

Cover Photo:

Located east of Yosemite National Park, Mono Lake has been referred to as "California's Dead Sea" for its alkaline waters. Microbes isolated from Mono Lake have been sequenced and analyzed by the JGI to help understand how these organisms have adapted to thrive where oxygen-rich waters provided by freshwater springs interface with the oxygen poor and salty waters of the lake. (Jon Bertsch)

Impact Section Case Study credits (clockwise from top left): Fruiting bodies of *L. bicolor* colonizing seedlings of Douglas fir (photograph courtesy of *D. Vairelles*); Soybean helix (Roy Kaltschmidt, Berkeley Lab); SEM of wood being decayed by the white rot fungus Punctularia strigoso-zonata (Robert Blanchette, University of Minnesota); Ivotuk range, Alaska (LANL–Cathy Wilson); Fistulated cow (Jonas Løvaas Gjerstad); poplar leaf (David Gilbert, JGI).







1	JGI Mission
2	Director's Perspective
8	Achieving the DOE Mission
10	Organizational Structure
12	Impact 2018
18	Case Study: JGI's Community Science Program at Fifteen
20	Science: A Year in Review
21	Discovery
30	Bioenergy
46	Biogeochemistry
54	Computational Infrastructure
56	Appendices
57	Appendix A: Acronyms at a Glance
59	Appendix B: Glossary
62	Appendix C: 2018 User Program Supported Proposals
68	Appendix D: Advisory and Review Committee Members
70	Appendix E: 13th Annual Genomics of Energy and Environment Meeting
74	Appendix F: 2018 Publications







Vision

The vision of the U.S. Department of Energy (DOE) Joint Genome Institute (JGI) is to become the leading integrative genome science user facility enabling researchers to solve the world's evolving energy and environmental challenges.

Mission

The mission of the JGI is to provide the global research community with access to the most advanced integrative genome science capabilities in support of the DOE's research mission.



Nigel Mouncey, Director, DOE Joint Genome Institute

Director's **Perspective**





In 2018, the U.S. Department of Energy (DOE) Joint Genome Institute (JGI) made unprecedented progress towards realizing its vision as the leading integrative genome science user facility enabling researchers to solve the world's evolving energy and environmental challenges. We sustained our commitment to provide the global research community with access to unrivaled research infrastructure in support of the DOE's mission. Bringing together our core capabilities in DNA sequencing and synthesis, high-performance computing, and metabolomics, and building upon our multidisciplinary partnerships with other user facilities has strengthened the JGI's position to make impactful, sustainable contributions in support of the nation's innovation ecosystem, examples of which I have summarized below.

Over the last year, the JGI continued to grow its productive user and science programs in alignment with its scientific directions. Following our successful Triennial Science & Operations Review in December 2017, the JGI embarked on developing a new strategic plan that is more responsive to the needs of the research community and better positions us to assert leadership in emerging opportunities. Today's pace of sequence data generation outweighs our abilities to ascribe function and derive biological insights. Major breakthroughs in advancing our understanding of biology are thus only possible through the integrative use of complementary technologies. The JGI is well positioned to address specific challenges in the area of biological systems science, such as plant and microbial metabolism and interaction, engineering of diverse organisms, integrative analysis of complementary data sets, and establishing experimentally-validated links between genotype and phenotype. The JGI's new 5-Year Strategic Plan: Beyond Basepairs — A Vision for Integrative and Collaborative Genome Science, published at the end of January 2019, was developed through internal discussions, input from our Advisory Committees, and experts across our core areas at a workshop in April 2018. The implementation of the mission will be facilitated by a set of overarching guiding principles, applicable to a wide range of scientific activities, platforms and user groups. To focus our efforts and chart progress, a series of 2- and 5-year milestones have been articulated.

As a user facility, the JGI enables the advancement of science by our users and in fiscal year 2018, the JGI served a worldwide community of 1,882 users from academia, government and industry. In support of their projects we produced 53 plant genome annotations and 2,190 resequenced plant genomes, 222 annotated fungal genomes and 2,557 microbial genomes (including single-cells), as well as 1,898 metagenomic data sets. As we continue to advance our functional genomics, we generated 2,994 plant transcriptomic and 692 metatranscriptomic datasets. In total this led the JGI to produce a record sequencing output of 225 trillion bases of sequence data in FY18. Additionally, our DNA synthesis science program produced 510 million bases of DNA constructs for our users. Data from the sequencing projects are made freely accessible to the broader scientific community via the JGI's Genome Portal, as well as adding value through the analysis tools hosted by JGI's Integrated Microbial Genomes and Microbiomes data system, Phytozome for plant genomes, and MycoCosm for fungal genomes. These data

platforms continue to attract new users, with >11,000 new portal users who accessed our data portals in FY18 alone and with more than 100,000 visitors for data access and analysis. This underscores the importance of these portals and associated databases and tools as essential computational infrastructure for the broader energy and environmental research community. We have raised the visibility of JGI's capabilities, expertise and data sets with users from industry and our Industry Engagement Program has had a successful first year under Tootie Tatum's leadership, meeting with dozens of companies and completing three Sponsored Partnership Projects.

The demand for our user programs is increasing. In 2018, we received 94 proposals submitted to the large-scale Community Science Program (CSP), as well as 174 proposals in response to the New Investigator (previously Small-Scale) and DNA Synthesis calls and 49 proposals to the joint Facilities Integrating Collaborations for User Science (FICUS) call, a user program collaboratively offered by the JGI and the Environmental Molecular Sciences Laboratory (EMSL) at Pacific Northwest National Laboratory and, as of 2017, also with the National Energy Research Scientific Computing Center (NERSC) and the DOE Systems Biology Knowledgebase (KBase). Across all of these programs, a total of 102 proposals were approved and accepted after peer review and are now in progress.

The impact of the research performed through these proposals is reflected by the large number of scientific publications originating from the JGI and its users. In 2018, the JGI contributed to 232 peer-reviewed publications, including 34 in *Science* and the *Nature* family of journals. Six of our scientists were included in the annual "Highly Cited Researchers" list by Clarivate Analysis (formerly Thomson Reuters) and several of our scientists serve on editorial boards of leading scientific journals. Nikos Kyrpides was recognized for outstanding science as recipient of ASM's Roger Porter Award. Selected scientific publication highlights in 2018 involving JGI researchers include:

- Liverwort Insights to Land Plant Evolution. The common liverwort has no roots or vascular tissues for nutrient transport. It serves as a living link to the transition from the algae that found its way out of the ocean to the established multitude of land plants. JGI scientists were part of an international team led by researchers at Australia's Monash University, and Japan's Kyoto University and Kindai University that analyzed the genome sequence of the common liverwort (Marchantia polymorpha) to identify genes and gene families that were deemed crucial to plant evolution and have been conserved over millions of years and across plant lineages. (Cell, October 5, 2017)
- Virophages Double Using Freshwater Lake Time-Series. Viruses exist amidst micro- and macro-organisms, usually in a 10-fold excess, and are made up of various sizes ranging from giant viruses, to much smaller viruses known as virophages (which live in giant viruses and use their machinery to replicate and spread.) Using metagenome data sets collected over several years in northern freshwater lakes, a team led by researchers at The Ohio State University and the JGI uncovered 25 novel sequences of virophages. The identification of these novel sequences effectively doubles the number of virophages known since their discovery a decade ago. (Nature Communications, October 11, 2017)
- Humongous Fungus' Size Determined by Breadth of Gene Families. Armillaria fungi are among the most devastating fungal pathogens, causing root rot disease in more than 500 plant species found in forests, parks and vineyards. As white rot fungi, they are capable of breaking down all components of plant cell walls, a capability that interests bioenergy researchers looking for methods to cost-effectively convert plant biomass into alternative fuels. An international team sequenced and analyzed four Armillaria fungi and then compared these genomes with those of related fungi to better understand the evolution of Armillaria's abilities to spread and infect, and effectively break down all components of plant cell walls. (Nature Ecology & Evolution, October 30, 2017)





JGI and KBase staff gathered to sign the ultimate beam before it is hoisted into place on top of the Integrative Genomics Building (IGB), where JGI and KBase will be co-located. (Paul Mueller, Berkeley Lab)

- Microbial Metabolism in Biocrusts. Arid lands cover some 40 percent of the Earth's terrestrial surface and are too dry to sustain much vegetation. In spite of this, the uppermost millimeters of soil, or biocrusts, are home to diverse communities of microorganisms including fungi, bacteria, and archaea. Understanding how microbial communities in the biocrusts adapt to their harsh environments could provide important clues to help shed light on the roles of soil microbes in the global carbon cycle. Working with colleagues at Lawrence Berkeley National Laboratory (Berkeley Lab), JGI researchers reported that specific compounds are transformed by and strongly associated with specific bacteria in biocrust using a suite of tools called "exometabolomics." (Nature Communications, January 2, 2018)
- Reference Catalog of Rumen Microbiome. The digestive tracts of ruminant (cud-chewing) animals such as cattle, sheep, and goats convert lignocellulosic plant matter to short-chain fatty acids used for nourishment with unparalleled efficiency, thanks to the activity of symbiotic microbes in the rumen. Rumen microbes play a vital role in allowing ruminant livestock to break down the food they eat. The process is also the single largest human-influenced source of the greenhouse gas methane. An international team led by New Zealand researchers reported a reference catalog of rumen microbial genomes and isolates, one of the largest targeted cultivation and sequencing projects to date, produced through the coordinated efforts of rumen microbiology researchers worldwide. (Nature Biotechnology, March 19, 2018)
- **Accelerating Experiment-Based Gene Function Assignments. While advances in sequencing technologies have enabled researchers to access the genomes of thousands of microbes and make them publicly available, no similar shift has occurred with the task of assigning functions to the genes uncovered. Working with Berkeley Lab colleagues, JGI researchers helped develop a workflow that enables large scale, genome-wide assays of gene importance across many conditions. The resulting study, considered the largest functional genomics study of bacteria published to date, combined high-throughput genetics and comparative genomics to identify mutant phenotypes for thousands of genes with previously unknown functions. (Nature, May 16, 2018)

- A Tool for Recovering Soil Metagenomes. Through the JGI's Emerging Technologies Opportunity Program, a team led by University of California, Berkeley researchers reported on the development and validation of a dereplication, aggregation and scoring tool (DAS Tool) to optimize binning of metagenome-assembled genomes (MAGs). Using DAS Tool on data from soil samples, researchers were able to reconstruct 79 minimally contaminated (<5%) draft genomes to >70% completeness. Of those 79 genomes, 26 were high-quality draft genomes with >90% completeness. (*Nature Microbiology*, May 28, 2018)
- Reference Sequence for Sugarcane. An international team led by researchers from the French Agricultural Research Centre for International Development (CIRAD) and including JGI scientists worked on sequencing and assembling fragments of sugarcane chromosomes into the first monoploid reference of the gene-rich part of the sugarcane genome. Their approach relied in part on having a sequence for sugarcane's relative sorghum, a JGI Plant Flagship Genome sequence, and knowing that most genes in sorghum occurred roughly in the same order in sugarcane. (Nature Communications, July 6, 2018)



Since 2014, more than 20 UC Merced students have contributed to the research of JGI scientists through a flagship education program.

I appreciate that these foundational contributions to the literature are catalyzed by long-term commitment of resources and steadfast determination on behalf of the JGI staff to support our users' success. Of related importance is the critical role that we play in mentoring the next generation of scientists and engineers. One of the hallmarks of JGI is the diversity of our workforce and that of our collaborators around the world. That said, we recognize that our workforce is still not representative of the rich diversity particularly here in California. In July, the JGI celebrated the culmination of the fifth summer of its flagship education program with University of California, Merced. More than 20 UC Merced students have now contributed to the research of 13 JGI scientists since the program's inception in 2014. JGI's partnership with UC Merced represents a model for building awareness and opportunities for populations of students traditionally not on a trajectory for science careers. The JGI's education outreach efforts extend down into high school as well. Though a partnership with Biotech Partners, a nonprofit that provides underserved youth in the Bay Area with personal, academic and professional development experiences that increase participation in higher education and access to fulfilling science careers, JGI hosted five students from Antioch High School this last summer.

Through our membership in the Society for Science at User Research Facilities (SSURF) — a not-for-profit organization working to advance awareness of user science — I had an opportunity this summer to participate in SSURF's annual meeting, "America's Evolving Scientific Infrastructure." This coincided with my first foray into the halls of Congress which I found both humbling and energizing to meet with those responsible for setting science

policy and funding directions for the nation and representing districts where JGI's users make a tangible impact. We found receptive ears for the examples we provided describing the value that the National Laboratories and User Facilities contribute in driving the nation's innovations and how we enable groundbreaking science.

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. As such, the JGI relies heavily on computational processing and data analysis at all levels of the organization, from project management and the production aspects of sequencing through to the assembly, annotation, and biological analysis of gene and genome function, variation, and evolution. In 2018, we consolidated core components of our computation into a new department, Data Science and Informatics (DSI), led by JGI's Chief Informatics Officer, Kjiersten Fagnan. This department will ensure that all JGI's data are high-quality and accessible (consistent with the FAIR data principles) and that data management systems are designed with user-driven focus. The Advanced Analysis group within DSI has strengthened the partnership between the JGI and KBase to share and integrate genomic and experimental data of increasing complexity, co-develop tools and workflows, and ensure cooperativity between our respective systems by engaging in a series of joint development and outreach activities. This partnership will further flourish upon our co-localization in the new Integrative Genomics Building (IGB) later in 2019. Collectively, the JGI's investments in infrastructure and data management, computational tool development, and forward-thinking approaches to large-scale data mining have paved the way to translate the complexities of the world around us into biological discoveries.

This last year was one marked by a significant transition. For more than a dozen years, Dan Drell served as the JGI program manager in the DOE Office of Biological and Environmental Research (BER). After 30 years in government service, Dan retired in September 2018. We have been the fortunate beneficiaries of his sage guidance in assuring the JGI's sustained fidelity with DOE mission and relevance to emerging fields of scientific inquiry. With the baton now passed along to our new program manager Ramana Madupu, we have already made great strides in improving the integration of JGI's data with KBase's analysis tools and workflows to provide additional value to our research communities.

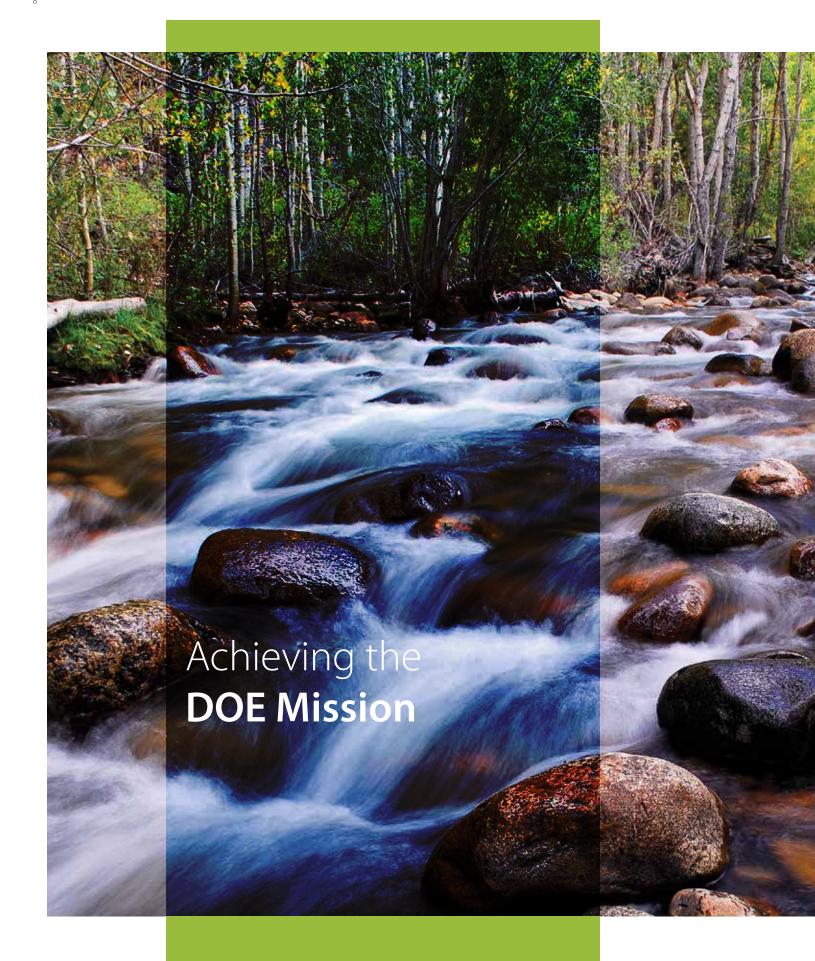
The JGI, as a national user facility, serves a diverse scientific community. We are committed to conducting groundbreaking research in genomic science and technology, focusing on DOE mission-relevant topics. As leaders and team members, we rise to serve as stewards of our nation's investment and are responsible for ensuring that it remains as a valuable national asset. People are our most valuable resource, and the talented staff at the JGI are central to achieving our mission. Our scientific and technical strategies are complemented by an expanded OurJGI initiative, encompassing converging bottom-up and top-down activities to focus efforts on evolving the culture of the JGI to increase inclusivity, safety, skill and career development, teamwork, communication, respect and scientific excellence. This becomes even more central as the JGI makes its move from Walnut Creek to the Berkeley Lab campus and our new home in the Integrative Genomics Building later in 2019. Our move will enable closer interaction with our Berkeley Lab colleagues, and proximity to the Biosciences Area capabilities and expertise. So that we stay apprised of diversity best practices, the JGI's Diversity & Inclusion (D&I) Working Group recruited a 12-person Advisory Board comprised of diversity thought-leaders in academia, industry, and non-profits and convened early in the new fiscal year, at the JGI's first D&I retreat.

The JGI enjoyed yet another highly-productive year in 2018, delivering an essential suite of state-of-the-art capabilities to enable our global community of users to expand the frontiers of energy and environmental science and address scientific grand challenges. Our science, our capabilities, our users, our new strategic plan and OurJGI culture all ensure the future success of the JGI.

Nigel J. Mouncey, DPhil

Director, DOE Joint Genome Institute

Nigel J. Normer





The 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 the BER's bioenergy and environmental missions. These missions mirror the following DOE and national priorities:

- Develop renewable and sustainable sources of biofuels from plant biomass 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 the DOE has stewardship responsibilities.

Bioenergy

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

Biogeochemistry

Many DOE-relevant environmental processes are controlled by complex, interconnected biogeochemical reactions. The JGI engages in projects that can enable 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 biogeochemical cycles. 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.

