioxide fixation in high-temperature microbial communities. *Front Microbiol.* 2017 Feb 3;8:88. doi: 10.3389/fmicb.2017.00088. eCollection 2017.
- Jiao JY et al. Complete genome sequence of *Jiangella gansuensis* strain YIM 002(T) (DSM 44835(T)), the type species of the genus *Jiangella* and source of new antibiotic compounds. *Stand Genomic Sci.* 2017 Feb 3;12:21. doi: 10.1186/s40793-017-0226-6.
- Jiménez DJ et al. Draft genome sequence of *Coniochaeta ligniaria* NRRL 30616, a lignocellulolytic fungus for bioabatement of inhibitors in plant biomass hydrolysates. *Genome Announc.* 2017 Jan 26;5(4). pii: e01476-16. doi: 10.1128/genomeA.01476-16.

- Jungbluth SP et al. Genomic comparisons of a bacterial lineage that inhabits both marine and terrestrial deep subsurface systems. *PeerJ*. 2017 Apr 6;5:e3134. doi: 10.7717/peerj.3134. eCollection 2017.
- Kamke J et al. Gene and transcript abundances of bacterial type III secretion systems from the rumen microbiome are correlated with methane yield in sheep. *BMC Res Notes*. 2017 Aug 8;10(1):367. doi: 10.1186/s13104-017-2671-0.
- Kamke J et al. Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a Sharpea-enriched microbiome characterised by lactic acid formation and utilisation. *Microbiome*. 2016 Oct 19;4(1):56.
- Kantor RS et al. Genome-resolved meta-omics ties microbial dynamics to process performance in biotechnology for thiocyanate degradation. *Environ Sci Technol*. 2017 Mar 7;51(5):2944-2953. doi: 10.1021/acs.est.6b04477. Epub 2017 Feb 13.
- Khdhiri M et al. The tale of a neglected energy source: Elevated hydrogen exposure affects both microbial diversity and function in soil. *Appl Environ Microbiol*. 2017 Jun 1;83(11). pii: e00275-17. doi: 10.1128/AEM.00275-17. Epub 2017 May 17.
- Klonowska A et al. High-quality draft genome sequence of *Rhizobium mesoamericanum* strain STM6155, a *Mimosa pudica* microsymbiont from New Caledonia. *Stand Genomic Sci*. 2017 Jan 17;12:7. doi: 10.1186/s40793-016-0212-4.
- Kublanov IV et al. Genomic analysis of *Caldithrix abyssi*, the thermophilic anaerobic bacterium of the novel bacterial phylum Caldithrichaeota. *Front Microbiol*. 2017 Feb 20;8:195. doi: 10.3389/fmicb.2017.00195.