Lee Vining Creek starts as snowmelt in the Ansel Adams Wilderness and then descends more than 1,600 feet, making its way to the alkaline and extremely salty waters of Mono Lake, California. Extensive studies have been done on microbes isolated from Mono Lake sediments, some sequenced at the JGI, to learn more about how they are coping with the presence high dissolved concentrations of arsenic in the environment. (*Jon Bertsch*)

Organizational **Structure**

Senior Management Team



Nigel Mouncey JGI Director



Susannah Tringe Deputy, User Programs



Len Pennacchio Deputy, Genomic Technologies



Kjiersten Fagnan **Chief Informatics** Officer/ Data Science & Informatics Lead



Axel Visel Deputy, Science Programs



Ray Turner Deputy, Operations



David Gilbert Senior Manager, Communications & Outreach



Ronan O'Malley Sequencing Technologies



Atif Shahab Institutional Informatics



Dan Rokhsar Eukaryote Super Program



Igor Grigoriev Fungal Program



Jeremy Schmutz Plant Program



Jan-Fang Cheng **Functional** Genomics

Alex Copeland

Genome Assembly



Steven Wilson Systems Engineering



Nikos Kyrpides Prokaryote Super Program



Tanja Woyke Microbial Program



Emiley Eloe-Fadrosh Metagenome Program



Hugh Salamon Advanced Analysis



Yasuo Yoshikuni DNA Synthesis Science Program



Trent Northen Metabolomics Group

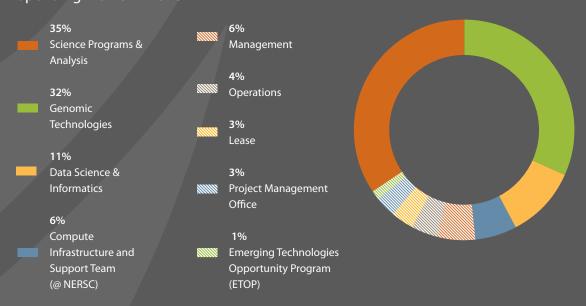


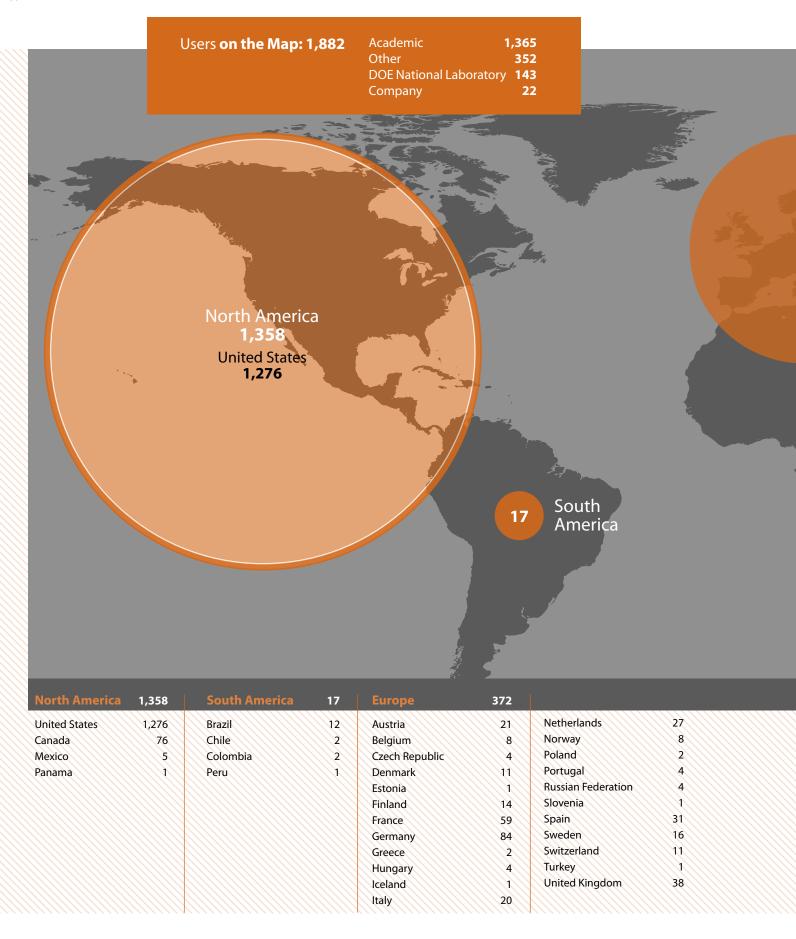
Primary Users Fiscal Year 2018

This category captures the primary users of the JGI, which include PIs and their collaborators on all user projects that were active during FY 2018. 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 **FY2018**





Users of JGI Tools & Data

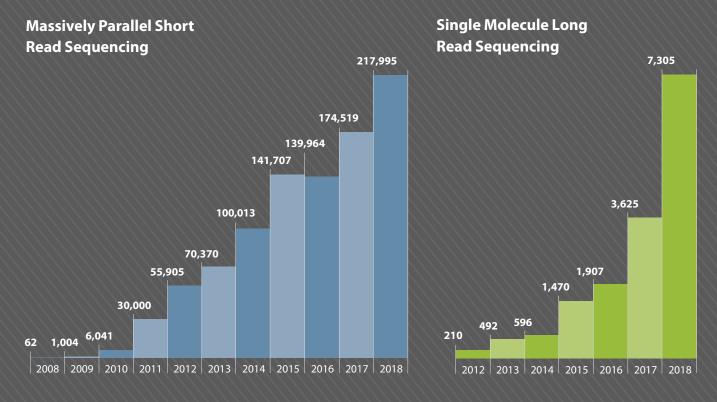
The JGI produces high-quality data that are made available to the community through our data portals. External Data Users are not included in the primary Data User count because their projects were not conducted as part of JGI's user programs. When Data Users log in to the systems and download data, this activity is tracked in order to help us understand which data sets are of greatest value to our Data Users. Additionally, the JGI's data management system is able to restore data from HPSS upon user request and this activity is logged. In 2018, JGI users downloaded more than 2.1 million files and a total of 5.6 PB of data.

Workshops and Meetings		
Workshops and Meetings	Participants	
Genomics of Energy & Environment 13th Annual User Meeting Participants:	602	
VEGA Symposium	134	
Other Workshop Participants	879	

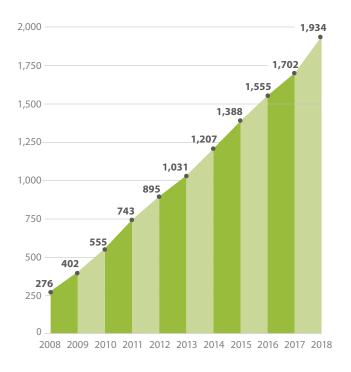
Sequence Output

(in billions of bases or GB)

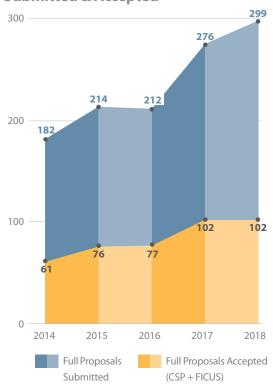
The JGI supports short and long read sequencers, where a read refers to a sequence of DNA bases. Short read sequencers produce billions of 300 base reads used for quantification such as in gene expression analysis. Long read sequencers currently average 12,000 base reads and are used for *de novo* genome assembly. Combined short read and long read totals per year give JGI's annual sequence output. The total sequence output in 2018 was 225,299 GB.



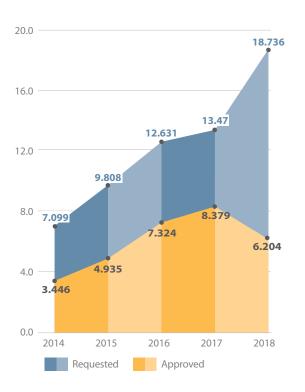
Cumulative Number of Scientific Publications



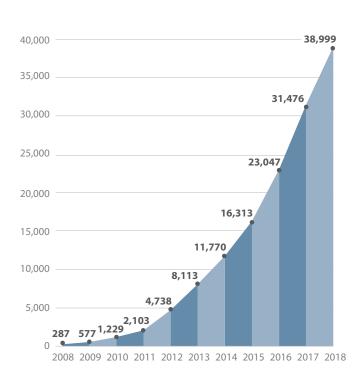
User Program **Growth Submitted & Accepted**



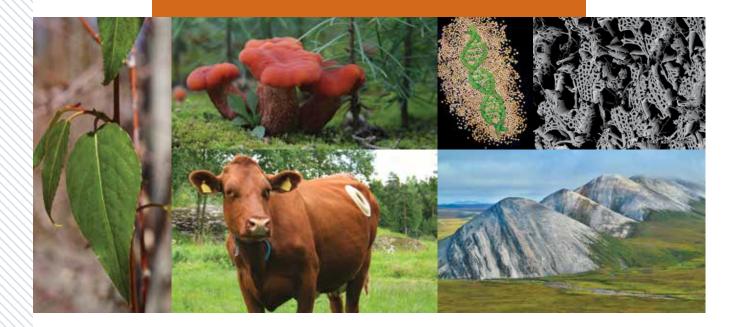
User Demand: Synthetic Biology Requested & Approved (Mb)



Cumulative Number of **Projects Completed**



Case Study: JGI's Community Science Program at Fifteen



The JGI Community Sequencing Program (CSP) debuted in January 2004 after the JGI transitioned to a national user facility following the completion of their contribution to the Human Genome Project. Researchers in disciplines aside from biomedicine could propose projects involving at least 10 million bases of sequence; approval required vetting for scientific merit. For the first round of applications, the DOE allocated roughly 12 billion bases of sequencing capacity, then more than half of the JGI's yearly production total..

The proposals approved for the most recent round of the CSP, now the Community Science Program as the JGI's capabilities have expanded far beyond DNA sequencing, were capped at 2.5 trillion bases or Terabases (Tb) of sequence data per proposal, though collaborative proposals could request as much as 40 Tb. The CSP also accepts smaller scale proposals year-round from investigators and research initiatives new to the JGI, and for DNA synthesis science.

In providing high-quality reference genomes and data analysis tools to CSP users, the JGI has helped research communities flourish. Among the more notable CSP proposals approved and published over the years are:

Pioneer Tree Genome. (*Science*, September 2006) The international consortium selected a tree with a genome of less than a billion bases. Since the paper's publication, the clonally propagated sapling on the JGI's front lawn has grown in parallel with various genomics technologies harnessed by the worldwide research community availing of the poplar genome resources.

The Mechanisms of Plant-Fungi Symbiosis. (*Nature*, March 2008). Plant health is dependent on a number of factors, including their interactions with ectomycorrhizal (ECM) fungi. When the *Laccaria bicolor* establishes a partnership with plant roots, a mycorrhizal root is created. The fungus within the root is protected from competition with other soil microbes and gains preferential access to carbohydrates within the plant. The underground networks that link ECM fungi and their host plant roots impact not just the plant's health and tolerance to stressors such as drought or disease, but the global carbon cycle as well. Building upon this first ECM symbiosis study, researchers have since learned more about the evolution of this lifestyle.

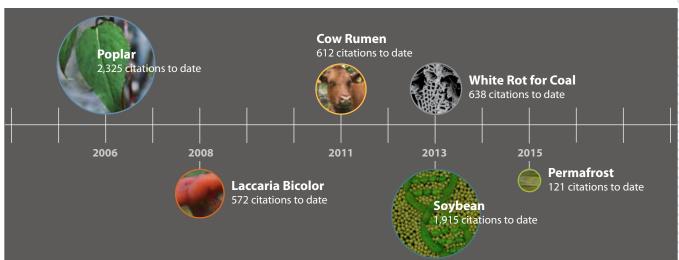
Mining the Cow Rumen Microbiome. (*Science*, January 2011) While cows seem to spend hours chewing grass, the microbes in their guts are actually responsible for breaking down the plant mass to extract the nutrients. Among the challenges associated with developing sustainable, alternative fuels from plants has been the costs associated with the processes involved in converting plant biomass first into sugars and then into biofuels. This work paved the way for the international effort to catalog rumen microbial genomes and isolates, and link microbial genomes to rumen function, known as the Hungate1000 collection (see page 31).

How an Ancestral Fungus May Have Influenced Coal Formation. (Science, June 2012). Researchers suggested the evolution of white rot fungi affected coal formation as dead plant matter could be completely broken down into its basic chemical components and then released into the atmosphere. Unlike brown rot fungi, white rot can efficiently break down lignin, which helps keep plant cell walls rigid, a factor in the JGI's decision to sequence the first white rot genome, *Phanerochaete chrysosporium*, in 2004.

A Genome for a Food and Fuel Crop. (*Nature*, October 2013) With more than a billion bases in its genetic code, soybean was the first legume species with a published complete draft genome sequence. Soybeans are the single largest source of biodiesel in the United States, and the genome sequence helps researchers identify genes that are essential for protein and oil content. Additionally, the soybean genome became a key reference for more than 20,000 legume species while also allowing researchers to explore nitrogen-fixing symbiosis, critical for successful agricultural crop rotation strategies.

Permafrost Microbes in a Changing Climate. (*Nature*, March 2015) If the permafrost soils storing billions of tons of carbon thaw completely, the result may be the largest contribution of carbon transferred to the atmosphere by a single terrestrial process. Over several years, JGI researchers and collaborators have applied multiple omics techniques to look at the short-term and long-term effects of thawing permafrost. Among the JGI's collaborators are researchers at the Next-Generation Ecosystem Experiments (NGEE Arctic) project, a consortium of academic institutions and national laboratories developing a process-driven ecosystem model to better predict the evolution of Arctic ecosystems in a changing climate.

Proposal Citations (2004–2018)









Discovery



Marsarchaeota in Yellowstone Offers Clue to Early Earth Conditions

Iron cycling is thought to have been essential to the development of life on Earth. The extreme environments found in Yellowstone National Park include a wide range of temperatures, levels of acidity, and geographical features, and the archaeal lineages found there and elsewhere offer researchers an analog for studying early Earth conditions.

Through a combination of sequencing tools and techniques applied to samples collected from acidic ironoxide microbial mats in Yellowstone National Park over time, researchers led by Montana State University's Bill Inskeep, a longtime collaborator of the JGI have discovered and characterized a novel phylum-level lineage of archaea with at least two major subgroups, dubbed Marsarchaeota. In a report published in Nature Microbiology, Inskeep and his team describe a candidate phylum-level lineage of aerobic archaea found in iron-oxide microbial mats. The reddish hues caused by the presence of iron led the team to name the archaeal lineage for the planet Mars.

To help determine where Marsarchaeota might fall amidst other known archaeal lineages, the team relied on a combination of phylogenomic analyses, transcriptomics for microbial genomic activity (gene expression), and direct microscopy. Through the JGI's Community Science Program (CSP), the researchers used metagenome assemblies, transcriptomes, and single amplified genomes (SAGs) from samples collected at several locations to thoroughly characterize the archaeal lineage, information that they believe will lend insights into discussions on the origin of archaea. The researchers report that the Marsarchaeota are a sister group to the archaeal lineage Geoarchaeota that Inskeep's team previously identified and characterized, also with the JGI's help. Additionally, the Marsarchaeota comprise 20 to 50 percent of the iron-oxide microbial mat communities in the 60-80°C temperature range.

The discovery of aerobic, thermophilic Marsarchaeota in these microbial mats provides the team with clues on how early life evolved on Earth, as iron is believed to have played a key role in redox processes important in the formation and evolution of early life. Additionally, high-temperature (thermophilic) microbes are of interest for their potential use in a number of biotechnological applications. Molecular biology has benefited from heat-stable enzymes derived from thermophiles, stabilizing the polymerase chain reaction technology that enables researchers to generate thousands to millions of copies of a particular DNA sequence from just a few pieces of DNA.

The iron (Fe)-oxide terraces at Echinus Geyser form from the oxidation of ferrous Fe, and the temperature across the terraces ranges from ~ 60-70°C, while pH values vary from 3.4 to 3.6. A very thin (1–2 mm) layer of water flowing over the Fe-oxide terraces from the outflow channel at Echinus Geyser, located in Yellowstone National Park's Norris Geyser Basin, is thought to provide the needed oxygen to create habitats suitable for the Marsarchaeota. (Bill Inskeep)



A lingering question is how Marsarchaeota can access oxygen in low-oxygen habitats such as these microbial mats. Inskeep suggested that the thin film of water running over the iron-oxide microbial mats at the geyser sites provides just enough oxygen to the Marsarchaeota. "Proving that is an entirely different process," he added.

The Marsarchaeota furnish greater definition of the tree of life's roots on Earth. Their broad distribution suggests that iron-oxidizing habitats similar to those in Yellowstone from which the Marsarchaeota were isolated and identified may have been important to the early evolution of these microbes.

Evolutionary Changes in the Genetic Code of Yeasts

Yeasts are some of the most important microbes used in biotechnology, but only a fraction have been harnessed for biotechnological applications. Researchers studying various nonconventional yeast species aim to capitalize on yeast physiology and genetic features to drive biotechnology. In the future, yeasts may play a large role in developing palm oil substitutes, ethanol products, and feedstocks, for example. (Watch Tom Jeffries, president of Xylome and professor emeritus at the University of Wisconsin–Madison, on the importance of nonconventional yeasts for biotechnological applications at the 13th Annual JGI Genomics of Energy and Environment Meeting at http://bit.ly/JGI2018Jeffries.)

Yeasts are also unique in terms of their genetic code. In nuclear genomes, a genetic code change where the amino acid assignment of a sense codon is swapped for a different amino acid is very rare. Until 2016, the only such example in eukaryotes was the reassignment of the three-letter code CUG from the amino acid leucine (its usual meaning) to the amino acid serine in budding yeasts.

Then in 2016, JGI researchers discovered a similar switch in Pachysolen tannophilus. A close relative of the well-known yeast genus Candida, P. tannophilus is a recently sequenced yeast that can ferment the wood sugar, xylose. As reported then in the Proceedings of the National Academy of Sciences (PNAS), the team discovered another reassignment of the codon CUG. This time, CUG was changed from serine to alanine in P. tannophilus. This change was only the second reported case of a non-stop codon reassignment. Scientists are not sure why or how the change came about, but knowing if yeasts' genetic codes are the same is important for gene expression experiments. (Watch the paper's first author Robert Riley discuss the CUG reassignment at the 11th Annual JGI Genomics of Energy and Environment Meeting: http://bit.ly/JGI2016Riley.)

Building on this information, another group of scientists followed up, investigating the phylogenetic relationships in yeasts with standard and non-standard genetic codes. In a report published during the first half of 2018 in Nature Communications, the researchers looked at the genomes of 52 yeast species, including seven newly sequenced species, using whole-genome data and mass spectrometry to determine phylogeny and genetic codes, respectively. Within this data set, the researchers observed all three CUG codon reassignments: CUG-Ser, CUG-Ala, and CUG-Leu.

The researchers propose that natural selection caused by a toxin from a virus-like element (VLE) acting specifically against the ancestral tRNALeu(CAG) may explain the CUG codon's instability in yeasts. The researchers believe that a VLE with a tRNALeu(CAG)-specific toxin infected the common ancestor of five clades of yeasts. In response, yeast lineages either changed their genetic codes or altered the sets of tRNALeu genes they maintain. If this hypothesis is correct, these genetic code changes represent a profound defense mechanism.



Attendees of JGI's Microbial Genomics & Metagenomics (MGM) Workshop work through hands-on exercises. (Frederik Schulz, JGI)



The team used data from a three-year metagenomic time series collected from Trout Bog Lake, a small acidic bog in Wisconsin. (Image courtesy of McMahon Lab)

Doubling the Virophage Database

In freshwater lakes, microbes regulate the flow of carbon and determine if the bodies of water serve as carbon sinks or carbon sources. Algae and cyanobacteria in particular can trap and use carbon, but their capacity to do so may be affected by viruses. Viruses exist amidst all bacteria, usually in a 10-fold excess, and are made up of various sizes ranging from giant viruses to much smaller viruses known as virophages (which live in giant viruses and use their machinery to replicate and spread.) Virophages can change the way a giant virus interacts with its host eukaryotic cell. For example, if algae are co-infected by a virophage and giant virus, the virophage limits the giant virus's ability to replicate efficiently. This reduces the impact a giant virus has on the diversion of nutrients, allowing the host algae to multiply, which could lead to more frequent algal blooms.

Using metagenome data sets collected over several years in northern freshwater lakes, a team led by researchers at The Ohio State University and the JGI uncovered 25 novel sequences of virophages. Reported October 11, 2017, in *Nature Communications*, the identification of these novel sequences effectively doubles the number of virophages known since their discovery a decade ago.

"Usually metagenome data sets are one-offs," said JGI scientist and first author Simon Roux. "People had started to see virophages in metagenomes, but no one had a long time series until now. Was it here once? Always? We never really knew this, but it's a critical piece of information to understand their importance."

The work stemmed from a JGI CSP proposal involving northern freshwater lakes by Trina McMahon of the University of Wisconsin–Madison. Samples of microbial communities in Lake Mendota and Trout Bog Lake were regularly collected over several years as part of the National Science Foundation–funded North Temperate Lakes Long Term Ecological Research (LTER) project. Sequencing and analyzing these metagenomes from the three-year and the five-year time series is allowing researchers to identify the community members, determine their metabolic pathways, and follow changes in communities over several years.

Beyond looking at the microbial communities, McMahon and Rex Malmstrom, head of the JGI Micro-Scale Applications group, asked collaborator Matt Sullivan at The Ohio State University if he'd be interested in using the same metagenomic data sets to look at the lakes' viral ecology. Roux started mining the data sets while still a postdoctoral fellow with the Sullivan Lab. "I knew there were lots of viruses in the sequence data, but not that some of the viruses were themselves hosts to other viruses," said Malmstrom. "With time series data, we could do more than assemble genomes and build phylogenetic trees: The data allowed us to examine genetic variation within populations and look for co-occurrence and abundance patterns between virophages and their giant virus hosts. With so many time points in the data set, you can find strong connections."

McMahon, whose CSP data sets were the basis of this work, said that having the viral ecology information helps form a more complete picture of the ecosystem. "We are thrilled to have one more piece of the puzzle. Viruses are clearly playing a major role in shaping community composition, and therefore function, of the whole lake ecosystem," she said.

"My own lab lacks the expertise to tackle viruses alone. Hence, the collaboration with Simon and Matt Sullivan is so important. Our long-term goal is to learn enough about the forces controlling community assembly and dynamics, as well as the ecological traits of each lineage, to create more predictive models about how freshwater lakes will respond to climate and land-use change, at an ecosystem scale," she added.

Aside from doubling the number of virophages in public databases, the time series allowed Roux and his colleagues to see the viruses' ecological profiles — if the presence of the viruses had been influenced by factors such as the seasons or abundance of particular microbes. Through co-occurrence analysis, the researchers associated the virophages with sequences of known lineages of giant viruses, and proposed the existence of three new groups of candidate giant viruses infected by virophages. These co-occurrence analyses also allowed the researchers to find putative associations between the giant virus sequences and specific eukaryotic hosts.

"These findings are correlation-based," noted Roux, "but the study is a good example of a metagenomics use case. Metagenomes helped us not only discover new viral diversity and determine what it should do in the ecosystem, but it also helps us design a hypothesis and follow-up experiments about virus-host interactions so we're not just throwing out a wide net blindly."

Workflow Advances Experiment-Based Gene Function Assignments

In the air, beneath the ocean's surface, and on land, microbes are the minute but mighty forces regulating much of the planet's biogeochemical cycles. While advances in sequencing technologies have enabled researchers to access the genomes of thousands of microbes and make them publicly available, no similar shift has occurred with the task of assigning functions to the genes uncovered.

To help overcome this bottleneck, JGI researchers were among the scientists at Lawrence Berkeley National Laboratory (Berkeley Lab) who developed a workflow that enables large-scale, genome-wide assays of gene importance across many conditions. Tested on nearly three dozen bacteria from various genera, the workflow described May 16, 2018, in Nature combined high-throughput genetics and comparative genomics to identify mutant phenotypes for thousands of genes with previously unknown functions.

"This is by far the largest functional genomics study of bacteria ever published," said study senior author and Berkeley Lab biologist Adam Deutschbauer of the Biosciences Area's Environmental Genomics and Systems Biology Division. "This is the first really large, systematic experimental effort to try to assign functions to bacterial genes of unknown function. We are tackling the problem that biology is up against and recognizes: It is super easy to sequence, but we cannot currently assign confident functions for the majority of genes identified by sequencing. Our experimental data provide an anchor that other researchers could use to make a more informed inference about protein function."

The team worked with 32 bacteria, including plant growth-promoting bacteria and a cyanobacterium relevant for biofuels production, as well as bacteria involved in bioremediation. "Typically, researchers work on functional analysis of individual genomes, from a limited number of 'workhorse' bacteria" noted JGI scientist and study co-author Matt Blow. "This is because of the limited capacity of functional analysis approaches compared with high-throughput sequencing. Here, you have data from 32 different bacteria at once, capturing more microbial diversity."

To more efficiently generate mutant libraries for each bacterium, the team refined a DNA barcode sequencing approach known as RB-TnSeq (randomly barcoded transposon sequencing) that Deutschbauer had begun developing while still part of Adam Arkin's lab.

"The technology behind this project was developed to elucidate the genetic functions of all the organisms we were collecting in the field," said Arkin, Co-Director of the Scientific Focus Area ENIGMA program working on developing laboratory and computational tools that link molecular functions within individual members to integrated activities of microbial communities. "We believe that to understand means given appropriate data, you should be able to predict, control, and design behavior in the system of interest. The implications of this work are that it could be scaled with proper investment and coordination and could have substantial benefit for understanding the genetic potential of the earth."

Deutschbauer pointed out that the resulting large data set allowed the team to glean insights from conserved phenotypes across organisms. The data set also allowed the team to look for co-fitness patterns among the genes — cases where two genes had similar patterns of phenotypes across all conditions — a correlation that suggested they might be part of the same pathway. For example, the team found that genes with the uncharacterized protein domain UPF0126 were important for growth on glycine in 11 different bacteria, suggesting that this protein domain is involved in transporting glycine across the cell membrane. Studying such conserved associations, he added, demonstrates the value in identifying phenotypes for homologous genes across multiple bacterial species.

"A comparative functional genomics study of bacteria was not really possible before because large genetic data sets were available for only a few bacteria, and the ones that did exist were not typically generated with the same technology, same methodology, and same metadata, so it's hard to do comparisons," he said. "Although we experimentally studied a relatively small number of bacteria compared to the diversity present in nature, our data are of relevance across all bacteria."

The data set is publicly accessible for comparative analyses at fit.genomics.lbl.gov. Arkin also sees future benefits toward integrating this data set into systems like the JGI's Integrated Microbial Genomes and Microbiomes (IMG/M) system and the DOE Systems Biology Knowledgebase (KBase), the first large-scale bioinformatics system that allows users to upload, analyze, and share information within a single integrated environment. Arkin is KBase's Chief Executive Officer and lead primary investigator. "These data sets provide a fantastic opportunity for innovations in data science to predict biological function," he said.



JGI co-author Matt Blow, shown here with a student intern. (Marilyn Chung, Berkeley Lab)

Benchmarking Computational Methods for Metagenomes

For more than a decade, the JGI has been enabling researchers to study uncultured microbes unable to grow in the lab, using state-of-the-art approaches such as metagenomics and the development of computational tools to uncover and characterize microbial communities from the environment. To tackle assembling metagenomes into a set of overlapping DNA segments that together represent a consensus region of DNA (a contig), then binning these contigs into genome bins, and finally conducting taxonomic profiling of genome bins, analysts around the world have developed an array of different computational tools. However, little consensus existed on how to evaluate the performance of these tools.

In *Nature Methods* on October 2, 2017, a team including JGI researchers described the results of the Critical Assessment of Metagenome Interpretation (CAMI) Challenge, the first ever community-organized benchmarking assessment of computational tools for metagenomes. The CAMI Challenge was led by Alexander Sczyrba, head of the Computational Metagenomics group at Bielefeld University and a former JGI postdoctoral fellow, and Alice McHardy, head of the Computational Biology of Infection Research Lab at the Helmholtz Centre for Infection Research.

"It is very difficult for researchers to find out which program to use for a particular data set and analysis based on the results from method papers," said McHardy. "The data sets and evaluation measures used in evaluations vary widely. Another issue is that developers usually spend a lot of time benchmarking the state-of-the-art when assessing the performance of novel software that way. CAMI wants to change these things and involves the community in defining standards and best practices for evaluation and to apply these principles in benchmarking challenges."

More than 40 teams signed up for the CAMI Challenge, and the organizers received 215 submissions from 25 programs around the world, though only 17 teams were willing to have their software implementations published. The organizers evaluated computational tools in three categories. Half a dozen assemblers and assembly pipelines were evaluated on assembling genome sequences generated from short-read sequencing technologies. In the binning challenge, five genome binners and four taxonomic binners were evaluated on criteria including the tools' efficacy in recovering individual genomes. Finally, 10 taxonomic profilers with various parameter settings were evaluated on how well they could predict the identities and relative abundances of the microbes and circular elements.

"The JGI has a strong interest in benchmarking of tools and technologies that would advance the analysis of metagenomes and improve the quality of data we provide to the users. Having published the very first study on the use of simulated data sets for benchmarking of metagenomics tools from the JGI, it is great to see how this methodology has expanded over the years and now through this study is evolving into a model for standardized community efforts in the field," said Nikos Kyrpides, JGI Prokaryote Super Program head.

Cataloging Candidate Genes for Plant Microbiome Studies

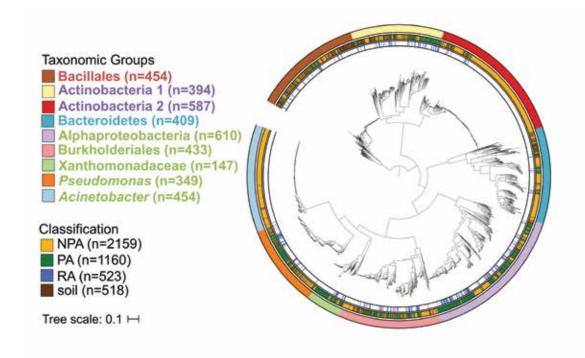
As the global population rises, estimated to hit nearly 10 billion by 2050, so does the need to boost crop yields and produce enough plant material for both food and sustainable alternative fuels. To help improve crop breeding strategies and overcome challenges such as making plants more tolerant of marginal lands, and stresses such as drought and low nutrient availability, researchers are focusing on understanding and promoting beneficial plant-microbe relationships.

Published December 18, 2017, in *Nature Genetics*, a team led by researchers at JGI and the Howard Hughes Medical Institute at the University of North Carolina at Chapel Hill (UNC) have exploited a catalog of bacterial genomes to identify and characterize candidate genes that aid bacteria in adapting to plant environments, specifically genes involved in bacterial root colonization.

Most of the studies in the field to date have focused on the community structure of the plant microbiome, i.e. "who is there," and less on the function, i.e. "what they are doing, how and when they are doing it." Previous studies that have considered function have mainly looked at a single host-microbe interaction, such as the one between an *Arabidopsis* plant and a pathogen.

"If we want to engineer the right microbiome to support plant growth, we need to understand the real function of the microbiome and not just sequence marker genes," said study co-first author and JGI research scientist Asaf Levy. "Here we used a massive genomic and computational effort to address the fundamental and important question: 'How does the plant microbiome interact with the plant?""

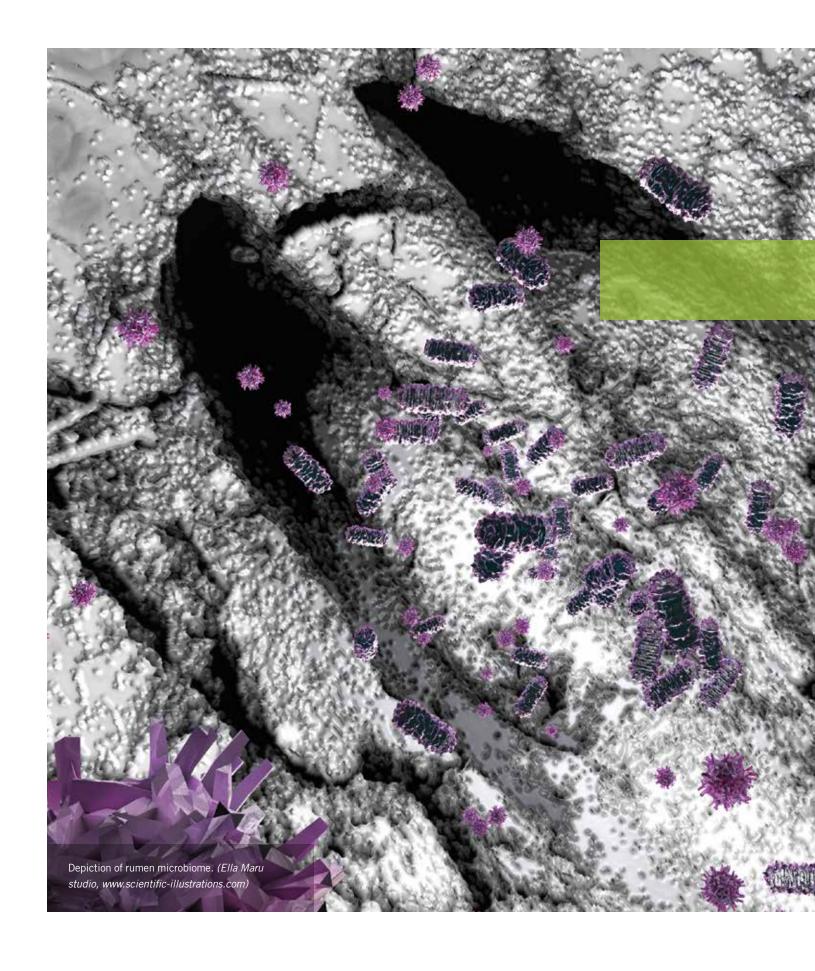
Most of the interaction between microbes and plants occurs at the interface between the roots and soil. Researchers from UNC, Oak Ridge National Lab, and the Max Planck Institute isolated novel bacteria from the root environment of *Brassicaceae* (191), poplar trees (135), and maize (51). The genomes of these 377 bacterial isolates, plus an additional 107 single bacterial cells from roots of *A. thaliana*, were then sequenced, assembled, and annotated at the JGI.



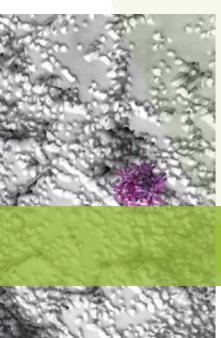
Phylogenetic tree of over 3,800 high quality and non-redundant bacterial genomes. Outer ring denotes the taxonomic group, central ring denotes the isolation source, and inner ring denotes the root-associated genomes within plant-associated genomes. Taxon names are color-coded based on phylum: green - Proteobacteria, red - Firmicutes, blue -Bacteroidetes, purple - Actinobacteria. (Asaf Levy, JGI and Isai Salas Gonzalez, University of North Carolina)

The authors then combined the new genomes with thousands of publicly available genomes that represent the major groups of plant-associated bacteria, and included bacteria from multiple plant and non-plant environments, such as the human gut, for comparison. The resulting database of 3837 genomes, 1160 of which are from plants, was used in a comparative genomics analysis.

The researchers then identified genes that are enriched in the genomes of plant-associated and root-associated organisms. "It's very important for us to understand what genes and functions microbes use to colonize plants because only then might we have a chance to rationally devise useful 'plant probiotics' to help us raise more food and energy crops with fewer chemical inputs such as fertilizers and pesticides or fungicides," said study senior author Jeff Dangl, a Howard Hughes Medical Institute investigator and the UNC John N. Couch Professor of Biology.







Bioenergy



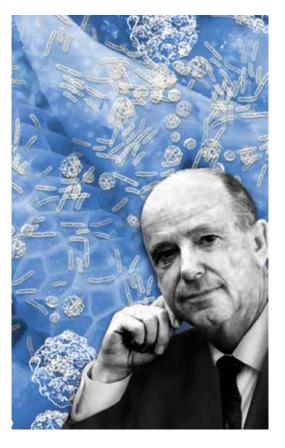
The digestive tracts of ruminant animals, such as cattle, sheep, and goats, convert lignocellulosic plant matter to short-chain fatty acids used for nourishment with unparalleled efficiency, thanks to the activity of symbiotic microbes in the rumen. Rumen microbes play a vital role in allowing ruminant livestock to break down the food they eat and produce milk, meat, and wool, which help support the livelihoods and food security of over a billion people worldwide. The process, however, is also the single largest human-influenced source of the greenhouse gas methane (CH₄), with these animals releasing approximately 138 million U.S. short tons of CH₄ into the atmosphere each year.

Understanding the diversity and function of the rumen microbiome is a critical step toward developing technologies and practices that support efficient global food production from ruminants while mitigating methane emissions. Additionally, there is considerable interest in identifying biotechnologically relevant enzymes for the conversion of plant feedstocks to biofuel and bioproducts.

Reported March 19, 2018, in Nature Biotechnology, an international team led by William (Bill) Kelly, formerly at AgResearch New Zealand's Grasslands Research Centre, and including JGI scientists presented a reference catalog of rumen microbial genomes and isolates cultivated and sequenced from the Hungate 1000 collection. One of the largest targeted cultivation and sequencing projects to date, the collection was produced through the coordinated efforts of rumen microbiology researchers worldwide. At the beginning of the project, there were reference genomes for only 14 bacteria and one methanogen. The Hungate catalog now contains a total of 501 genomes: 410 newly generated from this study, plus an additional 91 already publicly available from other studies.

"The JGI is a world leader in conducting, enabling, and democratizing sequencebased research — and one of the few places that does science at this large scale. Beyond the sequence generation, data processing, and big compute resources, we bring significant experience and expertise to help bridge the gap from sequence to biology," said Rekha Seshadri, JGI computational biologist and co-first author of the paper.





The project was named in honor of the late Robert Hungate (1906–2004), a pioneering microbiologist who invented the widely used method of cultivating strictly anaerobic bacteria that now bears his name.

The work, noted Kelly, is the culmination of a JGI CSP 2012 proposal that originated at a meeting of rumen microbiologists held in New Zealand in February 2011. Former JGI Director Eddy Rubin was one of the attendees at the meeting, as the JGI had just published the first rumen metagenome study in *Science*. His perspective on the scale of sequencing that was possible encouraged the development of an ambitious project.

"This project is a great example of how science can progress rapidly if we do things together," added study co-first author Sinead Leahy, now with the New Zealand Agricultural Greenhouse Gas Research Centre. "The project helps further fill the gap in our knowledge of anaerobic microorganisms as it specifically targets the specialized anaerobic rumen environment and reports genome sequences from strains that have yet to be taxonomically assigned or phenotypically characterized, but which are among the most abundant rumen microorganisms."

The Hungate catalog encompasses 75 percent of genus-level taxa reported from the rumen. The researchers were able to assign individual microbes to the major metabolic pathways involved in rumen function. They reported that in total, the catalog of genomes encodes nearly 33,000 degradative carbohydrate-active enzymes (CAzymes) which can break down plant cell walls. Other metabolic highlights and evolutionary vignettes of the rumen microbiome are presented in the manuscript. The researchers noted an interesting instance of evolution by gene loss of the universally conserved enolase, the penultimate enzyme in glycolysis, the metabolic pathway that converts glucose to pyruvate. Rumen-specific adaptations, such as *de novo* synthesis of vitamin B_{12} and potential vertical inheritance of the rumen microbiome, are discussed inn the manuscript.

To test the value of the Hungate Collection as a resource that underpins metagenomic analysis, 1.4 million coding sequences from the reference genomes were searched against ~1.9 billion coding sequences from over 8,000 varied metagenomic samples, stored in the IMG/M database. The IMG/M system supports annotation, analysis, and distribution of microbial genomes and microbiomes.

The majority of Hungate genomes were indeed present in available rumen metagenomes. "However, there was significant overlap with the human microbiome: Almost a third of the species were detected in human digestive system samples, inadvertently increasing the reference set for the study of the human microbiome as well. IMG is a comprehensive resource of sequence data integrated with environmental metadata without which these observations would not have been made," said Seshadri. The importance of integrating microbiome data across all habitat types to enable novel correlations and discovery is one of the main pillars of the proposed National Microbiome Data Collaborative, in which the JGI plays a leadership role.

The Hungate Collection was conceived as a community resource. Access to bacterial cultures can be requested from AgResearch New Zealand. All available genomic data and annotations are available through the IMG/M portal. Additionally, all 410 genomes sequenced in the study can be downloaded through a dedicated portal.

Large-Scale Maize Study Identifies Rhizosphere Core Microbial Community

A plant's health is affected not only by conditions such as water and temperature, but by the microorganisms that live around its roots. The rhizosphere microbiome, as this microbial community is known, regulates nutrient availability to the plant from the soil, and can affect plant growth and yields.

In the June 25, 2018, issue of PNAS, researchers reported on the results of a large-scale field study that partially replicates earlier trials to identify soil microbes that colonize plants and can be associated with particular traits. The work was conducted by an international team led by scientists at the Max Planck Institute (MPI) for Developmental Biology, the Howard Hughes Medical Institute at the University of North Carolina at Chapel Hill, and Cornell University, and includes JGI researchers.

"This is an extremely large and thorough temporal survey of the maize rhizosphere microbiota," noted study senior author Ruth Ley of the MPI. "The data set constitutes a rich resource for soil microbiologists and potentially plant breeders, once we zero in on the microbial traits that we'd like to breed for, to reduce dependence on fossil fuels in agriculture."

The study builds on a previous study, also reported in PNAS, in which the team used 500 samples from 27 maize lines growing in five fields across three states. All of those samples were collected at a single time point. This time around, the team collected nearly 5,000 samples from a subset of the same maize lines growing in just one field over an entire growing season. "Scaling from 500 to 5,000 is challenging for sample processing, and the bioinformatics is also quite challenging, as we had over half a billion 16S sequences," Ley said. "First author Tony Walters did a brilliant job with this."

The information allowed the team to associate abundances of microbial populations with plant genotype, while also distinguishing the effects of conditions such as plant age and weather. The large-scale field study allowed the team to identify 143 heritable microbes, whose population variations across samples were partially driven by



differences in plant genotype. Additionally, the team identified a core rhizosphere microbiome consisting of seven operational taxonomic units, all within the Proteobacteria phylum, found in every single sample.

"Scaling up allowed us to understand the relative importance of plant genetics, environment, and time," said Cornell University's Ed Buckler, whose team provided expertise on maize genetics and the field trials. "We knew each mattered, but this really provides perspective on how much each matters."

The work was conducted as part of the JGI's Rhizosphere Grand Challenge pilot projects involving maize and the model plant *Arabidopsis*. These projects highlighted the JGI's capabilities and expertise that could be harnessed for scientific grand challenges that were assigned to the DOE's BER back in 2010.

"The overarching goal of the Rhizosphere Grand Challenge as it was conceived was to determine the major drivers of microbial community composition in plant-associated microbial communities, for example, plant compartment, soil type, plant age, and plant genotype," said JGI User Programs Deputy Susannah Tringe. "The previous paper used 454 pyrotag sequencing to demonstrate effects of plant genotype on microbial community structure in the rhizosphere, but lacked the statistical power to make specific genotype-phenotype links. The current study used Illumina tag sequencing, all done at the JGI, for deeper sampling and much higher temporal resolution for improved statistical analysis."