- Lam KK et al. BIGMAC: Breaking inaccurate genomes and merging assembled contigs for long read metagenomic assembly. *BMC Bioinformatics*. 2016 Oct 28;17(1):435.
- Lamit LJ et al. Patterns and drivers of fungal community depth stratification in Sphagnum peat. *FEMS Microbiol Ecol*. 2017 Jul 1;93(7). doi:10.1093/femsec/fix082.
- Lastovetsky OA et al. Lipid metabolic changes in an early divergent fungus govern the establishment of a mutualistic symbiosis with endobacteria. *Proc Natl Acad Sci U S A*. 2016 Dec 27;113(52):15102-15107. doi: 10.1073/pnas.1615148113.
- Laviad-Shitrit S et al. High quality permanent draft genome sequence of *Chryseobacterium bovis* DSM 19482T, isolated from raw cow milk. *Stand Genomic Sci*. 2017 May 8;12:31. doi: 10.1186/s40793-017-0242-6. eCollection 2017.
- Lee CR et al. Young inversion with multiple linked QTLs under selection in a hybrid zone. *Nat. Ecol. Evol*. 2017 Apr 3;1:119. doi: 10.1038/s41559-017-0119.
- Li G et al. The sequences of 1,504 mutants in the model rice variety *Kitaake* facilitate rapid functional genomic studies. *Plant Cell*. 2017 Jun 2. pii: tpc.00154.2017. doi: 10.1105/tpc.17.00154.
- Li H et al. Lignocellulose pretreatment in a fungus-cultivating termite. *Proc Natl Acad Sci U S A*. 2017 May 2;114(18):4709-4714. doi: 10.1073/pnas.1618360114. Epub 2017 Apr 19.
- Lux M et al. acdc — Automated Contamination Detection and Confidence estimation for single-cell genome data. *BMC Bioinformatics*. 2016 Dec 20;17(1):543. doi: 10.1186/s12859-016-1397-7.
- Magain N et al. Conserved genomic collinearity as a source of broadly applicable, fast evolving, markers to resolve species complexes: a case study using the lichen-forming genus *Peltigera* section Polydactylon. *Mol Phylogenet Evol*. 2017 Dec;117:10-29. doi: 10.1016/j.ympev.2017.08.013. Epub 2017 Aug 30.
- Mansueto L et al. Rice SNP-seek database update: new SNPs, indels, and queries. *Nucleic Acids Res*. 2017 Jan 4;45(D1):D1075-D1081. doi: 10.1093/nar/gkw1135. Epub 2016 Nov 29.
- Masumba EA et al. QTL associated with resistance to cassava brown streak and cassava mosaic diseases in a bi-parental cross of two Tanzanian farmer varieties, Namikonga and Albert. *Theor Appl Genet*. 2017 Oct;130(10):2069-2090. doi: 10.1007/s00122-017-2943-z. Epub 2017 Jul 13.