Liverwort Genes and Land Plant Evolution



A Marchantia polymorpha thallus in the vegetative form. Cup-shaped structures on the surface are gemma cups (cupules), reproductive organs producing asexual propagules (gemmae). (Shohei Yamaoka, Kyoto University)

Though the common liverwort is found around the world, it's easy to overlook — the plant can fit in the palm of one's hand and appears to be comprised of flat, overlapping leaves. Despite their unprepossessing appearance, these plants without roots or vascular tissues for nutrient transport are living links to the transition from the algae that found its way out of the ocean to the established multitude of land plants.

As reported in the October 5, 2017, issue of *Cell*, an international team including JGI researchers analyzed the genome sequence of the common liverwort (*Marchantia polymorpha*) to identify genes and gene families crucial to plant evolution and have been conserved over millions of years and across plant lineages. The work was led by researchers at Monash University in Australia and at Kyoto University and Kindai University in Japan.

"Early plants like the liverwort are what set the world up for land plants. Without them, we wouldn't have plants more than two feet from the ocean and fresh water," said JGI Plant Program head Jeremy Schmutz. "In going back to liverworts, we find genes shared with grasses that are candidate genes for biofuel generation crops. Land plants began with the same parts present in *Marchantia* today, so the changes are all due to factors such as evolution, polyploidy, gene exchange, and rounds of selection. We want to know what genes do, and we do this by translating function across genomes using conserved sequences. Smaller genomes with less complexity — such as those in a basal or early plant model like liverwort — give us the ability to identify ancestral genes for a gene or gene family. We identify gene function in a plant and determine how this gene works, and then we identify other genes by understanding the evolutionary history of the gene or gene family across the history of plants."

The liverwort's genome sequencing and annotation was done through the JGI's CSP and allows for genomic comparisons with other early plant lineages sequenced and analyzed by the JGI: the spikemoss *Selaginella moellendorffi* and the moss *Physcomitrella patens*. One of the most important biochemical pathways concerns production of the hormone auxin, which is critical for regulating plant growth and development. The team identified a minimal but complete pathway for auxin biosynthesis in the liverwort. Another finding suggests that the genes encoding enzymes producing "sunscreen" that allowed early plants to tolerate ultraviolet light may have been transferred from ancient soil microbes.

One of the team's most important findings concerns plant cell wall development. The variety of genes encoding enzymes for plant cell wall development found in *Marchantia* emphasizes the importance of plant cell walls for the transition to land plants. The team identified early lignin biosynthesis genes similar to those in *Physcomitrella*. While the researchers identified genes involved in plasmodesmata formation (plasmodesmata are membrane channels involved in nutrient and signal molecule transfers,) a pathway that is involved in cell division, they also found that liverworts retain the vestiges of cell division pathways predating land plant-specific pathways.

Another important finding involves water retention and distribution. Early plants had to develop strategies for dealing with drought and desiccation, and many of these same strategies are still employed by modern plants. Abscisic acid is a plant stress hormone that regulates when a plant goes dormant because water is in short supply. The researchers found homologous genes for abscisic acid biosynthesis and were also able to identify when specific receptors became critical to land plant families.

Schmutz pointed out that through the CSP, the JGI's exploration of plant evolutionary history is expanding, leading to the development of a comparative genomics framework, including genomes from early plant lineages like the liverwort, that benefits the plant research community at large. By learning the original functions of genes, elucidated from the genomes of earlier, simpler plants and cells, scientists can more easily solve for the functions of related genes seen in more complex plants that may help address DOE missions in bioenergy and environmental processes.

The *Marchantia polymorpha* v3.1 genome data are publicly available at Phytozome, the JGI's plant comparative genomics portal, which provides users and the broader plant science community with a hub for accessing, visualizing, and analyzing JGI-sequenced plant genomes, as well as selected genomes and data sets that have been sequenced elsewhere.

Succulent Genes for Water Use Efficiency



Kalanchoë fedtschenkoi (lavender scallops). (Forest and Kim Starr, Flickr CC BY 2.0)

In the presence of sufficient water and light, most plants conduct photosynthesis through what is known as the C3 pathway. As plants spread out and adapted to live in a variety of environments, they developed alternate photosynthesis pathways, known as C4 and crassulacean acid metabolism (CAM), to make use of limited nutrients. To understand how many plant lineages have independently transitioned from C3 to CAM photosynthesis, researchers sequenced and analyzed the genome of *Kalanchoë fedtschenkoi* (lavender scallops). Comparing this plant's genome to those of other plants that also conduct CAM photosynthesis allowed the team to identify the genes and protein sequences involved in *K. fedtschenkoi's* evolution to a flowering plant that efficiently uses limited water resources to conduct photosynthesis.

Some plants can conduct C4 photosynthesis in water-limited conditions; a different enzyme collects carbon dioxide from the air to form a four-carbon chain that lends itself to the pathway name. In water-poor conditions, some plants collect and process carbon at night rather than during the day through CAM photosynthesis, named for its discovery in succulents. In the December 1, 2017, issue of *Nature Communications*, a team led by Oak Ridge National Laboratory researchers and including JGI scientists sequenced and analyzed *K. fedtschenkoi's* genome to better understand how this plant transitioned from C3 to CAM photosynthesis. CAM photosynthesis is found across 36 plant families and is thought to have evolved independently from multiple C3 lineages. As the first CAM eudicot to have its genome sequenced, *K. fedtschenkoi* offers researchers an emerging model to trace the evolution of CAM photosynthesis through these plant lineages.

As the first CAM eudicot to have its genome sequenced, *Kalanchoë* offers researchers a reference to trace the evolution of CAM photosynthesis in this group of flowering plants. Coffee, bean and sunflower also number among the eudicots. Using recently sequenced genomes of related CAM plants—the pineapple and the moth orchid—the team identified sequences involved in the evolution of plant lineages. The researchers found changes in protein sequences and in gene expression changes related to rescheduling a plant's daily cycles that affect genes involved in traits such as when the leaf pores (stomata) open and heat stress response. While they found convergent changes to either protein sequences or gene expression across genomes in comparing the three CAM genomes, they did not find the same changes in all three genomes simultaneously.

Armed with the *K. fedtschenkoi* genome data, researchers can learn more about how CAM genes are regulated and functionally characterize CAM-related genes, such as by generating loss-of-function mutants. The information can help accelerate genetic improvement of plants to make them more tolerant of stresses such as drought and boost crop yields in non-CAM plants for fuel and food production on marginal lands.

When One Reference Genome Is Not Enough

Much of the research in the field of plant functional genomics to date has relied on approaches based on single reference genomes. But by itself, a single reference genome does not capture the full genetic variability of a species. A pan-genome, the non-redundant union of all the sets of genes found in individuals of a species, is a valuable resource for unlocking natural diversity. However, the computational resources required to produce a large number of high-quality genome assemblies has been a limiting factor in creating plant pan-genomes.

Having plant pan-genomes for important food and fuel crops would enable breeders to harness natural diversity to improve traits such as yield, disease resistance, and tolerance of marginal growing conditions. In a paper published December 19, 2017, in *Nature Communications*, an international team led by JGI researchers gauged the size of a plant pan-genome using *Brachypodium distachyon*, a wild grass widely used as a model for grain and biomass crops. As one of the JGI's Plant Flagship Genomes, *B. distachyon* ranks among the most complete plant reference genomes.

"A vast number of genes are not captured in a single reference genome," added study senior author John Vogel, head of the JGI's Plant Functional Genomics group. "Indeed, about half of the genes in the pan-genome are found in a variable number of lines." Working toward the primary goal of accurately estimating the size of a plant pan-genome, Vogel and his colleagues performed whole-genome de novo assembly and annotation of 54 geographically diverse lines of *B. distachyon*, yielding a pan-genome containing nearly twice the number of genes found in any individual line.

"The genome of a species is a collection of genomes, each with their own unique twist," added JGI bioinformaticist and study first author Sean Gordon. "Now that we know focusing on a single reference genome leads to incomplete and biased estimates of genetic diversity and ignores genes potentially important for breeding applications, we should better incorporate multiple references in future studies of natural diversity."

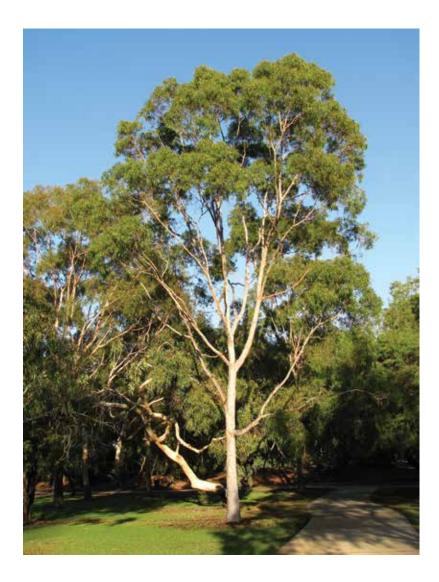


Moreover, genes found in only some lines tend to contribute to biological processes (e.g., development or disease resistance) that may be beneficial under some environmental conditions, whereas genes found in every line usually underpin essential cellular processes (e.g., glycolysis and iron transport).

"This means that the variable genes are being preferentially retained if they are beneficial under some conditions. These are exactly the types of genes that breeders need to improve crops," Vogel said.

In addition, genes found in only a subset of lines displayed faster rates of evolution, lay closer to transposable elements (thought to play a key role in pan-genome evolution), and were less likely to be found in the same chromosomal location as functionally equivalent genes in other grasses. The sequence assemblies, gene annotations, and related information can be downloaded from BrachyPan: brachypan.jgi.doe.gov. The B. distachyon genome is available on the JGI plant portal Phytozome: phytozome.jgi.doe.gov.

Brachypodium plants. (Marilyn Chung, Berkeley Lab)



Corymbia citriodora subspecies citriodora is a native of north Queensland in Australia. (Mervyn Shepherd)

Corymbia Genome Expands Terpene Synthesis Gene Family

A diverse group of plant-produced organic compounds, terpenes are ubiquitous and play key roles in plant growth, defense, and environmental interactions. Terpenes are also economically important because of their use in industrial materials and pharmaceutical products and as biofuel precursors. Collectively, hundreds of terpene compounds have been characterized from eucalypts, a group of 900 tree species belonging to the Myrtaceae (myrtle) family and containing the closely related genera *Angophora, Corymbia*, and *Eucalyptus*.

The genus *Corymbia* is endemic to northern Australia but is increasingly farmed in other countries for essential oil production. The recent assembly of two *Corymbia citriodora* subspecies variegata genomes allowed researchers to study the conservation and evolution of the genes responsible for terpene synthase (TPS) enzyme production. This family of enzymes is critical to the synthesis and broad diversity of terpenes. Until recently, studies of the TPS gene family were confined to two *Eucalyptus* species: *E. grandis* and *E. globulus*. The annotation of two *C. citriodora* subspecies provides an excellent opportunity to investigate the conservation and evolution of this important gene family across eucalypt lineages. Since terpenes serve as feedstocks for biofuel production, a greater understanding of terpene synthesis in plants will be important for alternative fuel development in the future.

Though the closely related Eucalyptus and Corymbia species number among the eucalypts, they inhabit different environments. Eucalyptus species prefer cooler, more temperate or sub-tropical environments, while Corymbia are more abundant in the drier parts of Australia with lower quality soils and even in the desert areas with poor rainfall. That said, *Corymbia* can also thrive in areas that receive a lot of rain. Previous analysis of the *E. grandis* reference genome, an international effort by a team that included researchers at the JGI, revealed the largest number of TPS genes of any currently sequenced plant, a number closely followed by E. globulus. Occurring in clusters or duplicate arrays, these genes are prone to rapid genetic expansion, which partially explains the great variety of terpene products in nature. As part of a proposal by the DOE's Joint BioEnergy Institute (JBEI), the JGI worked on resequencing several eucalypt genomes to establish the feasibility of genome-wide association studies for genetic traits that are desirable from a biofuels production perspective. By using genomic database alignment tools, researchers searched for TPS genes in *C. citriodora*. They then compared the list of putative genes from *C. citriodora* to known TPS gene sequences from *Eucalyptu*s species and other plants. The locations of TPS genes and gene clusters were mapped against those of E. grandis to find differences in genome organization between the two species. The work was reported in the journal Heredity.

To understand the gene expression and function of these TPS genes, Australian researchers sequenced mRNA from different tissues of C. citriodora. From these samples, a total of 127 TPS loci were found, many of which had high sequence similarity to TPS genes from other plants. The researchers reported 102 total putative functional TPS genes in C. citriodora, which is high compared with other plants, and comparable to the number of TPS genes in E. grandis. Notably, the specific types of TPS genes found in C. citriodora suggest that these plants synthesize a high level of secondary metabolites, which play a part in biotic and abiotic stress responses.

Overall, this study improves our understanding of the TPS gene family and shows us that a large TPS gene family is well-conserved across eucalypts. Improved knowledge of the evolution and selection of this gene family may help researchers manipulate TPS genes to increase terpene production for biofuel development.

First Monoploid Reference Sequence of Sugarcane

Most species are diploids and have two sets of chromosomes, one from each parent. In contrast, many crops have multiple sets of chromosomes and are polyploids. Their complex genomes are more difficult to sequence and assemble, limiting the use of modern, genomic breeding in these crops. For diploid species, sequencing programs generally focus on a genotype with two identical sets of chromosomes to produce a monoploid reference sequence. For the highly polyploid sugarcane, an international team of researchers has successfully assembled a first monoploid reference sequence using a targeted approach focusing on the gene-rich part of the genome by harnessing information from a sequenced related species: sorghum.

Eighty percent of the world's sugar comes from sugarcane. That makes the plant of interest to bioenergy researchers who want to develop sustainable alternative fuels. While more than half of the ethanol produced worldwide is from the United States, a quarter is produced by Brazil from sugarcane. Improving sugarcane breeding methods using molecular biology techniques has been hampered by the crop's highly polyploid genome, which makes sequencing and assembly of the genome extremely challenging. The reference sequence is useful for mapping the genes involved in sugar production and for identifying different variants on different chromosomes, information that can be used to assemble a more complex and more realistic polyploid sugarcane genome now underway through the JGI CSP. A genome sequence for sugarcane is a critical factor in developing sugarcane as a sustainable bioenergy feedstock.

Sugarcane is a cultivated crop that very efficiently converts solar energy into plant biomass, which can then be crushed to extract the sugar-laden juice. This juice can be further purified to produce biofuels. A C4 plant like sorghum and maize, sugarcane can grow on nutrient-poor soils and does not need to be planted annually. The modern sugarcane cultivars are hybrids, each with more than 100 chromosomes; assembling a sugarcane

genome is incredibly complex given its estimated size of 10 billion bases (more than three times the size of the human genome). As part of a proposal by the JBEI, the JGI was part of an international team led by researchers from the French Agricultural Research Centre for International Development (CIRAD), who worked on sequencing and assembling fragments of sugarcane chromosomes into the first monoploid reference of the gene-rich part of the sugarcane genome. Reported in *Nature Communications*, their approach relied in part on having a sequence for sugarcane's relative sorghum, a JGI Plant Flagship Genome sequence, and knowing that a high level of colinearity existed between the two crops, which meant most genes in sorghum occur in roughly the same order in sugarcane.

To create this reference sequence, the team relied on bacterial artificial chromosomes (BACs), vectors that can hold large DNA segments such as an entire sugarcane gene and its associated regulatory elements. Sequenced segments of sugarcane-genome-containing BACs were aligned on a sorghum sequence and the overlapping segments used to define a minimum tiling path of sequences in a single contiguous region. The JGI sequenced half of the 4,500 BACs sequenced for this project that were colinear to the gene-rich part of sorghum used as a reference. The final sugarcane sequence generated is 382 Mb; 25,316 protein-coding gene models were predicted with more than 80 percent found to be colinear to comparable regions in sorghum.

Based in part on these positive results, in 2017, the JGI approved a CSP proposal from the international team to complete the first draft genome sequence of the polyploid sugarcane variety R570. Among the numerous benefits expected from a genome sequence are better understanding the roles of genes in traits such as sucrose accumulation and disease, as well as targeting genes to improve biomass and sugar yield for biofuel production.



Sugarcane. (Rufino Uribe, Flickr CC-SA 2.0)



Clusters of fruiting bodies emerge on and around trees in Armillariainfected areas in the fall. (Virág Tomity)

White Rot Fungi's Size Explained by Breadth of Gene Families Involved

Among the contenders for the world's largest living organism is a specimen of the fungus Armillaria ostoyae, first discovered two decades ago and thought to be a few millennia old then. The fungus is so large, it is spread over nearly four square miles — a space equivalent to one-sixth of Manhattan, or nearly 8,300 Olympic-sized swimming pools — and weighs as much as three blue whales.

Sheer size aside, Armillaria fungi dominate in an entirely different category: They are among the most devastating fungal pathogens, causing root rot disease in more than 500 plant species found in forests, parks, and vineyards. As white rot fungi, they are capable of breaking down all components of plant cell walls — cellulose, hemicellulose, and lignin — which interests bioenergy researchers looking for methods to cost-effectively convert plant biomass into alternative fuels. Reported in the October 30, 2017, issue of Nature Ecology & Evolution, an international team led by László G. Nagy of the Biological Research Center at the Hungarian Academy of Sciences and including JGI researchers sequenced and analyzed four Armillaria fungi, including A. ostoyae. The team then compared these genomes with those of related fungi to better understand the evolution of Armillaria's abilities to spread and infect, and effectively break down all components of plant cell walls.

"Armillaria species are some of the most devastating forest pathogens, responsible for forest decline in many temperate regions. There is thus a considerable interest in developing strategies against Armillaria spp, toward which understanding how they function in nature might be the first step," said study senior author Nagy. "We are interested in how Armillaria use plant cell wall degrading enzymes when confronted with potential host plants."

According to study co-author James Anderson from the University of Toronto, Armillaria species are extremely common in northern temperate forests and have nearly identical fruit-body morphology but different lifestyles. For example, A. gallica is primarily a degrader of hardwood and is not a pathogen of conifers. In contrast, A. ostoyae can be a highly aggressive root rot pathogen of firs, pines, and other conifers, causing up to 100 percent mortality of conifer seedlings.

"Both of these fungi have a major effect on the composition of forest tree species and on carbon cycling," said Igor Grigoriev, JGI Fungal Program Head and a co-author on the study. "Both can help us better understand mechanisms of lignocellulose degradation. Moreover, these were among the first representatives of the Physalacriaceae family and sequenced as a part of the JGI's 1000 Fungal Genomes initiative to produce reference genomes from each of more than 500 recognized families of fungi to fill in gaps in the Fungal Tree of Life."

Aside from *A. ostoyae*, the team also sequenced and analyzed the genomes of *A. cepistipes*, *A. gallica*, and *A. solidipes*. These genomes were then compared with 22 fungal genomes, many previously sequenced and annotated by the JGI. The researchers cataloged 20 gene families related to pathogenicity in the fungi, and identified enriched plant cell wall degrading enzyme families, the better to efficiently break down and access nutrients in dead wood. To help explain the unusually large fungal genomes in the *Armillaria* genus, they also found duplicated genes, suggesting *Armillaria* evolved primarily through gene family expansion and not transposable elements or "jumping genes." The *Armillaria* fungal genomes are all available on the JGI fungal genomics portal MycoCosm, along with the fungal genome sequences used for comparison.

Nagy pointed out that the research also sheds light on one of the long-standing questions in biology: the evolution of multicellularity. "Our comparative genomics and RNA-Seq data suggest that the development of rhizomorphs — shoestring-like structures that spread through the substrate in search for new food sources and can cross several feet underground — have a lot in common with that of fruiting bodies, both being complex multicellular structures," he said.

Just as well-organized teams can accomplish more than even talented individuals, the evolution of multicellularity is of great interest because multicellular organisms can carry out functions beyond the reach of single cells. In addition, *A. ostoyae's* collection of plant biomass-degrading enzymes could provide candidates for use with bioenergy feedstocks to generate biofuels and bioproducts that would be difficult to generate economically using more conventional approaches.

How Dry Rot Adapted to a New Ecological Habitat



Unlike white rots, brown rots break down only cellulose and hemicellulose, leaving lignin behind. The brown rot *Serpula lacrymans* is typically found in spruce and other conifers in boreal forests. As these trees have been harvested for constructing buildings, the dry rot fungus has migrated indoors and crossed borders, adapting to thrive in manmade environments. Due to its aggressive capacity to damage the wood in homes, bioenergy researchers have been interested in harnessing *S. lacrymans* toward breaking down plant mass for conversion to sustainable biofuels and bioproducts.

As reported January 5, 2018, in *The ISME Journal*, a team led by University of Oslo scientists and including JGI researchers compared the genomes of two strains of *S. lacrymans* — var. *lacrymans* from Europe and var. *shastensis* from North America — against a third fungal *Serpula* species, *S. himantioides* from Europe, that was sequenced and analyzed by the JGI.

The dry rot *Serpula lacrymans var. lacrymans*, shown attacking a dresser *(Inger Skrede)*

By comparing genetic information from similar organisms, researchers have gained insights on why *S. lacrymans* is so destructive in houses. The researchers found that *S. lacrymans* var. *lacrymans* has become an ecological specialist adapted to its indoor home, and has lost its ability to harness other woody substrates different brown rots could access. While *S. lacrymans* var. *shastensis* has a similar genome, the team suggests that it has not adapted like the other strain because its local environment has been less affected by human encroachment.

The team also did head-to-head, or confrontation, experiments involving these three *Serpula* fungi against three other brown rot fungi to see how the *Serpula* fungi have adapted in an environment with little to no competitors for resources. On wood blocks of pine, fir, and spruce, the researchers grew a *Serpula* species and a non-*Serpula* brown rot. The researchers found that both varieties of *S. lacrymans* were less aggressive at colonizing the wood blocks compared with the other brown rots, though *S. lacrymans* var. *lacrymans* decomposed more spruce than the other brown rot fungi. Additionally, *S. himantioides* outcompeted all the brown rots it was paired with, aggressively colonizing more of the wood blocks than its partner brown rots.

The team's results reflect the evolutionary gains and losses of *S. lacrymans* var. *lacrymans* in becoming a threat to homeowners. The brown rot has adapted to thrive on the limited nutrients found in wood inside homes in an environment that offers limited interactions with wild relatives.

Fungal Reproduction Regulated by Bacteria

a b

In heritable mutualisms, hosts pass on beneficial symbionts between generations. The origin of this relationship, though, is often antagonistic, and the parasite first needs to secure its own transmission before working with the host. Using the mutualistic relationship between the plant pathogenic fungus *Rhizopus microsporus* (Rm) and *Burkholderia endobacteria*, a collaborative effort led by researchers at Cornell University and JGI scientists was conducted to understand how the antagonistic-to-mutualistic transition occurs.

Sexual reproduction in *Rhizopus* microspores.

(a) Successful mating between fungi harboring bacteria. (b) Lack of sex between mates cured of endobacteria. (*Stephen Mondo*)

Rhizopus is a fungal pathogen of crops, including the candidate bioenergy feedstocks sunflower and maize, and a part of the oil-producing Mucoromycotina group, about which little is known. The Cornell team cultivated and experimented with the fungi and bacteria, while the JGI team sequenced and annotated a host genome (Rm ATCC 52813) as part of the 1000 Fungal Genomes project. *Burkholderia* is recognized as a mutualist but was predicted to have evolved from a parasitic interaction with their soil fungus host, *Rhizopus microsporus*.

Researchers found that endobacteria establishing control over reproduction was a likely key to this evolutionary transition. Using this model, researchers also generated the first transcriptomic data set of sexual reproduction in early fungi and discovered genes that are critical for this process.

As reported in the November 29, 2017, issue of *Nature Communications*, the team found that the fungus is highly dependent on the *Burkholderia endobacteri*a to proliferate both sexually and asexually. This dependence is consistent with the addiction model of mutualism evolution; in this case, the endobacteria control the expression of ras2-1, a gene crucial to reproductive development, making the fungus reliant on the continued presence of the bacteria. By studying the Rm-*Burkholderia* symbiosis model, the team was able to reconstruct the reproductive pathways in several branches of the fungal kingdom, generating the first transcriptomic data set of sexual reproduction in early fungi to find sex-relevant genes across fungi. The team also uncovered candidate genes, conserved across all Mucoromycotina, that appear to be involved in identifying pheromones that are critical for this fungal reproductive pathway.

Without the mutualistic Rm-*Burkholderia* relationship, *Rhizopus* is neither a plant pathogen nor able to reproduce with ease. So far, all endosymbionts discovered in fungi have been shown to substantially affect host lipid metabolism. As these oil-producing fungi are potential sustainable sources of alternative fuels, understanding fungal-bacterial relationships can shed light on how these interactions influence lipid production in Mucoromycotina and their potential for industrial use.

All in the Family: Focused Genomic Comparisons

Colonies of Aspergillus A. campestris, A. ochraceoroseus, and A. steynii. (Kirstine Ellen Lyhne, Technical University of Denmark)



Found in microbial communities around the world, *Aspergillus* fungi are pathogens, decomposers, and important sources of biotechnologically-important enzymes. Each *Aspergillus* species is known to contain more than 250 CAzymes, which break down plant cell walls and are of interest to DOE researchers working on the industrial production of sustainable alternative fuels using candidate bioenergy feedstock crops. Additionally, each fungal species is thought to contain more than 40 secondary metabolites, small molecules with the potential to act as biofuel and chemical intermediates.

In a study published the week of January 8, 2018, in the *Proceedings of the National Academy of Sciences*, a team led by researchers at the Technical University of Denmark, the JGI, and the DOE's JBEI report the first results of a long-term plan to sequence, annotate, and analyze the genomes of 300 *Aspergillus* fungi. These findings are a proof of concept of novel methods to functionally annotate genomes to more quickly identify genes of interest

"This is the first outcome from the large-scale sequencing of 300+ Aspergillus species," said study co-author Igor Grigoriev, head of the JGI Fungal Genomics Program. "With the JGI's strategic shift toward functional genomics, this study illustrates several new approaches for functional annotation of genes. Many approaches rely on experiments and go gene by gene through individual genomes. Using Aspergillus, we're sequencing a lot of closely related genomes to highlight and compare the differences between genomes. A comparative analysis of closely related species with distinct metabolic profiles may result in a relatively small number of species-specific secondary metabolism genes clusters to be mapped to a relatively small number of unique metabolites."

In the study, the team sequenced and annotated six Aspergillus species. A comparative analysis involving these genomes and other Aspergillus genomes — several of which were sequenced by the JGI — was then conducted, and allowed the team to identify biosynthetic gene clusters for secondary metabolites of interest.

"One of the things we found to be interesting here was the diversity of the species we looked at — we picked four that were distantly related," said study senior author Mikael R. Andersen, a professor at the Technical University of Denmark. "With that diversity comes also chemical diversity, so we were able to find candidate genes for some very diverse types of compounds. Moreover, we also showed how to solidify said predictions for a given compound by sequencing additional genomes of species known to produce the compound. By looking for genes found in all producer species, we can elegantly pinpoint the genes." Andersen spoke about the project at the 12th Annual JGI Genomics of Energy and Environment Meeting. Watch his talk at http://bit.ly/JGI2017Andersen.

Grigoriev added that to date, about 30 *Aspergillus* genomes have been published, an additional 25 genomes are publicly available from the JGI fungal genomes portal MycoCosm at **genome.jgi.doe.gov/Aspergillus**, and over 100 genomes are being sequenced and analyzed

Bacterial Biosensors

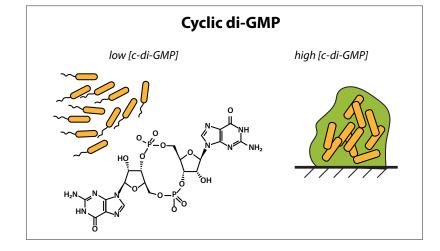
Cyclic di-GMP (guanine monophosphate) is found in nearly all types of bacteria and interacts with cell signaling networks that control many basic cellular functions. It plays an important role in regulating microbial cellulose production and biofilm formation, which affects a number of environments, including plants, soil, and the gut.

Scientists have tried to study cyclic di-GMP and cellular networks with fluorescent biosensors before, but such sensors require external illumination, making them difficult to use inside deep tissues. Additionally, all plants, many bacterial cultures, and environmental samples auto-fluoresce, making fluorescent biosensors ineffective.

To better understand the dynamics of this molecule, University of California, Berkeley, and JGI researchers developed the first chemiluminescent biosensors for measuring cyclic di-GMP in bacteria through work enabled by the JGI's CSP. The results were described July 20, 2018, in ACS Chemical Biology.

Using fluorescent biosensors to monitor the intracellular signaling network is inefficient on auto-fluorescing samples, rendering this approach useless when applied to plants and several bacterial cultures and environmental samples. Chemiluminescent biosensors typically use luciferases, enzymes that drive oxidative reactions to produce light. Engineered to produce light when bound to a molecule of interest, such sensors allow researchers to easily detect and measure their target molecule. There are two main types of luciferase-based biosensors: (1) complementation of split luciferase (CSL); and, (2) bioluminescence resonance energy transfer (BRET)-based biosensors. In general, CSL biosensors produce large signal changes, but their signal intensity is low due to weak reconstitution of luciferase activity. In contrast, BRET biosensors have high signal intensity but produce small signal changes in response to the molecule of interest. The team adapted a previously developed combination CSL-BRET approach called "Nano-lantern" to develop their new biosensor for cyclic di-GMP. The biosensor can produce both large signal changes and high signal intensity for imaging.

This new technology enables researchers to directly measure cyclic di-GMP levels in environmental samples or clinical isolates. As cyclic di-GMP also regulates enzymes involved in cellulose production in microbes, these biosensors may eventually lead to ways to improve cellulose-based biofuel synthesis. Additionally, having the biosensors opens the door to measuring real-time changes in cyclic di-GMP via live cell imaging.



The molecule cyclic di-GMP plays a key role in controlling cellulose production and biofilm formation. (Image courtesy of Hammond Lab, University of California, Berkeley)





Biogeochemistry

Building Sphagnum Genomic Resources

Sphagnum (peat moss) is an unassuming plant, but it thrives in nutrient-poor, acidic, and waterlogged environments, occupying every continent except Antarctica. Its impact on global carbon cycling and climate is estimated to be larger than any other single plant genus. Building on the availability of genomic information from two distinct environments, researchers partnering with the JGI are strengthening the case for Sphagnum as a versatile model system for studies of carbon cycling in diverse environments

Peatlands make up nearly 3 percent of Earth's land surface and accumulate partially decomposed matter. Researchers estimate that they hold about 25 percent of the world's soil carbon. In high latitudes, sphagnum can also act as a protective layer insulating permafrost from the warming regional temperatures. For this reason, the DOE is overseeing the SPRUCE (Spruce and Peatland Responses Under Changing Environments) project, assessing how northern peatland ecosystems respond to increases in temperature and elevated concentrations of atmospheric carbon dioxide.

In New Phytologist, researchers led by David Weston of Oak Ridge National Laboratory make the case for sphagnum as "an unparalleled model system for ecological and evolutionary genomics." Understanding the nature and genetic basis of functional traits in sphagnum growth and decomposition is critical to predicting the ecosystem response to climate change. Through the CSP, researchers are aiming to have two high-quality reference genomes for peat-forming species that occupy different microhabitats: S. fallax and S. magellanicum. The draft genome for S. fallax is already available on the JGI plant portal Phytozome, and the S. magellanicum genome is currently being assembled. These resources would provide the research community with the ability to address questions including sphagnum's associations with methane oxidizing bacteria, nutrient uptake, and productivity and decomposition rates.

Peatlands, like other wetlands, can act as a carbon sink or carbon source. Sulfur cycling is a carbon flux regulator since sulfate-reducing microbes prevent methane production by routing the carbon away from methanogens. In *The ISME Journal*, a team led by researchers from the University of Konstanz and the University of Vienna reported recovering draft metagenome-assembled genomes for seven novel species of Acidobacteria through an approved CSP proposal. The researchers found that these species encode a sulfate reduction pathway, but also have genes for sulfide oxidation. The findings lead them to speculate that these microbes could use the same pathway for either sulfur reduction or sulfur oxidation. The pathway's reversibility could be a factor in determining if the peatlands act as a carbon sink or a carbon source.



Sphagnum fallax (Image courtesy of Jonathan Shaw, Duke University)

Tracking Microbial Diversity through the Terrestrial Subsurface

The majority of microbial life deep underground has yet to be characterized. A pressing question in subsurface microbiology is how these organisms, and their capacities for carbon, nitrogen, and sulfur cycling, are distributed along vertical transects underground. In marine or freshwater aquatic systems, sampling microbial communities along depth gradients is fairly straightforward. For the terrestrial subsurface, however, it's difficult to obtain samples without contamination from drilling fluids or equipment.

In collaboration with a team led by longtime JGI collaborator Jill Banfield of the University of California, Berkeley, and Cathy Ryan of the University of Calgary in Canada, JGI researchers investigated samples collected over the course of a Utah geyser's complex, five-day eruption cycles. Located in Paradox Basin, Crystal Geyser is a CO₂-driven, cold-water geyser fed by a series of underlying aquifers separated by layers of shale and mudstone. Its five-day eruption cycle has three phases: minor eruptions, major eruptions, and recovery. In each eruption phase, the microbial communities sampled were sourced from groundwater at different depth intervals. The results were reported January 29, 2018, in *Nature Microbiology*.

"For microbiology and geochemistry, it's a really cool system because through its cycle the geyser feeds off these different aquifers and through these eruptions you can get access to the stratified system and the microbes that come up," said JGI Microbial Program head Tanja Woyke.

Genome-resolved metagenomics, single-cell genomics, and geochemical analyses were integrated to show that samples taken during each phase contain microbial communities that are distinctive in terms of both composition and metabolic function. For Banfield, the key achievement is that cultivation-independent genomic methods directly linked highly novel groups of organisms to their sources below ground. This was possible due to the unprecedented combination of genomics methods with geochemistry and hydrological information. She also noted it is important that major methods in microbial ecology, genome-resolved metagenomics, and single-cell genomics were applied to the same samples so that these key methods could be compared.

Prior work by Banfield's group hinted at the enormous phylogenetic diversity of previously unknown bacteria and archaea in this system, including deeply branching organisms of the Candidate Phyla Radiation (CPR) and the DPANN superphylum of extremophile archaea named for the first five groups discovered: Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, and Nanohaloarchaea.

"Ecosystems in the deep terrestrial subsurface represent the final frontier in the exploration of the diversity of life on our planet. Little is known about how these ecosystems are structured and how organisms in these

environments make their living without sunlight," said study co-first author Alexander Probst, a former postdoctoral researcher in Banfield's lab who is now at the University of Duisburg-Essen in Germany.

In the current study, genomes were reconstructed for 505 different bacterial and archaeal organisms from 104 different phylum-level lineages, including nine potentially novel phyla. Approximately 57 percent of the organisms were significantly enriched in a specific phase of the geyser's eruption cycle, and thus could be sourced to a particular groundwater depth interval.

The shallow groundwater community was comprised primarily of one *Sulfurimonas* species, with a handful of other bacteria and a few archaea. The majority of the microbes in the intermediate depths and associated with the geyser itself belonged to the CPR, but the most abundant was a member of Gallionellaceae. And the deepest groundwater had the greatest concentration of DPANN archaea and Candidatus "Altiarchaeum," a remarkable archaeon with miniature grappling hooks over its cell surface.



Many of these subsurface-dwelling microbes are presumed to be symbiotic; the dearth of resources available from the environment means that they have to rely more on one another to survive. Organisms that are able to generate energy through chemical reactions (such as hydrogen or iron oxidation) and then use that energy to build up carbon are considered "primary producers" that sustain the other bacteria and archaea in their communities.

Banfield's group also previously reported that organisms in the Crystal Geyser system employ three different carbon fixation mechanisms. In the current study, they associated the mechanism with the highest energy requirement with the shallow groundwater community, while the mechanism with the lowest energy requirement was associated with the deep groundwater community.

Moreover, noted Probst, the analysis identified genomes of hundreds of putative symbiotic organisms, including a potential symbiont of one the most abundant organisms in these subsurface fluids, an archaeon that binds CO₂ in the subsurface and converts it to organic matter.

Crystal Geyser during its "major eruption" period, when sufficient discharge is produced to cause overland flow to the Green River. (Cathy Ryan)

Uncovering the Role of Nitrite-Oxidizing Bacteria in the Marine Carbon Cycle

Most of the ocean is shrouded in darkness. Sunlight penetrates the first 200 meters only of ocean waters, and no light reaches the water deeper than 1,000 meters. Beneath the ocean's surface, the sun's rays can only go so far. Some light reaches the mesopelagic region, which lies 200–1,000 meters beneath the waters, but no light reaches the bathypelagic (1,000–4,000 meters) or abyssopelagic (4,000–6,000 meters) regions. Despite the lack of sunlight, these waters are teeming with microbial life and contribute significantly to regulating the planet's global cycles.

This "dark ocean" is estimated to be 90 percent of the ocean's volume, though very little is known about the microbes that are involved in trapping and utilizing the carbon that reaches these regions, keeping it out of the planet's atmosphere. In the November 24, 2017, issue of *Science*, a team led by Bigelow Laboratory for Ocean Sciences researchers and including JGI scientists identified the most abundant and globally distributed nitrite-oxidizing bacteria (NOB) in the oceans through single-cell genomics and community meta-omics.



View of Saanich Inlet. (Courtesy of the British Columbia Ministry of Transportation, Flickr CC BY-NC-ND 2.0)

The researchers generated and analyzed nearly 3,500 SAGs of bacteria and archaea from some 40 samples collected from around the world. From these SAGs, they identified over a hundred NOB, with the bulk belonging to the *Nitrospinae* phylum and a handful belonging to the *Nitrospirae* phylum.

The results indicated that these are the predominant NOB in the ocean. Additionally, based on the partial NOB genomes the researchers assembled, they found genes with the capacity for carbon fixation, indicating these bacteria are involved in regulating the carbon cycle, and genes suggesting that the NOBs are strictly specialized, producing energy solely from nitrite oxidation.

Finally, the team extrapolated the amount of carbon fixed by NOBs, based on the amounts of dissolved inorganic carbon from incubated samples and scaling that calculation to the volume of the ocean's mesopelagic region where the largest concentration of NOBs are found. Though NOBs account for less than 5 percent of the populations in each of the various levels of the so-called "dark ocean," they capture, or "fix," close to half of the inorganic carbon in the waters.

The researchers' calculations suggest that NOBs capture as much as 1 Petagram (1 billion metric tons) of carbon annually. This is similar to prior estimates of the total carbon fixation in the dark ocean and indicates that NOBs play a much larger role in regulating the marine carbon cycle than previously thought.

A Pan-Genome for Antarctic Archaea

Conditions such as extreme cold, high levels of salinity, and the sheer distance from other parts of the world have kept microbial populations on Antarctica distinct and unique. In collaboration with the JGI, University of New South Wales microbial ecologist Rick Cavicchioli and his team have been studying the haloarchaea that thrive in the very salty waters of Deep Lake to better understand how they've adapted to these conditions, information that could have biotechnological applications. The work was reported June 20, 2018, in *Microbiome*.

Haloarchaea flourish in hypersaline environments. Researchers are interested in understanding how these microbes have learned to adapt from marine to hypersaline conditions by studying the microbial communities in

Antarctic lakes, some of which have salinities 10 times the level found in seawater. By collecting and sequencing dominant haloarchaeal sequences from six hypersaline lakes, the team focused on understanding the genomic variation in haloarchaea across East Antarctica.

The implications of the current study are important both to understand genomic variation in Antarctic lake environments and to characterize regional and global biogeography of haloarchaea and the structure and extent of the Antarctic microbial pan-genome. Since these microbes thrive under extreme temperature and salinity conditions, understanding their adaptations offers insights into microbial evolution, as well as how Antarctic microbial communities contribute to global biogeochemical cycles and possibly cold-temperature enzymes for biotechnological use.

Cavicchioli and his team compared two strains of Halorubrum lacusprofundi from Deep Lake and Rauer 1 Lake, one of the lakes in the nearby Rauer Islands. Additionally, they compared metagenome data generated from samples collected at four Rauer lakes to assess population level genomic variation. The genome and metagenome data are available through the JGI's IMG/M platform.

These data allowed researchers to define a haloarchaea pan-genome. Described as the total pool of genetic material comprised by all members of a species, the pan-genome of haloarchaea contains the parts of the genome shared across all haloarchaea, along with the pool of genes considered the flexible genome content that was likely acquired due to gene transfer events. Such events can occur in response to viral infections and are important for virus-host interactions. Analysis of strain variation showed that haloarchaea generally have a primary replicon (chromosome) that is highly conserved and secondary replicons that are highly variable but organized in blocks, or "islands," suggesting that the acquisition via horizontal gene transfer mechanisms involved significant sized pieces.

Much of the strain variation appeared related to defense against viruses. Experiments infecting H. lacusprofundi with an Antarctic halovirus demonstrated differences in resistance between two strains. Perhaps the largest implication of this study is the evidence that H. litchfieldiae and H. lacusprofundi are found across all six lakes, suggesting that these species are endemic to Antarctica and distinct from other hypersaline environments.



Antarctica's Deep Lake. (Rick Cavicchioli)

DAS Tool for Genome Reconstruction from Metagenomes

Through the JGI's Emerging Technologies Opportunity Program (ETOP), researchers have developed and improved upon a tool that combines existing DNA sequence binning algorithms, allowing them to reconstruct more near-complete genomes from soil metagenomes compared with other methods.