- Matrosova VY et al. High-quality genome sequence of the radioresistant bacterium *Deinococcus ficus* KS 0460. *Stand Genomic Sci.* 2017 Jul 28;12:46. doi: 10.1186/s40793-017-0258-y. eCollection 2017.
- Melton ED et al. Draft genome sequence of *Dethiobacter alkaliphilus* strain AHT1T, a gram-positive sulfidogenic polyextremophile. *Stand Genomic Sci.* 2017 Sep 21;12:57. doi: 10.1186/s40793-017-0268-9. eCollection 2017.
- Men Y et al. Metagenomic and metatranscriptomic analyses reveal structure and dynamics of a dechlorinating community containing *Dehalococcoides mccartyi* and corrinoid-providing microorganisms under cobalamin-limited condition. *Appl Environ Microbiol.* 2017 Mar 31;83(8). pii: e03508-16. doi: 10.1128/AEM.03508-16. Print 2017 Apr 15.
- Meyer KM et al. Conversion of Amazon rainforest to agriculture alters community traits of methane-cycling organisms. *Mol Ecol.* 2017 Mar;26(6):1547-1556. doi: 10.1111/mec.14011.
- Min B et al. FunGAP: Fungal Genome Annotation Pipeline Using Evidence-based Gene Model Evaluation. *Bioinformatics.* 2017 Sep 15;33(18):2936-2937. doi: 10.1093/bioinformatics/btx353. Epub 2017 Jun 5.
- Miroshnikov KK et al. Draft Genome Sequence of *Methylocapsa palarum* NE2T, an Obligate Methanotroph from Subarctic Soil. *Genome Announc.* 2017 Jun 15;5(24). pii: e00504-17. doi: 10.1128/genomeA.00504-17.
- Mock T et al. Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature.* 2017 Jan 26;541(7638):536-540. doi: 10.1038/nature20803.
- Moghaddam SM et al. Genome-Wide Association Study Identifies Candidate Loci Underlying Agronomic Traits in a Middle American Diversity Panel of Common Bean. *Plant Genome.* 2016 Nov;9(3). doi: 10.3835/plantgenome2016.02.0012.
- Mondo SJ et al. Widespread adenine N6-methylation of active genes in fungi. *Nat Genet.* 2017 Jun;49(6):964-968. doi: 10.1038/ng.3859. Epub 2017 May 8.
- Moraes LE et al. Resequencing and annotation of the *Nostoc punctiforme* ATTC 29133 genome: facilitating biofuel and high-value chemical production. *AMB Express.* 2017 Dec;7(1):42. doi: 10.1186/s13568-017-0338-9.
- Morales J et al. Development of a toolbox to dissect host-endosymbiont interactions and protein trafficking in the trypanosomatid *Angomonas deanei*. *BMC Evol Biol.* 2016 Nov 11;16(1):247.
- Morrell WC et al. The Experiment Data Depot: A Web-Based Software Tool for Biological Experimental Data Storage, Sharing, and Visualization. *ACS Synth Biol.* 2017 Sep 8. doi: 10.1021/acssynbio.7b00204. [Epub ahead of print]
- Mujic AB et al. Comparative Genomics of the Ectomycorrhizal Sister Species *Rhizopogon vinicolor* and *Rhizopogon vesiculosus* (Basidiomycota: Boletales) Reveals a Divergence of the Mating Type B Locus. *G3 (Bethesda).* 2017 Jun 7;7(6):1775-1789. doi: 10.1534/g3.117.039396.
- Mukai T et al. Transfer RNAs with novel cloverleaf structures. *Nucleic Acids Res.* 2017 Mar 17;45(5):2776-2785. pii: gkw898. doi: 10.1093/nar/gkw898. Epub 2016 Oct 5.
- Mukai T et al. RNA-Dependent Cysteine Biosynthesis in Bacteria and Archaea. *MBio.* 2017 May 9;8(3). pii: e00561-17. doi: 10.1128/mBio.00561-17.
- Mukherjee S et al. Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. *Nucleic Acids Res.* 2017 Jan 4;45(D1):D446-D456. doi: 10.1093/nar/gkw992. Epub 2016 Oct 27.
- Mukherjee S et al. 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nat Biotechnol.* 2017 Jul;35(7):676-683. doi: 10.1038/nbt.3886. Epub 2017 Jun 12.
- Myers CJ et al. A standard-enabled workflow for synthetic biology. *Biochem Soc Trans.* 2017 Jun 15;45(3):793-803. doi: 10.1042/BST20160347.
- Nagy LG et al. Genetic bases of fungal white rot wood decay predicted by phylogenomic analysis of correlated gene-phenotype evolution. *Mol Biol Evol.* 2017 Jan;34(1):35-44. doi: 10.1093/molbev/msw238. Epub 2016 Nov 10.
- Nouioui I et al. High-quality draft genome of *Nakamurella lactea* type strain, a rock actinobacterium, and emended description of *Nakamurella lactea*. *Stand Genomic Sci.* 2017 Jan 6;12:4. doi: 10.1186/s40793-016-0216-0.