Understanding how individual microbes interact with each other in a community, and characterizing their individual contributions, is essential to understanding their ecosystem functions in processes ranging from nutrient cycling to plant health and growth. Developing a scalable computational approach that provides researchers with the ability to recover and reconstruct individual genomes, particularly from incredibly complex soil microbial communities, is crucial to understanding how these microbes respond and adapt to environmental changes.

Researchers have been utilizing metagenomics as a way to study a collection of hard-to-culture microbes, but the efficacy of this approach would be greatly increased by the ability to reconstruct complete genomes of individual, unknown microbes in the community. Since the inception of ETOP in 2013, a team of researchers has been working on improving methods for isolating and characterizing entire microbial genomes from sequences generated from environmental samples. The work aligns with the ETOP mission of providing JGI users with the most current technology and expertise to address pressing energy and environmental scientific challenges.

In *Nature Microbiology*, a team led by longtime JGI collaborator Jill Banfield of the University of California, Berkeley, and involving JGI researchers reported on the development and validation of a dereplication, aggregation, and scoring tool (DAS Tool). To validate DAS Tool, the team applied it to assemblies from both simulated and real microbial communities and examined the bins obtained. The simulated communities were those created for the CAMI Challenge (see page 27) with varying numbers of genomes to demonstrate its capacity to tackle metagenomes with low (40 genomes) to high (596 genomes) complexity. The environmental data sets included metagenomic data from Crystal Geyser in Utah (another collaboration between the Banfield Lab and the JGI), human microbiomes, natural oil seeps in Santa Barbara, California, and soil from the Angelo Coast Range Reserve in California.

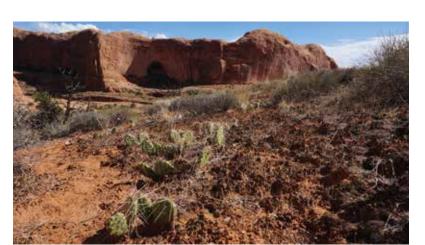
Using DAS Tool on data from soil samples, researchers were able to reconstruct 79 minimally contaminated (<5%) draft genomes to >70% completeness. Of those 79 genomes, 26 were high-quality draft genomes with >90% completeness. These results, noted JGI User Programs Deputy Susannah Tringe, suggest that extracting high-quality genomes from soil metagenome data is no longer "nearly impossible."

The team reported that "almost always, [DAS Tool] extracted considerably more genomes from complex metagenomes than any of the single binning tools alone," primarily because of its capacity to harness the strengths of multiple binning algorithms. The researchers added that DAS Tool's flexibility in integrating results from both manual and automated binning methods also extends to harnessing future binning tools. DAS Tool is an example of how JGI users can benefit from the ETOP program, as tools such as this can help advance genome-centric analyses.

DAS Tool will significantly enhance the ability to assemble often-fragmentary DNA sequences from environmental samples of very complex microbial communities into complete or near-complete genomes to aid characterization of community members and their genome-encoded metabolic capabilities and interactions. Supported by the JGI's ETOP, this new tool will now be available for JGI users interested in exploring microbial communities engaged in mission-relevant energy and environmental processes.

Microbial Metabolism in Biocrusts

Arid lands cover some 40 percent of Earth's terrestrial surface and are too dry to sustain much vegetation. In spite of this, they are home to diverse communities of microorganisms — including fungi, bacteria, and archaea — that dwell together within the uppermost millimeters of soil. These biological soil crusts, or biocrusts, can exist for



Biocrust among one of its many natural habitats, near the Corona Arch, Moab, Utah (Tami Swenson) extended periods in a desiccated, dormant state. When it does rain, the microbes become metabolically active, setting in motion a cascade of activity that dramatically alters both the community structure and the soil chemistry.

"These biocrusts and other soil microbiomes contain a tremendous diversity of both

microbes and small molecules [metabolites]. However, the connection between the chemical diversity of soil and microbial diversity is poorly understood," said Trent Northen, a Berkeley Lab senior scientist affiliated with the Biosciences Area's Environmental Genomics and Systems Biology division and the JGI.

As reported January 2, 2018, in *Nature Communications*, Berkeley Lab researchers led by the Northen Lab reported that specific compounds are transformed by and strongly associated with specific bacteria in native biological soil crust (biocrust) using a suite of tools Northen calls "exometabolomics." Understanding how microbial communities in the biocrusts adapt to their harsh environments could provide important clues to help shed light on the roles of soil microbes in the global carbon cycle.

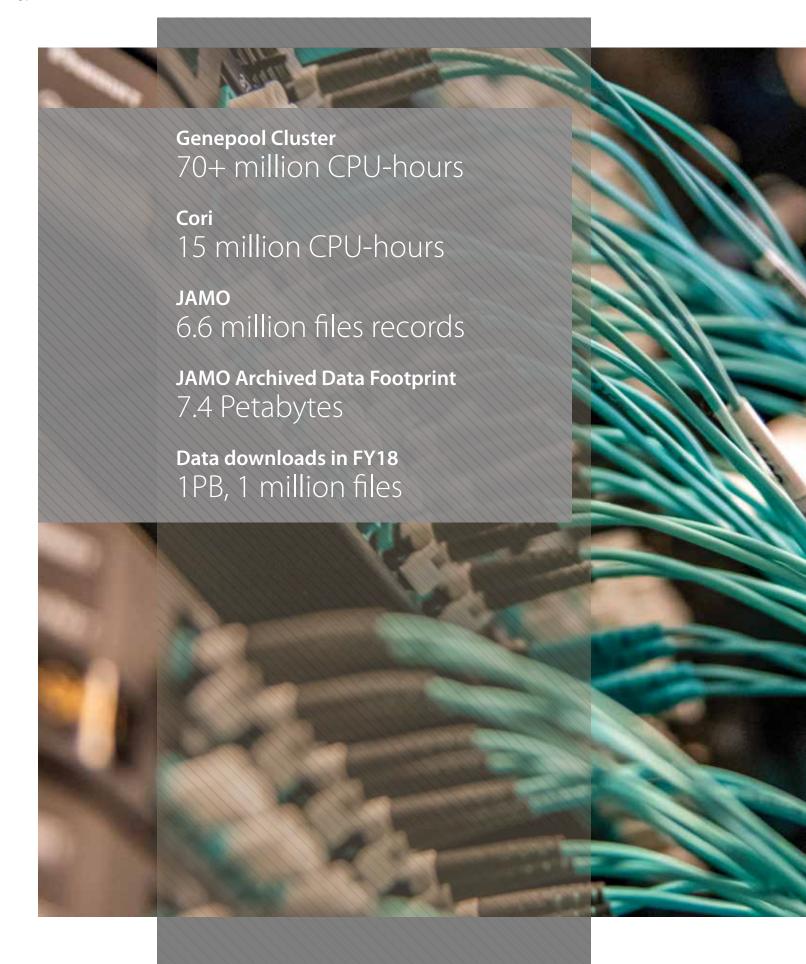
The work follows a 2015 study that examined how specific small molecule compounds called "metabolites" were transformed in a mixture of bacterial isolates from biocrust samples cultured in a milieu of metabolites from the same soil. "We found that the microbes we investigated were 'picky' eaters," Northen said. "We thought we could use this information to link what's being consumed to the abundance of the microbes in the intact community, thereby linking the biology to the chemistry."

In the new study, the investigators set out to determine whether the microbe-metabolite relationships observed in the simplified test-tube system could be reproduced in a more complex soil environment. Biocrusts from the same source—representing four successive stages of maturation—were wet, and the soil water was sampled at five time points. The samples were analyzed by liquid chromatography-mass spectrometry to characterize the metabolite composition (metabolomics), and biocrust DNA was extracted for shotgun sequencing to measure single copy gene markers for the dominant microbe species (metagenomics).

"When we compare the patterns of metabolite uptake and production for isolated bacteria that are related to the most abundant microbes found in the biocrusts, we find that, excitingly, these patterns are maintained," said Northen. That is, increased abundance of a given microbe is negatively correlated with the metabolites that they consume and positively correlated with metabolites that they release.

When active, biocrusts take up atmospheric carbon dioxide and fix nitrogen, contributing to the ecosystem's primary productivity. They also process organic matter in soil, modifying key properties related to soil fertility and water availability.

"This study suggests that laboratory studies of microbial metabolite processing can help understand the role of these microbes in carbon cycling in the environment. This study gets us closer to understanding the complex food webs that are vital in nutrient dynamics and overall soil fertility," said study first author Tami Swenson, a scientific engineering associate in the Northen lab. The team is currently working on expanding these studies to capture a greater fraction of microbial diversity. Ultimately, this may enable the prediction of nutrient cycling in terrestrial microbial ecosystems, and perhaps even manipulation by adding specific metabolites.





Computational Infrastructure

The JGI established a Data Science and Informatics (DSI) department in order to build seamless connectivity between JGI and the broader computational ecosystem. Led by Chief Informatics Officer Kjiersten Fagnan, the DSI is responsible for the software and hardware infrastructure built to support JGI's user community, operations, and scientific research.

The JGI's capacity as a next-generation genomics user facility has generated petabytes of high-quality sequence data and analysis. To support this workload, the JGI invested significant resources in its high-performance compute cluster, Denovo; 192 nodes of capacity in the Cori supercomputer, as well as storage and web infrastructure. The JGI also procured 20 1.5TB large-memory nodes to support the assembly and analysis of large datasets.

The DOE Knowledgebase (KBase) and the 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 application program interface (API) to an Elastic Search instance built on top of the JGI Archive and Metadata Organizer (JAMO) developed by the Sequence Data Management Group's Harika Tandra and Chris Beecroft. The JGI and KBase deployed the JGI pipelines for metagenome assembly and an open source version of metagenome annotation 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 software development kit (SDK). The KBase team ran a number of workshops at the 2018 Genomics of Energy & Environment Meeting where users were able to gain hands-on experience with the platform, data resources, and tools. Following on the successful completion of the first six JGI-NERSC Facilities Integrating Collaborations for User Science (FICUS) proposals, the JGI, KBase, and NERSC held a joint proposal call to help scientists perform large-scale computational analyses of sequence data.

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. In 2010, all of the JGI's computational resources were moved to NERSC, enabling its researchers and users to devote more of their time to cutting-edge genomics research. Additionally, the infrastructure advancements to Denovo and other JGI portals mean rapid and smooth access for users across the globe.

The JGI used more than 10 million central processing unit hours on NERSC's petascale supercomputer, Cori, which was acquired in 2016. Cori represents a \$70-million investment in data-intensive and high-performance computing infrastructure. Many of these calculations could not have been completed on the Genepool cluster because of the computing scale required. NERSC has also deployed a Containers-as-a-Service (CaaS) platform based on Docker container technology. It can be used to implement web sites and science gateways, workflow managers, databases and key-value stores, and all sorts of network services that can access NERSC systems and storage on the back end. This will enable the JGI to use more robust and scalable back ends for data analysis platforms.

In April 2018, the JGI and the Biosciences Engineering Division and the Computing Sciences Area of Berkeley Lab organized a Machine Learning Hackathon. The first three days were hands-on sessions led by instructors from The Data Incubator, and the final two days involved groups working together to use machine learning to investigate different hypotheses. The projects ranged from microbiome analysis to new classification algorithms for viruses.

Finally, the JGI partnered with the Oak Ridge Leadership Compute Facility (OLCF) at Oak Ridge National Laboratory (ORNL) in Tennessee to establish a second copy of irreplaceable sequence data. In addition to providing a secure location for JGI data, the repository is accessible to scientists that are utilizing OLCF resources like Summit, the fastest supercomputer in the world. The DSI's Beecroft and Seung-Jin Sul led this effort.



Appendix A

Acronyms at a Glance

AF	Alignment fraction	ECP	Exascale Computing Project
ANI	Average Nucleotide Identity	EMSL	Environmental Molecular Sciences
BER	DOE Office of Biological and Environmental Research	ЕТОР	Laboratory (at PNNL) Emerging Technologies Opportunity Program
BERAC	Biological and Environmental Research Advisory Committee	FACS	Fluorescence Activated Cell Sorting
BESC	BioEnergy Science Center (ORNL)	FICUS	Facilities Integrating Collaborations for
BGC	Biosynthetic Gene Cluster	FIGU	User Science
BLAST	Basic Local Alignment Search Tool	FISH	Fluorescence In Situ Hybridization
BOOST	Build Optimization Software Tools for DNA Synthesis	GEBA	Genomic Encyclopedia of Bacteria and Archaea
BRC	•	GLBRC	Great Lakes Bioenergy Research Center
DNC	Bioenergy Research Center (i.e., BESC, CABBI, CBI, GLBRC, JBEI)	GOLD	Genomes OnLine Database
CABBI	Center for Advanced Bioenergy and	HPC	High Performance Computing
	Bioproducts Innovation	HPSS	High Performance Storage System
CAMI	Critical Assessment of Metagenome Interpretation	IEP	Industry Engagement Program
СВІ	Center for Bioenergy Innovation	IGB	Integrative Genomics Building
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats	IMG/M	Integrated Microbial Genomes & Microbiomes system
CRADA	Cooperative Research & Development	ISM	Integrated Safety Management
	Agreement	ITS	Integrated Tracking System
CRAGE	Chassis-Independent Recombinase Assisted Genome Engineering	JAMO	JGI Archive and Metadata Organizer
CRT	Computational Research and Theory Building (LBNL)	JBEI	Joint BioEnergy Institute
		KBase	DOE Systems Biology Knowledgebase
CSP	Community Science Program	LANL	Los Alamos National Laboratory
DOE	Department of Energy	LBNL	Lawrence Berkeley National Laboratory
DSI	Data Science and Informatics	LLNL	Lawrence Livermore National Laboratory

MAG	Metagenome-Assembled Genome
MIMAG	Minimum Information about a Metagenome-Assembled Genome
MISAG	Minimum Information about a Single Amplified Genome
MGM	Microbial Genomics & Metagenomics
NERSC	National Energy Research Scientific Computing Center
NGEE	Next-Generation Ecosystem Experiments
NREL	National Renewable Energy Laboratory
ORNL	Oak Ridge National Laboratory
РМО	Project Management Office
PNNL	Pacific Northwest National Laboratory
SAC	Scientific Advisory Committee
SAG	Single Amplified Genome
SFA	Scientific Focus Area
SIP	Stable Isotope Probing

Work Initiation Process

WIP

Appendix B

Glossary

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

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

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

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

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

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 or sequester contaminants and other unwanted substances in waste and other substances.

Biosynthetic gene cluster: Genes responsible for the production of secondary metabolites are often clustered in their genomes and often contain all genes required for biosynthesis of precursors, assembly of the compound scaffold, tailoring of the scaffold and additional genes for resistance, export and regulation.

Bridge amplification: A proprietary technique used by Illumina sequencing platforms to generate singlestranded 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 by Chromatin Immunoprecipitation.

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 by DNA Affinity Purification.

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

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

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

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

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

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

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

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

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: At the 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.

Metabolites: The organic compounds occurring from an organism's metabolism. Secondary metabolites are all the other metabolites that are unique to a given species, and may also be known as the metabolites that derive from "specialized metabolism."

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.

Metaproteomics: Comprehensive characterization of all proteins expressed by environmental samples at a given point in time.

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

Microbe: Another name for a microorganism.

Microbiome: A 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.

Omics: A term used to encompass multiple biological fields of study ending in "omics," i.e., genomics, metabolomics, metagenomics, etc.

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 or membrane-bound organelles.

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

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

RB-TnSeq: Randomly-Barcoded Transposon Sequencing is a technique to generate sequencetagged 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, 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 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 among the planet's atmosphere, land, and oceans.

Synthetic biology: A field of research concerned with purposeful editing of biological systems. For the 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

2018 User Program Supported Proposals

Community Science Program (CSP)

INVESTIGATOR	AFFILIATION	DESCRIPTION
Banfield, Jill	University of California, Berkeley	LBNL Watershed Function SFA: Biogeochemical Dynamics from Genome to Watershed Scales
Beam, Jake	Bigelow Laboratory for Ocean Sciences	Emergent Properties of Marine and Freshwater Sediment Chemolithoautotrophic Microbial Communities
Berka, Randy	Novozymes	Genus-wide Genomics of the Biomass-Degrading and Plant-Beneficial <i>Trichoderma</i>
Bhaya, Devaki	Stanford University	Cyanophages and Cyanobacteria in Extreme Environments
Buckley, Dan	Cornell University	Microbial Metabolic Dependency and its Impacts on the Soil Carbon Cycle
Catalan, Pilar	Universidad de Zaragoza (Spain)	Genomic Characterization of the <i>Brachypodium Polyploid</i> Model to Unravel Bases of Success of Polyploidy in Flowering Plants
Coleman, Maureen	University of Chicago	Integrated Ecosystem Genomics across a Vast and Vital Freshwater System
Crump, Byron	Oregon State University	Combined Metagenomics and Metatranscriptomics to Evaluate Dissolved Organic Matter Metabolizing Microbes in Big Rivers of the U.S.
Cullen, Daniel	Forest Products Laboratory	Gene Expression in the Unusual White Rot Fungus Phlebiopsis gigantea
Dijkstra, Paul	Northern Arizona University	Partitioning Flux between Entner-Douderoff and Embden-Meyerhof-Parnas Glycolysis in Soil Communities
Duplessis, Sebastien	INRA (France)	Rust Pangenomics: Understanding the Diversity and Potential Impact of Rust Fungi — a Systematic Collection of Genomes & Transcriptomes
Fukami, Tadashi	Stanford University	Genomic Basis of the Ecological Success of Nectar Yeasts in their Carbon-Stressed and Nitrogen- Limited Environments

Community Science Program (CSP) (continued)

INVESTIGATOR	AFFILIATION	DESCRIPTION
Garre, Victoriano	University of Murcia (Spain)	Exploring the Role of DNA Methylation in Biofuel Production, Environmental Sensing and Development in Basal Fungi
Huffaker, Alisa	University of California, San Diego	Systems Analysis of Secondary-Metabolism- Mediated Microbial Community Interactions in Sorghum and Maize Enabled by Comparative Metabolomic and Transcriptomic Genome-Wide Association Studies
Kalluri, Udaya	Oak Ridge National Laboratory	Reciprocal Impacts of Modified Plant Cell Wall and Associated Microbiome
Kistler, Harold	USDA ARS Cereal Disease Lab, University of Minnesota	Mechanisms of Co-Evolutionary Adaptation of Soil Microbes
Leebens-Mack, James	University of Georgia	Open Green Genomes: a Framework for Comparative Plant Genomics
Liu, Yu	UT Southwestern Medical Center	Determination of Fungal Chromatin Regulatory Network and Its Impact on Gene Expression
Mandadi, Kranthi	Texas A&M University	Gene Atlas of Diverse Grass-Microbe Interactions in Brachypodium and Setaria
Martinez-Gomez, Norma	Michigan State University	Dissecting the Genetic and Functional Diversity of Rare Earth-dependent Plant-Microbiome Interactions by Sequencing Analysis
Martiny, Jennifer	University of California, Irvine	Investigation of Diel Variation in Microbial Decomposition Processes
Reese, Brandi	Texas A&M University- Corpus Christi	Dynamics of Seasonal and Diurnal Fluctuations in a Wetland Mangrove Ecosystem
Stuart, Rhona	Lawrence Livermore National Laboratory	Flipping the Switch: the Molecular Mechanisms Underlying Trophic Strategy Versatility in Parasitic Chytrids
Sullivan, Matt	Ohio State University	Elucidating Viral "Dark Matter" and Biogeochemical Impacts in Extreme Environments

Community Science Program (CSP) (continued)

INVESTIGATOR	AFFILIATION	DESCRIPTION
Tas, Neslihan	Lawrence Berkeley National Laboratory	Deciphering Microbial Functions at Soil-Aquatic Interfaces: Fate of Carbon Exports into High- Elevation Streams
Tiffin, Peter	University of Minnesota	Leveraging Natural Diversity to Identify the Genetic Basis of Microbial Success in Legume-Rhizobia Mutualism and Non-Host Environments
U'Ren, Jana	University of Arizona	Comparative and Population Genomics of Xylariaceae: Exploring the Roles of Endophytic Fungi in Lignocellulose Degradation, Nutrient Cycling, and Secondary Metabolite Production
Umen, James	Donald Danforth Plant Science Center	Comparative Genomics and Germ Plasm Diversity in Acutodesmus, a Green Microalgal Bioenergy Feedstock Candidate with Potential for Breeding and Hybridization
Yang, Xiaohan	Oak Ridge National Laboratory	High Quality Genome Sequencing of <i>Agave</i> tequilana, a Bioenergy Crop with High Drought Tolerance and Low Biomass Recalcitrance
Zhang, Ru	Donald Danforth Plant Science Center	High-Throughput Sequencing and Metabolomics Enabled Phenomics to Investigate Integration of Heat and Circadian Responses in the Model Green Alga Chlamydomonas reinhardtii