- Nzuki I et al. QTL Mapping for Pest and Disease Resistance in Cassava and Coincidence of Some QTL with Introgression Regions Derived from *Manihot glaziovii*. *Front Plant Sci.* 2017 Jul 21;8:1168. doi: 10.3389/fpls.2017.01168. eCollection 2017.
- Oberortner E et al. Streamlining the Design-to-Build Transition with Build-Optimization Software Tools. *ACS Synth Biol.* 2017 Mar 17;6(3):485-496. doi: 10.1021/acssynbio.6b00200. Epub 2016 Dec 22.
- Olm MR et al. The Source and Evolutionary History of a Microbial Contaminant Identified Through Soil Metagenomic Analysis. *MBio.* 2017 Feb 21;8(1). pii: e01969-16. doi: 10.1128/mBio.01969-16.
- Osman WAM et al. High-quality draft genome sequence of *Ensifer meliloti Mlalz-1*, a microsymbiont of *Medicago laciniata* (L.) miller collected in Lanzarote, Canary Islands, Spain. *Stand Genomic Sci.* 2017 Sep 25;12:58. doi: 10.1186/s40793-017-0270-2. eCollection 2017.
- Ovchinnikov S et al. Protein structure determination using metagenome sequence data. *Science.* 2017 Jan 20;355(6322):294-298. doi: 10.1126/science.aah4043.
- Paez-Espino D et al. Nontargeted virus sequence discovery pipeline and virus clustering for metagenomic data. *Nat Protoc.* 2017 Aug;12(8):1673-1682. doi: 10.1038/nprot.2017.063. Epub 2017 Jul 27.
- Paez-Espino D et al. IMG/VR: a database of cultured and uncultured DNA Viruses and retroviruses. *Nucleic Acids Res.* 2017 Jan 4;45(D1):D457-D465. doi: 10.1093/nar/gkw1030. Epub 2016 Oct 30.
- Patil Y et al. High-quality-draft genome sequence of the fermenting bacterium *Anaerobium acetethylicum* type strain GluBS11(T) (DSM 29698). *Stand Genomic Sci.* 2017 Feb 20;12:24. doi: 10.1186/s40793-017-0236-4.
- Pavlopoulos GA et al. Empirical Comparison of Visualization Tools for Larger-Scale Network Analysis. *Adv Bioinformatics.* 2017 Jul 18;2017:1278932. doi: 10.1155/2017/1278932.
- Peng M et al. The draft genome sequence of the ascomycete fungus *Penicillium subrubescens* reveals a highly enriched content of plant biomass related CAZymes compared to related fungi. *J Biotechnol.* 2017 Mar 20;246:1-3. doi: 10.1016/j.jbiotec.2017.02.012. Epub 2017 Feb 16.
- Proença DN et al. Draft genome sequence of the cellulolytic endophyte *Chitinophaga costaii* A37T2T. *Stand Genomic Sci.* 2017 Sep 6;12:53. doi: 10.1186/s40793-017-0262-2. eCollection 2017.
- Rabe F et al. A complete toolset for the study of *Ustilago bromivora* and *Brachypodium* sp. as a fungal-temperate grass pathosystem. *Elife.* 2016 Nov 11;5. pii: e20522. doi: 10.7554/eLife.20522.
- Raissig MT et al. Mobile MUTE specifies subsidiary cells to build physiologically improved grass stomata. *Science.* 2017 Mar 17;355(6330):1215-1218. doi: 10.1126/science.aal3254. Epub 2017 Mar 16.
- Reeve W et al. High-quality permanent draft genome sequence of the *Bradyrhizobium elkanii* type strain USDA 76(T), isolated from *Glycine max* (L.) Merr. *Stand Genomic Sci.* 2017 Mar 4;12:26. doi: 10.1186/s40793-017-0238-2.
- Rice MC et al. Complete Genome Sequence of *Nitrosomonas cryotolerans* ATCC 49181, a Phylogenetically Distinct Ammonia-Oxidizing Bacterium Isolated from Arctic Waters. *Genome Announc.* 2017 Mar 16;5(11). pii: e00011-17. doi: 10.1128/genomeA.00011-17.
- Rineau F et al. Comparative genomics and expression levels of hydrophobins from eight mycorrhizal genomes. *Mycorrhiza.* 2017 May;27(4):383-396. doi: 10.1007/s00572-016-0758-4. Epub 2017 Jan 9.
- Robicheau BM et al. The complete mitochondrial genome of the conifer needle endophyte, *Phialocephala scopiformis* DAOMC 229536 confirms evolutionary division within the fungal *Phialocephala fortinii* s.l. - *Acephala appalana* species complex. *Fungal Biol.* 2017 Mar;121(3):212-221. doi: 10.1016/j.funbio.2016.11.007.
- Roux S et al. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ.* 2017 Sep 21;5:e3817. doi: 10.7717/peerj.3817. eCollection 2017.
- Schulz F et al. Giant viruses with an expanded complement of translation system components. *Science.* 2017 Apr 7;356(6333):82-85. doi:10.1126/science.aa14657.
- Schwander T et al. A synthetic pathway for the fixation of carbon dioxide *in vitro*. *Science.* 2016 Nov 18;354(6314):900-904.