Small-Scale Microbial/Metagenome

INVESTIGATOR	AFFILIATION	DESCRIPTION
Atsumi, Shota	University of California, Davis	Genomic Analysis, and Reconstruction of Isobutyl Acetate Tolerance in Escherichia coli
Barnhart, Elliott	U.S. Geological Survey	Multi-omic Sequencing of Sulfate Transition Zones in the Terrestrial Subsurface with Recalcitrant Carbon
Epstein, Slava	Northeastern University	Microbial Stem Cell Hypothesis
Högfors-Rönnholm, Eva	Novia University of Applied Sciences (Finland)	Microbial Community Structure and Function of an Actual and Potential Acid Sulfate Soil
Izquierdo, Javier	Hofstra University	Plant Growth-Promoting Bacteria from the Rhizosphere of the Beachgrass Ammophila breviligulata

Small-Scale Microbial/Metagenome (continued)

INVESTIGATOR	AFFILIATION	DESCRIPTION
Liu, Xiao-Jun Allen	University of Massachusetts Amherst	Disentangling the Relative Contributions of the Microbiome and Physical Protection in Soil Response to Long-Term Environmental Stress
Mason, Olivia	Florida State University	Coupled Plant: Microbe Interactions Mediate Carbon and Biogeochemical Cycling in the Marsh Rhizosphere
McMahon, Katherine	University of Wisconsin- Madison	Biogeochemical Gradients Structure Mercury- Methylating Microbial Communities in a Reservoir System
Nicol, Graeme	Ecole Centrale de Lyon (France)	Determining the Interaction of Viruses with Prokaryotic Hosts Controlling Nitrogen Cycling in Soil
Peay, Kabir	Stanford University	How Does Precipitation Impact the Taxonomic and Functional Diversity of the Populus Trichocarpa Soil Microbiome?
Potnis, Neha	Auburn University	Unraveling the Diversity of Plant-Associated Saprophytic/Non-Pathogenic Bacteria and Their Role in Plant Health and Plant-Pathogen Interactions
Richardson, Ruth	Cornell University	Metagenomic Exploration of Microbial Communities involved in Carbon and Sulfur Cycling in Two Central New York State Peatlands
Selbmann, Laura	University of Tuscia (Italy)	Metagenomic Reconstruction of Endolithic Communities from Victoria Land, Antarctica
Thamatrakoln, Kim	Rutgers University	The Role of Light and Nutrient Limitation on Algal Host-Virus Interactions in Natural Populations and Subsequent Impacts on Carbon Export and the Biological Pump
Treusch, Alexander	University of Southern Denmark (Denmark)	Impacts of Climate Change Induced Flooding of Coastal Soils on Carbon Cycling and Sequestration – 3.0
Voges, Mathias	Stanford University	Elucidating the Biological Role of Root Exudates: Investigating the Transcriptional Response of Pseudomonas simiae WCS417 to Novel Specialized Plant Metabolites.
Ziels, Ryan	University of British Columbia (Canada)	Genome-Centric Metagenomics and Metatranscriptomics to Resolve Adaptive Capacities of Methane-Producing Biofilms to High Salinity Concentrations

DNA Synthesis

INVESTIGATOR	AFFILIATION	DESCRIPTION
Ando, Nozomi	Princeton University	Evolution of Allosteric Regulation in the Ribonucleotide Reductases: Investigating a Family of Ancient Enzymes
Bacik, John	Cornell University	Gene-to-Structure Study of Microbial Extracellular Electron Transport Mechanisms
Donia, Mohamed	Princeton University	A Natural Model System for Studying Functional Evolution and Synthetic Design Principles of Non- Ribosomal Peptide Synthetases
Erlemont, Renaud	CSU Long Beach	Prothunt: Environmental Multi Activity Proteins for Improved Biomass Deconstruction, from Sequenced Metagenomes to Protein Biochemistry
Maeda, Hiroshi	University of Wisconsin- Madison	Mapping Multi-Substrate Specificities of Aminotransferases Across the Plant Kingdom
McDermott, Timothy	Montana State University	Optimization of Methylamine Conversion to Methane via Synthetic Biology
Nair, Nikhil	Tufts University	Developing Bacillus Subtilis Endospore-Display as a Platform for Rapid Cell-Free Pathway Prototyping and Bioprocessing
Narayan, Alison	University of Michigan	Flavin-Dependent Monooxygenases as Tools for Chemistry
Neugebauer, Monica	University of California, Berkeley	Mining The Diversity Of Environmental Organisms To Discover Novel Halogenation Catalysts
Sammond, Deanne	National Renewable Energy Laboratory	Computationally Designed Cellulases to Decrease the Cost of Biofuels
Szymanski, Dan	Purdue University	A Systems Biology Approach to Analyze Plant Responses to Metabolic Stress
Tyo, Keith	Northwestern University	Biosynthesis of Bioprivileged, Linear Molecules via Novel Carboligase Reactions
Wackett, Lawrence	University of Minnesota	Synthetic Biology to Exploit Nature's Diversity of Hydrocarbon Structures for Fuels and Chemicals
Wang, Jue	University of Washington	Optimizing a Synthetic One-Carbon-Assimilation Pathway via Enzyme Bioprospecting and Engineering
Young, Eric	Worcester Polytechnic Institute	Engineering a Yeast Cell Factory in Saltwater
Zalatan, Jesse	University of Washington	Rewiring Endogenous Gene Expression Programs in Bacteria with CRISPR Activation

Emerging Technologies Opportunity Program (ETOP)

The ETOP's primary purpose is to develop and support selected new technologies that JGI could establish to add value to the high throughput sequencing it currently carries out for its users.

INVESTIGATOR	AFFILIATION	DESCRIPTION
Alper, Hal	University of Texas at Austin	Pre-Optimized Cell Free Lysates for Rapid Prototyping of Genes and Pathways
Jewett, Michael	Northwestern University	

Facilities Integrating Collaborations 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..

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 Department of Energy's Office of Biological and Environmental Research.

INVESTIGATOR	AFFILIATION	DESCRIPTION
Bowen, Jennifer	Northeastern University	Combining High Resolution Organic Matter Characterization and Microbial Meta-Omics to Assess the Effects of Nutrient Loading on Salt Marsh Carbon Sequestration
Mayes, Melanie	Oak Ridge National Laboratory	Linking Proteogenomics, Metabolomics, and Soil Organic Chemistry of Tropical Wetlands to a Soil Nutrient Cycling Model
Onstott, Tullis	Princeton University	Detecting Seismically-Sustained Deep Subsurface CH₄-Cycling Chemolithoautotrophic Microbial Communities Using Multi-Omic Analyses and NanoSIMS
Saleska, Scott	University of Arizona	Investigating the Carbon Cycling Implications of Changing Microbial Leaf Litter Decomposition across a Permafrost Thaw Gradient
Talbot, Jennifer	Boston University	Scaling Molecular Mechanisms of Mycorrhizal- Decomposer Interactions to Emergent Ecosystem Carbon Balance
Tiemann, Lisa	Michigan State University	Tracking Switchgrass Photosynthate via 13CO ₂ Pulse-Chase into the Rhizosphere Microbiome and Metabolome

Appendix D

Advisory and Review Committee Members

JGI Scientific Advisory Committee

The Scientific User Advisory Committee is a board convened by the JGI Director to provide a scientific and technical overview of the JGI. The board's responsibilities 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 the DOE Office of Biological and Environmental Research (BER) concurrence, set the final sequence allocation for this program.

Mark Adams, The Jackson Laboratory

Gloria Coruzzi, New York University

Jeff Dangl, University of North Carolina

Paul Flicek, EMBL-EBI

Brian Fox, GLBRC

Jeff Gilbert, Corteva AgriScience

Trina McMahon, University of Wisconsin-Madison

Sean McSweeney, Brookhaven National Laboratory

Debbie Yaver, Novozymes

Informatics Advisory Committee

Toby Bloom, New York Genome Center

Amber Boehnlein, Jefferson Lab

Austin Che, Ginkgo Bioworks

David Dooling, Atomist

Paul Flicek, European Bioinformatics Institute (EBI)

Dan Jacobson, Oak Ridge National Laboratory (ORNL)

Nadim Jessani, Genedata AG

David Lewis, Illumina

Anne Deslattes Mays, The Jackson Laboratory for Genomic Medicine

Katherine Riley, Argonne National Laboratory (ANL)

Fungal Program User Advisory Committee

Scott Baker, Pacific Northwest National Laboratory

Randy Berka, Archer Daniels Midland

Ronald de Vries, Westerdijk Fungal Biodiversity Institute

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

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

Stephen Goodwin, Purdue University

David Hibbett, Clark University

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

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

Joseph Spatafora, Oregon State University

Kathleen Treseder, University of California, Irvine

Adrian Tsang, Concordia University

Prokaryotic Super Program Advisory Committee

Jill Banfield, University of California, Berkeley

Ed DeLong, University of Hawaii at Manoa

Jonathan Eisen, University of California, Davis

George Garrity, Michigan State University

Steve Hallam, University of British Columbia

Phil Hugenholtz, University of Queensland

Janet Jansson, Pacific Northwest National Laboratory

Kostas Konstantinidis, Georgia Tech

Trina McMahon, University of Wisconsin-Madison

Monica Medina, Penn State University

Mary Ann Moran, University of Georgia

Nancy Moran, University of Texas at Austin

Jennifer Pett-Ridge, *La*wrence Livermore National Laboratory

,

Rich Roberts, New England Biolabs

Lynn Schriml, *University of Maryland*

Ramunas Stepanauskas, Bigelow Laboratory for Ocean Sciences

Matt Sullivan, The Ohio State University

Plant Program User Advisory Committee

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, Danforth Center

Sabeeha Merchant, University of California, Los Angeles

Stephen Moose, University of Illinois

Sue Rhee, Carnegie Institution for Science

Bob Schmitz, University of Georgia

Gary Stacey, University of Missouri, DOE Biological and Environmental Research Advisory Committee

DNA Synthesis Science User Advisory Committee

Richard Bailey, independent consultant

Doug Cameron, Firstgreen Partners

Sunil Chandran, Amyris Inc.

James Flatt, Synthetic Genomics

Jay Keasling, Lawrence Berkeley National Laboratory

Megan Palmer, Center for International Security and Cooperation (CISAC), Stanford University

Elizabeth Sattely, Stanford University

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

David Weller, USDA-ARS

Diversity & Inclusion Advisory Board

Cat Adams, UC Berkeley

Debbie Brockett, Elemental Excelerator

Raul Cano, The BioCollective

Jonathan Eisen, UC Davis

Barbara Gee, Management Consultant

Nan Ho, Dean of Math, Science, Engineering & Public Safety, Las Positas College

Jim Hollis, Calculus Roundtable

Koni Patterson, Altria Client Services

Matt Perry, Kapor Center for Social Justice

Geri Richmond, University of Oregon

Neha Sampat, GenLead BelongLab

Appendix E

2018 Genomics of Energy and Environment Meeting

With more than 600 attendees, the 13th Annual JGI Genomics of Energy and Environment Meeting took place March 13–16, 2018.

Keynote Speakers



"We're interested in breadth of genomic and ecological diversity, niche differentiation and metabolic drivers of AOM syntrophy,"

— Victoria Orphan

The opening keynote from MacArthur Fellow Victoria Orphan of Caltech provided an overview of the ecosystems biology approach her group is taking to link uncultivated microorganisms to biogeochemical processes. Orphan described several approaches her lab is taking to better understand the

relationship between the global methane budget and anoxic methanogens (AOM), organisms in seep sediments that make a living oxidizing methane coupled to sulfur reduction. "We're interested in breadth of genomic and ecological diversity, niche differentiation and metabolic drivers of AOM syntrophy," she told the crowd.

Watch her talk at bit.ly/JGI2018Orphan.



Nigel Mouncey, Kristala Prather, Susannah Tringe

For keynote speaker **Kristala Prather**, a professor of chemical engineering at Massachusetts Institute of Technology (MIT), the challenge of synthetic biology is navigating that complex network to maximize productivity while minimizing unwanted byproducts and unnecessary growth. The flagship project in Prather's lab has been engineering microorganisms like E. coli and yeast to synthesize glucaric acid, considered by the DOE to be a top "value added chemical from biomass."

Watch her talk at bit.ly/JGI2018Prather.

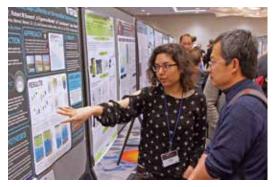


"There's nothing wrong with publishing papers, per se, however we also need to find ways to reward scientists who score higher on quality, who share more openly, and whose work is more reproducible and has more translational impact when it is applied."

— John Ioannidis

John loannidis of the Stanford School of Medicine and co-director of the Meta-Research Innovation Center at Stanford (METRICS) delivered the closing keynote address on the issue of reproducibility — or, rather, the lack thereof—that plagues the scientific research enterprise. In recent years, there's been growing concern about this topic, reflected in an increasing number of conversations and publications across all 22 major scientific disciplines. There's no real consensus about what "reproducible research" means, but there are at least three broad aspects. Ultimately, said loannidis, the goal is the same: to enhance trust in scientific findings.

Watch his talk at bit.ly/JGI2018Ioannidis.



JGI's Zhong Wang listens to UC Merced graduate student Sabah UI-Hasan, who had a 2015 JGI/UC Merced fellowship.



Han Wosten revisited his JGI collaboration to sequence the fungus *Schizophyllum commune*.

Videos from the Meeting and the VEGA Symposium are available on the JGI's YouTube channel at http://bit.ly/JGI2018videos.



Merce Montoliu Nerin talked about fungi.



Algal researcher Susan Brawley highlighted efforts of her JGI collaborators.

Other Featured Speakers (in order of appearance):

Tim James, University of Michigan

Mercè Montoliu Nerín, Uppsala University (Sweden)

Evan DeLucia, University of Illinois/CABBI

John Cushman, University of Nevada

Jeremy Schmutz, JGI

Susan Brawley, University of Maine

Han Wösten, Utrecht University (Netherlands)

Julia Anstett, University of British Columbia (Canada)

Jizhong Zhou, University of Oklahoma

Cynthia Collins, Rensselaer Polytechnic Institute

Maureen Coleman, University of Chicago

Lynn Schriml, University of Maryland

Kriti Sharma, University of North Carolina at Chapel Hill

Krista McGuire, University of Oregon

James Evans, EMSL

Ty Samo, Lawrence Livermore National Laboratory
Ru Zhang, Danforth Center
Andrzej Joachimiak, Argonne National Laboratory
Jennifer Pett-Ridge, Lawrence Livermore National Laboratory
Sheena Becker, DowDuPont
Mike Jewett, Northwestern University
Daniela Pignatta, Indigo Agriculture
lan Wheeldon, University of California, Riverside
Amy Banta, Stanford University
Jessica Vera, GLBRC, University of Wisconsin-Madison
Manuel Kleiner, North Carolina State University

Tom Jeffries, *Xylome*Aymeric Leveau, *John Innes Centre (UK)*

The first ever Viral EcoGenomics and Applications (VEGA) Symposium was held during the JGI's 13th Annual Genomics of Energy & Environment Meeting.







Top: Eugene Koonin; Bottom (left to right): Marilyn Roossinck, Joe DeRisi.



Eric Delwart of Blood Systems Research Institute (BSRI) poses a question considered by (left to right): Bonnie Hurwitz, University of Arizona; Xutao Deng, BSRI; VEGA co-organizer Phil Hugenholtz, Director of the Australian Centre for Ecogenomics at the University of Queensland; and, Nikos Kyrpides, JGI Prokaryote Super Program Head and lead organizer of the VEGA Symposium.

Keynote Speakers

The symposium opened with a keynote delivered by NCBI bioinformatician Eugene Koonin, who immediately declared viruses are the most abundant organisms, and by extension, the most important.

Watch his talk at bit.ly/JGI2018KooninVEGA.

Marilyn Roossinck of Penn State University delivered a keynote addressing plant viruses, reminding everyone that viruses are not always pathogens.

Watch her talk at bit.ly/JGI2018RoossinckVEGA.

Joe DeRisi, co-director of the Chan Zuckerberg Hub and also a professor at the University of California, San Francisco, used his keynote address to discuss the applications of identifying viruses.

Watch his talk at bit.ly/JGI2018DeRisiVEGA.

Other Featured Speakers (in order of appearance):

Lisa Zeigler-Allen, J. Craig Venter Institute

Wen-Tso Liu, University of Illinois

Matt Sullivan, Ohio State University

Rene Kallies, Helmholtz Centre for Environmental Research

David Paez-Espino, JGI

Rodney Brister, NCBI

Siobain Duffy, Rutgers University

Joanne Emerson, UC Davis

Vivek Mutalik, LBNL

Evelien Adriaenssens, University of Liverpool

Britt Koskella, UC Berkeley

Simon Roux, JGI

David Paez, JGI

David Mead, Varigen Biosciences

Sylvain Moineau, Laval University (Canada)

Philip Hugenholtz, The University of Queensland (Australia)

Appendix F

2018 Publications

Ai C et al. Evolution of copper arsenate resistance for enhanced enargite bioleaching using the extreme thermoacidophile Metallosphaera sedula. *J Ind Microbiol Biotechnol.* 2017 Dec;44(12):1613-1625. doi: 10.1007/s10295-017-1973-5.

Aime MC et al. Phylogenetics and Phylogenomics of Rust Fungi. *Adv Genet*. 2017;100:267-307. doi: 10.1016/bs.adgen.2017.09.011. Epub 2017 Oct 26.

Amos GCA et al. Comparative transcriptomics as a guide to natural product discovery and biosynthetic gene cluster functionality. *Proc Natl Acad Sci USA*. 2017 Dec 26;114(52):E11121-E11130. doi: 10.1073/pnas.1714381115

Anantharaman K et al. Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. *ISME J.* 2018 Jun;12(7):1715-1728. doi: 10.1038/s41396-018-0078-0. Epub 2018 Feb 21.

Angle JC et al. Methanogenesis in oxygenated soils is a substantial fraction of wetland methane emissions. *Nat Commun.* 2017 Nov 16;8(1):1567. doi: 10.1038/s41467-017-01753-4.

Aserse AA et al. Draft genome sequences of Bradyrhizobium shewense sp. nov. ERR11T and Bradyrhizobium yuanmingense CCBAU 10071T. *Stand Genomic Sci.* 2017 Dec 5;12:74. doi: 10.1186/s40793-017-0283-x. eCollection 2017.

Azad A et al. HipMCL: a high-performance parallel implementation of the Markov clustering algorithm for large-scale networks. *Nucleic Acids Res.* 2018 Apr 6;46(6):e33. doi: 10.1093/nar/gkx1313.

Bachy C et al. Transcriptional responses of the marine green alga Micromonas pusilla and an infecting prasinovirus under different phosphate conditions. *Environ Microbiol.* 2018 Aug;20(8):2898-2912. doi: 10.1111/1462-2920.14273. Epub 2018 Sep 9.

Bahri BA et al. Natural variation in genes potentially involved in plant architecture and adaptation in switchgrass (Panicum virgatum L.). *BMC Evol Biol.* 2018 Jun 14;18(1):91. doi: 10.1186/s12862-018-1193-2.

Bakermans C. Adaptations to marine versus terrestrial low temperature environments as revealed by comparative genomic analyses of the genus Psychrobacter. *FEMS Microbiol Ecol.* 2018 Jul 1;94(7). doi: 10.1093/femsec/fiy102.

Balasundaram SV et al. The fungus that came in from the cold: dry rot's pre-adapted ability to invade buildings. *ISME J.* 2018 Mar;12(3):791-801. doi: 10.1038/s41396-017-0006-8. Epub 2018 Jan 5.

Becraft ED et al. Rokubacteria: Genomic Giants among the Uncultured Bacterial Phyla. *Front Microbiol.* 2017 Nov 28;8:2264. doi: 10.3389/fmicb.2017.02264. eCollection 2017.

Becraft ED et al. Genomic Comparison of Two Family-Level Groups of the Uncultivated NAG1 Archaeal Lineage from Chemically and Geographically Disparate Hot Springs. *Front Microbiol*. 2017 Oct 31;8:2082. doi: 10.3389/fmicb.2017.02082. eCollection 2017

Bell E et al. Biogeochemical Cycling by a Low-Diversity Microbial Community in Deep Groundwater. *Front Microbiol.* 2018 Sep 7;9:2129. doi: 10.3389/fmicb.2018.02129. eCollection 2018.

Beller HR et al. Discovery of enzymes for toluene synthesis from anoxic microbial communities. *Nat Chem Biol.* 2018 May;14(5):451-457. doi: 10.1038/s41589-018-0017-4. Epub 2018 Mar 19.