- Session AM et al. Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature*. 2016 Oct 20;538(7625):336-343. doi: 10.1038/nature19840.
- Shih PM et al. A robust gene-stacking method utilizing yeast assembly for plant synthetic biology. *Nat Commun*. 2016 Oct 26;7:13215. doi: 10.1038/ncomms13215.
- Silas S. On the Origin of Reverse Transcriptase-Using CRISPR-Cas Systems and Their Hyperdiverse, Enigmatic Spacer Repertoires. *MBio*. 2017 Jul 11;8(4). pii: e00897-17. doi: 10.1128/mBio.00897-17.
- Simon L et al. *Paenibacillus aquistagni* sp. nov., isolated from an artificial lake accumulating industrial wastewater. *Antonie Van Leeuwenhoek*. 2017 Sep;110(9):1189-1197. doi: 10.1007/s10482-017-0891-x. Epub 2017 May 29.
- Singer E et al. Capturing the genetic makeup of the active microbiome *in situ*. *ISME J*. 2017 Sep;11(9):1949-1963. doi: 10.1038/ismej.2017.59. Epub 2017 Jun 2
- Škraban J et al. Draft genome sequence of *Chryseobacterium limigenitum* SUR2T (LMG 28734T) isolated from dehydrated sludge. *Braz J Microbiol*. 2017 Jul 18. pii: S1517-8382(16)31337-5. doi: 10.1016/j.bjbm.2017.03.009. [Epub ahead of print]
- Spatafora JW et al. The Fungal Tree of Life: from Molecular Systematics to Genome-Scale Phylogenies. *Microbiol Spectr*. 2017 Sep;5(5). doi: 10.1128/microbiolspec.FUNK-0053-2016.
- Spatafora JW et al. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia*. 2016 Sep-Oct;108(5):1028-1046.
- Staley C et al. Diurnal cycling of rhizosphere bacterial communities is associated with shifts in carbon metabolism. *Microbiome*. 2017 Jun 24;5(1):65. doi: 10.1186/s40168-017-0287-1.
- Strepis N et al. Description of *Trichococcus iylis* sp. nov. by combined physiological and in silico genome hybridization analyses. *Int J Syst Evol Microbiol*. 2016 Oct;66(10):3957-3963. doi: 10.1099/ijsem.0.001294.
- Susanti D et al. Permanent Draft Genome Sequence of *Desulfurococcus amylolyticus* Strain Z-533T, a Peptide and Starch Degrader Isolated from Thermal Springs in the Kamchatka Peninsula and Kunashir Island, Russia. *Genome Announc*. 2017 Apr 13;5(15). pii: e00078-17 doi: 10.1128/genomeA.00078-17.
- Suzuki A et al. Genome organization of the vg1 and nodal3 gene clusters in the allotetraploid frog *Xenopus laevis*. *Dev Biol*. 2017 Jun 15;426(2):236-244. doi: 10.1016/j.ydbio.2016.04.014. Epub 2016 Oct 5.
- Theodosiou T et al. NAP: The Network Analysis Profiler, a web tool for easier topological analysis and comparison of medium-scale biological networks. *BMC Res Notes*. 2017 Jul 14;10(1):278. doi: 10.1186/s13104-017-2607-8.
- Thieme N et al. The transcription factor PDR-1 is a multi-functional regulator and key component of pectin deconstruction and catabolism in *Neurospora crassa*. *Biotechnol Biofuels*. 2017 Jun 12;10:149. doi: 10.1186/s13068-017-0807-z. eCollection 2017.
- Tian R et al. Evolution of a multi-step phosphorelay signal transduction system in *Ensifer*: Recruitment of the sigma factor RpoN and a novel enhancer binding protein triggers acid-activated gene expression. *Mol Microbiol*. 2017 Mar;103(5):829-844. doi: 10.1111/mmi.13592. Epub 2016 Dec 9.
- Tighe S et al. Genomic Methods and Microbiological Technologies for Profiling Novel and Extreme Environments for the Extreme Microbiome Project (XMP). *J Biomol Tech*. 2017 Apr;28(1):31-39. doi: 10.7171/jbt.17-2801-004. Epub 2017 Mar 10.
- Tringe SG. Single-cell genomics for the masses. *Nat Biotechnol*. 2017 Jul 12;35(7):635-636. doi: 10.1038/nbt.3914.
- Tsesmetzis N et al. MixS-HCR: a MixS extension defining a minimal information standard for sequence data from environments pertaining to hydrocarbon resources. *Stand Genomic Sci*. 2016 Oct 12;11:78.
- Turco GM et al. DNA methylation and gene expression regulation associated with vascularization in *Sorghum bicolor*. *New Phytol*. 2017 May;214(3):1213-1229. doi: 10.1111/nph.14448. Epub 2017 Feb 10.
- Twing KI et al. Serpentinization-Influenced Groundwater Harbors Extremely Low Diversity Microbial Communities Adapted to High pH. *Front Microbiol*. 2017 Mar 1;8:308. doi: 10.3389/fmicb.2017.00308. eCollection 2017.
- Uehling J et al. Comparative genomics of *Mortierella elongata* and its bacterial endosymbiont *Mycocoidus cysteinexigens*. *Environ Microbiol*. 2017 Aug;19(8):2964-2983. doi: 10.1111/1462-2920.13669. Epub 2017 Feb 1.

Verde I et al. The Peach v2.0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. *BMC Genomics*. 2017 Mar 11;18(1):225. doi: 10.1186/s12864-017-3606-9.

Vigneron A et al. Succession in the petroleum reservoir microbiome through an oil field production lifecycle. *ISME J*. 2017 Sep;11(9):2141-2154. doi: 10.1038/ismej.2017.78. Epub 2017 May 19.

Waite DW et al. Comparative Genomic Analysis of the Class Epsilonproteobacteria and Proposed Reclassification to Epsilonbacteraeota (phyl. nov.). *Front Microbiol*. 2017 Apr 24;8:682. doi: 10.3389/fmicb.2017.00682. eCollection 2017.

Wang L et al. Interactions among triphenyltin degradation, phospholipid synthesis and membrane characteristics of *Bacillus thuringiensis* in the presence of d-malic acid. *Chemosphere*. 2017 Feb;169:403-412. doi: 10.1016/j.chemosphere.2016.10.140. Epub 2016 Nov 22.

Warshan D et al. Feathermoss and epiphytic *Nostoc* cooperate differently: expanding the spectrum of plant-cyanobacteria symbiosis. *ISME J*. 2017 Dec;11(12):2821-2833. doi: 10.1038/ismej.2017.134. Epub 2017 Aug 11.

Woo HL et al. Near-Complete Genome Sequence of *Thalassospira* sp. Strain KO164 Isolated from a Lignin-Enriched Marine Sediment Microcosm. *Genome Announc*. 2016 Nov 23;4(6). pii: e01297-16. doi: 10.1128/genomeA.01297-16.

Woods D et al. Genetic architecture of flowering-time variation in *Brachypodium distachyon*. *Plant Physiol*. 2017 Jan;173(1):269-279. pii: pp.01178.2016. Epub 2016 Oct 14.

Wu W et al. Characterization of four endophytic fungi as potential consolidated bioprocessing hosts for conversion of lignocellulose into advanced biofuels. *Appl Microbiol Biotechnol*. 2017 Mar;101(6):2603-2618. doi: 10.1007/s00253-017-8091-1.

Xiong Y et al. A fungal transcription factor essential for starch degradation affects integration of carbon and nitrogen metabolism. *PLoS Genet*. 2017 May 3;13(5):e1006737. doi: 10.1371/journal.pgen.1006737. eCollection 2017 May.

Yi W et al. Triphenyltin degradation and proteomic response by an engineered *Escherichia coli* expressing cytochrome P450 enzyme. *Ecotoxicol Environ Saf*. 2017 Mar;137:29-34. doi: 10.1016/j.ecoenv.2016.11.012. Epub 2016 Nov 28.

Yu FB et al. Microfluidic-based mini-metagenomics enables discovery of novel microbial lineages from complex environmental samples. *Elife*. 2017 Jul 5;6. pii: e26580. doi: 10.7554/eLife.26580.



**Left:** Estelle Schaefer.

**Facing page, Top row:**

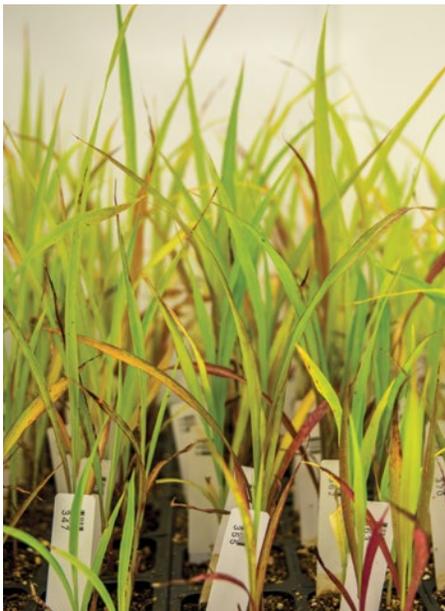
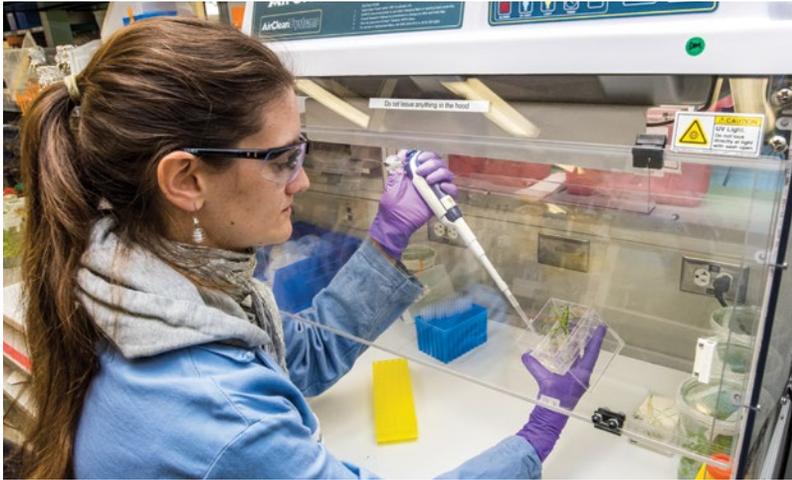
Joelle Schlapfer;

Middle row (left to right):

Hope Hundley and Jenifer

Johnson; Bottom row:

Peter Andeer and Peter Kim.



## Comments?

Massie S. Ballon, JGI Communications & Outreach  
mlballon@lbl.gov

David Gilbert, Senior Manager, JGI Communications & Outreach  
degilbert@lbl.gov

## Follow Us



NASA astronaut and microbiologist Kate Rubins, the first woman to sequence DNA in space, visited Berkeley Lab on May 11, 2017 and gave a talk about "Science in Extreme Environments." As her talk was hosted by the Molecular Foundry and the JGI, she came out to Walnut Creek later that day. **From left to right:** CASIS associate program scientist Liz Warren, JGI User Programs Deputy Susannah Tringe, NASA astronaut Kate Rubins, JGI Director Nigel Mouncey, JGI Science Programs Deputy Axel Visel, and JGI Genomic Technologies Deputy Len Pennacchio.



Ongoing construction of the Integrative Genomics Building (IGB) as of January 2018, future home of the JGI and KBase at Berkeley Lab.

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.



A U.S. Department of Energy  
National Laboratory  
Managed by the University of California



U.S. DEPARTMENT OF  
**ENERGY**

Office of Science



**Bio**Sciences