Benocci T et al. Enzymatic Adaptation of Podospora anserina to Different Plant Biomass Provides Leads to Optimized Commercial Enzyme Cocktails. *Biotechnol J.* 2018 Sep 17:e1800185. doi: 10.1002/biot.201800185. [Epub ahead of print]

Bhatnagar JM et al. Litter chemistry influences decomposition through activity of specific microbial functional guilds. *Ecol Monogr.* 2018 Aug;88(3):429-444. doi:10.1002/ecm.1303

Biswas R et al. Complete Genome Sequence of Thermoanaerobacterium sp. Strain RBIITD, a Butyrate-and Butanol-Producing Thermophile. *Genome Announc*. 2018 Jan 11;6(2). pii: e01411-17. doi: 10.1128/genomeA.01411-17.

Borton MA et al. Coupled laboratory and field investigations resolve microbial interactions that underpin persistence in hydraulically fractured shales. Proc Natl Acad Sci U S A. 2018 Jul 10;115(28):E6585-E6594. doi: 10.1073/ pnas.1800155115. Epub 2018 Jun 25.

Bowers RM et al. Sequencing of Genomes from Environmental Single Cells. Methods Mol Biol. 2018;1712:97-111. doi: 10.1007/978-1-4939-7514-3_8.

Bowman JL et al. Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. Cell. 2017 Oct 5;171(2):287-304.e15. doi: 10.1016/j. cell.2017.09.030.

Bryan AC et al. A Variable Polyglutamine Repeat Affects Subcellular Localization and Regulatory Activity of a Populus ANGUSTIFOLIA Protein. G3 (Bethesda). 2018 Jul 31;8(8):2631-2641. doi: 10.1534/g3.118.200188.

Bushnell B et al. BBMerge - Accurate paired shotgun read merging via overlap. PLoS One. 2017 Oct 26;12(10):e0185056. doi: 10.1371/journal. pone.0185056. eCollection 2017.

Butler JB et al. Annotation of the Corymbia terpene synthase gene family shows broad conservation but dynamic evolution of physical clusters relative to Eucalyptus. Heredity (Edinb). 2018 Jul;121(1):87-104. doi: 10.1038/s41437-018-0058-1. Epub 2018 Mar 10.

Camiolo S et al. An analysis of codon bias in six red yeast species. Yeast. 2019 Jan;36(1):53-64. doi: 10.1002/ yea.3359. Epub 2018 Oct 16.

Cardenas E et al. Effects of timber harvesting on the genetic potential for carbon and nitrogen cycling in five North American forest ecozones. Sci Rep. 2018 Feb 16;8:3142. doi:10.1038/s41598-018-21197-0

Carella P et al. Phytophthora palmivora establishes tissue-specific intracellular infection structures in the earliest divergent land plant lineage. Proc Natl Acad Sci USA. 2018 Apr 3. pii: 201717900. doi: 10.1073/ pnas.1717900115. [Epub ahead of print]

Carro L et al. Genome-based classification of micromonosporae with a focus on their biotechnological and ecological potential. Sci Rep. 2018 Jan 11;8(1):525. doi: 10.1038/s41598-017-17392-0.

Carella P et al. Phytophthora palmivora establishes tissue-specific intracellular infection structures in the earliest divergent land plant lineage. Proc Natl Acad Sci USA. 2018 Apr 3. pii: 201717900. doi: 10.1073/ pnas.1717900115. [Epub ahead of print]

Castanera R et al. Comparative genomics of Coniophora olivacea reveals different patterns of genome expansion in Boletales. BMC Genomics. 2017 Nov 16;18(1):883. doi: 10.1186/s12864-017-4243-z.

Ceballos SJ et al. Development and characterization of a thermophilic, lignin degrading microbiota. Process Biochem. 2017 Dec;63:193-203. doi:10.1016/j. procbio.2017.08.018

Chadha S et al. Genome-wide analysis of cytochrome P450s of Trichoderma spp.: annotation and evolutionary relationships. Fungal Biol Biotechnol. 2018 Jun 4;5:12. doi: 10.1186/s40694-018-0056-3. eCollection 2018.

Chafee M et al. Recurrent patterns of microdiversity in a temperate coastal marine environment. ISME J. 2018 Jan;12(1):237-252. doi: 10.1038/ismej.2017.165. Epub 2017 Oct 24.

Chang JJ et al. Biomimetic strategy for constructing Clostridium thermocellum cellulosomal operons in Bacillus subtilis. Biotechnol Biofuels. 2018 Jun 7;11:157. doi:10.1186/s13068-018-1151-7

Chen ECH et al. High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont Rhizophagus irregularis. New Phytol. 2018 Dec;220(4):1161-1171. doi: 10.1111/nph.14989. Epub 2018 Jan 22.

Chen LX et al. Metabolic versatility of small archaea Micrarchaeota and Parvarchaeota. ISME J. 2017 Dec 8. doi: 10.1038/s41396-017-0002-z.

Chernikova TN et al. Draft Genome Seguence of Monaibacterium marinum C7T, Isolated from Seawater from the Menai Straits, Wales, United Kingdom. Genome Announc. 2018 Feb 1;6(5). pii: e01444-17. doi: 10.1128/genomeA.01444-17.

Chiang E et al. Verrucomicrobia are prevalent in north-temperate freshwater lakes and display class-level preferences between lake habitats. PLoS One. 2018 Mar 28;13(3):e0195112. doi: 10.1371/journal. pone.0195112. eCollection 2018.

Choi DH et al. Draft genome sequence of Marinobacterium rhizophilum CL-YJ9T (DSM 18822T), isolated from the rhizosphere of the coastal tidal-flat plant Suaeda japonica. *Stand Genomic Sci.* 2017 Oct 30;12:65. doi: 10.1186/s40793-017-0275-x. eCollection 2017.

Coates RC et al. An integrated workflow for phenazine-modifying enzyme characterization. *J Ind Microbiol Biotechnol.* 2018 Jul;45(7):567-577. doi: 10.1007/s10295-018-2025-5. Epub 2018 Mar 15.

Cobo-Diaz, JF et al. Taxonomic and functional diversity of a Quercus pyrenaica Willd. rhizospheric microbiome in the Mediterranean mountains. *Forests*. 2017 Oct; 8(1): 390. doi: 10.3390/f8100390

Coninx L et al. The SIZRT1 Gene Encodes a Plasma Membrane-Located ZIP (Zrt-, Irt-Like Protein) Transporter in the Ectomycorrhizal Fungus Suillus luteus. *Front Microbiol.* 2017 Nov 28;8:2320. doi: 10.3389/fmicb.2017.02320. eCollection 2017.

Coradetti ST et al. Functional genomics of lipid metabolism in the oleaginous yeast Rhodosporidium toruloides. *Elife.* 2018 Mar 9;7. pii: e32110. doi: 10.7554/eLife.32110.

Cox RS et al. Synthetic Biology Open Language (SBOL) Version 2.2.0. *J Integr Bioinform*. 2018 Apr 2;15(1). pii: /j/jib.2018.15.issue-1/jib-2018-0001/jib-2018-0001.xml. doi: 10.1515/jib-2018-0001.

Crits-Cristoph A et al. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature*. 2018 Jun;558(7710):440-444. doi: 10.1038/s41586-018-0207-y. Epub 2018 Jun 13.

D'haeseleer P et al. Metagenomic analysis of intertidal hypersaline microbial mats from Elkhorn Slough, California, grown with and without molybdate. *Stand Genomic Sci.* 2017 Nov 15;12:67. doi: 10.1186/s40793-017-0279-6. eCollection 2017.

Dalcin Martins P et al. Viral and metabolic controls on high rates of microbial sulfur and carbon cycling in wetland ecosystems. *Microbiome*. 2018 Aug 7;6(1):138. doi: 10.1186/s40168-018-0522-4.

Daly P et al. Dichomitus squalens partially tailors its molecular responses to the composition of solid wood. *Env Microbiol.* 24 Sep 2018. doi: 10.1111/1462-2920.14416

de Freitas Pereira M et al. Secretome Analysis from the Ectomycorrhizal Ascomycete Cenococcum geophilum. *Front Microbiol.* 2018 Feb 13;9:141. doi: 10.3389/fmicb.2018.00141. eCollection 2018.

de Matos GF et al. Bradyrhizobium sacchari sp. nov., a legume nodulating bacterium isolated from sugarcane roots. *Arch Microbiol.* 2017 Jun 10;199:1251. doi:10.1007/s00203-017-1398-6

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.

Díaz-Cárdenas C et al. Draft genome sequence of Dethiosulfovibrio salsuginis DSM 21565T an anaerobic, slightly halophilic bacterium isolated from a Colombian saline spring. *Stand Genomic Sci.* 2017 Dec 20;12:86. doi: 10.1186/s40793-017-0303-x. eCollection 2017.

Ding Y et al. Selinene Volatiles Are Essential Precursors for Maize Defense Promoting Fungal Pathogen Resistance. *Plant Physiol.* 2017 Nov;175(3):1455-1468. doi: 10.1104/pp.17.00879. Epub 2017 Sep 20.

Dippel AB et al. Chemiluminescent Biosensors for Detection of Second Messenger Cyclic di-GMP. ACS Chem Biol. 2018 Jul 20;13(7):1872-1879. doi: 10.1021/acschembio.7b01019. Epub 2018 Mar 5.

Dossani ZY et al. A combinatorial approach to synthetic transcription factor-promoter combinations for yeast strain engineering. *Yeast*. 2017 Oct 30. doi: 10.1002/yea.3292.

Druzhinina IS et al. Massive lateral transfer of genes encoding plant cell wall-degrading enzymes to the mycoparasitic fungus Trichoderma from its plant-associated hosts. *PLoS Genet*. 2018 Apr 9;14(4):e1007322. doi: 10.1371/journal.pgen.1007322. eCollection 2018 Apr.

Duerkop BA et al. Murine colitis reveals a disease-associated bacteriophage community. *Nat Microbiol.* 2018 Sep;3(9):1023-1031. doi: 10.1038/s41564-018-0210-y. Epub 2018 Jul 23.

Dunivin TK et al. Community structure explains antibiotic resistance gene dynamics over a temperature gradient in soil. *FEMS Microbiol Ecol.* 2018 Mar 1;94(3). doi: 10.1093/femsec/fiy016.

Dyksma S et al. Evidence for H2 consumption by uncultured Desulfobacterales in coastal sediments. Environ Microbiol. 2017 Aug 03. doi: 10.1111/1462-2920.13880

Eichorst SA et al. Genomic insights into the Acidobacteria reveal strategies for their success in terrestrial environments. Environ Microbiol. 2018 J an 12. doi: 10.1111/1462-2920.14043. [Epub ahead

Emerson JB et al. Host-linked soil viral ecology along a permafrost thaw gradient. Nat Microbiol. 2018 Aug;3(8):870-880. doi: 10.1038/s41564-018-0190-y. Epub 2018 Jul 16.

Emery M et al. Preferential retention of genes from one parental genome after polyploidy illustrates the nature and scope of the genomic conflicts induced by hybridization. PLoS Genet. 2018 Mar 28;14(3):e1007267. doi: 10.1371/journal.pgen.1007267. [Epub ahead of print]

Eminoğlu A et al. Deletion of the hfsB gene increases ethanol production in Thermoanaerobacterium saccharolyticum and several other thermophilic anaerobic bacteria. Biotechnol Biofuels. 2017 Nov 30;10:282. doi: 10.1186/s13068-017-0968-9

Espínola F et al. Metagenomic Analysis of Subtidal Sediments from Polar and Subpolar Coastal Environments Highlights the Relevance of Anaerobic Hydrocarbon Degradation Processes. Microb Ecol. 2018 Jan;75(1):123-139. doi: 10.1007/s00248-017-1028-5.

Estrada-de Los Santos P et al. Whole Genome Analyses Suggests that Burkholderia sensu lato Contains Two Additional Novel Genera (Mycetohabitans gen. nov., and Trinickia gen. nov.): Implications for the Evolution of Diazotrophy and Nodulation in the Burkholderiaceae. Genes (Basel). 2018 Aug 1;9(8). pii: E389. doi: 10.3390/genes9080389.

Evans J et al. Extensive Genetic Diversity is Present within North American Switchgrass Germplasm. Plant Genome. 2018 Mar;11(1). doi: 10.3835/ plantgenome2017.06.0055.

Fernández-Bayo JD et al. Comparison of soil biosolarization with mesophilic and thermophilic solid digestates on soil microbial quantity and diversity. Appl Soil Ecol. 2017 Oct;119,183-191. doi:10.1016/j. apsoil.2017.06.016

Fischer MS et al. Regulation of Cell-to-Cell Communication and Cell Wall Integrity by a Network of MAP Kinase Pathways and Transcription Factors in Neurospora crassa. Genetics. 2018 Jun;209(2):489-506. doi: 10.1534/genetics.118.300904.

Fleming EJ et al. Insights into the fundamental physiology of the uncultured Fe-oxidizing bacterium Leptothrix ochracea. Appl Environ Microbiol. 2018 Feb 16. pii: AEM.02239-17. doi: 10.1128/AEM.02239-17. [Epub ahead of print]

Gach PC et al. Droplet microfluidics for synthetic biology. Lab Chip. 2017 Aug 10;17:3388-3400. doi:10.1039/C7LC00576H

Ganesh S et al. Single cell genomic and transcriptomic evidence for the use of alternative nitrogen substrates by anammox bacteria. ISME J. 2018 Jul 10. doi: 10.1038/ s41396-018-0223-9. [Epub ahead of print]

Gao J et al. Ecosystem Fabrication (EcoFAB) Protocols for The Construction of Laboratory Ecosystems Designed to Study Plant-microbe Interactions. J Vis Exp. 2018 Apr 10;(134). doi: 10.3791/57170.

Garcia SL et al. Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. ISME J. 2017 Dec 8. doi: 10.1038/ s41396-017-0001-0. [Epub ahead of print]

Garcia SL et al. Model Communities Hint at Promiscuous Metabolic Linkages between Ubiquitous Free-Living Freshwater Bacteria. mSphere. 2018 May 30;3(3). pii: e00202-18. doi: 10.1128/ mSphere.00202-18. Print 2018 Jun 27.

Garsmeur O et al. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. Nat Commun. 2018 Jul 6;9(1):2638. doi: 10.1038/s41467-018-05051-5.

Gasch AP et al. Single-cell RNA sequencing reveals intrinsic and extrinsic regulatory heterogeneity in yeast responding to stress. PLoS Biol. 2017 Dec 14;15(12):e2004050. doi: 10.1371/journal. pbio.2004050. eCollection 2017 Dec.

Gaskell J et al. Draft genome sequence of a monokaryotic model brown-rot fungus Postia (Rhodonia) placenta SB12. Genom Data. 2017 Aug 10;14:21-23. doi: 10.1016/j.gdata.2017.08.003. eCollection 2017 Dec.

Gilbert RA et al. Toward Understanding Phage:Host Interactions in the Rumen; Complete Genome Sequences of Lytic Phages Infecting Rumen Bacteria. *Front Microbiol.* 2017 Dec 5;8:2340. doi: 10.3389/fmicb.2017.02340. eCollection 2017.

Gluck-Thaler E et al. Specialized plant biochemistry drives gene clustering in fungi. *ISME J.* 2018 Jun;12(7):1694-1705. doi: 10.1038/s41396-018-0075-3.

Gordon SP et al. Extensive gene content variation in the Brachypodium distachyon pan-genome correlates with population structure. *Nat Commun.* 2017 Dec 19;8(1):2184. doi: 10.1038/s41467-017-02292-8.

Gryganskyi AP et al. Phylogenetic and Phylogenomic Definition of Rhizopus Species. *G3* (Bethesda). 2018 May 31;8(6):2007-2018. doi: 10.1534/g3.118.200235.

Guo J et al. Specialized proteomic responses and an ancient photoprotection mechanism sustain marine green algal growth during phosphate limitation. *Nat Microbiol.* 2018 Jul;3(7):781-790. doi: 10.1038/s41564-018-0178-7. Epub 2018 Jun 25.

Hahn MW et al. Reclassification of a Polynucleobacter cosmopolitanus strain isolated from tropical Lake Victoria as Polynucleobacter victoriensis sp. nov. *Int J Syst Evol Microbiol*. 2017 Oct 25. doi: 10.1099/ijsem.0.002421. [Epub ahead of print]

Handakumbura PP et al. SECONDARY WALL ASSOCIATED MYB1 is a positive regulator of secondary cell wall thickening in Brachypodium distachyon and is not found in the Brassicaceae. *Plant J.* 2018 Jul 27. doi: 10.1111/tpj.14047. [Epub ahead of print]

Harrington LB et al. A thermostable Cas9 with increased lifetime in human plasma. *Nat Commun*. 2017 Nov 10;8(1):1424. doi: 10.1038/s41467-017-01408-4.

Hausmann B et al. Peatland Acidobacteria with a dissimilatory sulfur metabolism. *ISME J.* 2018 Feb 23. doi: 10.1038/s41396-018-0077-1. [Epub ahead of print]

Hawley AK et al. Diverse Marinimicrobia bacteria may mediate coupled biogeochemical cycles along eco-thermodynamic gradients. *Nat Commun.* 2017 Nov 15;8(1):1507. doi: 10.1038/s41467-017-01376-9.

Hawley AK et al. A compendium of multi-omic sequence information from the Saanich Inlet water column. *Sci Data*. 2017 Oct 31;4:170160. doi: 10.1038/sdata.2017.160.

Herbold CW et al. Ammonia oxidising archaea living at low pH: Insights from comparative genomics. *Environ Microbiol.* 2017 Nov 02. doi:10.1111/1462-2920.13971

Hesse C et al. Genome-based evolutionary history of Pseudomonas spp. *Environ Microbiol.* 2018 Apr 6. doi: 10.1111/1462-2920.14130. [Epub ahead of print]

Hon S et al. Expressing the Thermoanaerobacterium saccharolyticum pfor A in engineered Clostridium thermocellum improves ethanol production. *Biotechnol Biofuels*. 2018 Sep 6;11:242. doi: 10.1186/s13068-018-1245-2

Hori C et al. Multi-omic analyses of extensively decayed Pinus contorta reveal expression of diverse array of lignocellulose degrading enzymes. *Appl Environ Microbiol.* 2018 Aug 10. pii: AEM.01133-18. doi: 10.1128/AEM.01133-18. [Epub ahead of print]

Huang YT et al. Genome-based analysis of virulence determinants of a Serratia marcescens strain from soft tissues following a snake bite. *Future Microbiol.* 2017 Nov 6. doi: 10.2217/fmb-2017-0202. [Epub ahead of print]

Huang YT et al. Draft genome sequence of carbapenem-resistant Shewanella algae strain AC isolated from small abalone (Haliotis diversicolor). J *Glob Antimicrob Resist*. 2018 Sep;14:65-67. doi: 10.1016/j.jgar.2018.06.005.

Jansson C et al. Climate-smart crops with enhanced photosynthesis. *J Exp Bot*. 2018 Jul 18;69(16):3801-3809. doi: 10.1093/jxb/ery213.

Jarett JK et al. Single-cell genomics of co-sorted Nanoarchaeota suggests novel putative host associations and diversification of proteins involved in symbiosis. *Microbiome*. 2018 Sep 17;6(1):161. doi: 10.1186/s40168-018-0539-8.

Jay ZJ et al. Marsarchaeota are an aerobic archaeal lineage abundant in geothermal iron oxide microbial mats. *Nat Microbiol.* 2018 Jun;3(6):732-740. doi: 10.1038/s41564-018-0163-1. Epub 2018 May 14.

Jayakody LN et al. Thermochemical wastewater valorization via enhanced microbial toxicity tolerance. *Energy Environ. Sci.* 2018 Apr 17;11:1625-1638. doi: 10.1039/C8EE00460A

Kappelmann L et al. Polysaccharide utilization loci of North Sea Flavobacteriia as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. ISME J. 2018 Aug 15. doi: 10.1038/s41396-018-0242-6. [Epub ahead of print]

Khdhiri M et al. Meta-omics survey of [NiFe]hydrogenase genes fails to capture drastic variations in H2-oxidation activity measured in three soils exposed to H2. Soil Biology and Biochemistry. 2018 Oct; 125: 239-243. doi: 10.1016/j.soilbio.2018.07.020.

Kijpornyongpan T et al. Broad genomic sampling reveals a smut pathogenic ancestry of the fungal clade Ustilaginomycotina. Mol Biol Evol. 2018 Aug 1;35(8):1840-1854. doi: 10.1093/molbev/msy072. Epub 2018 May 15.

Kimbrel JA et al. Microbial Community Structure and Functional Potential Along a Hypersaline Gradient. Front. Microbiol. 10 July 2018 doi:10.3389/ fmicb.2018.01492

Kjærbølling I et al. Linking secondary metabolites to gene clusters through genome sequencing of six diverse Aspergillus species. Proc Natl Acad Sci U S A. 2018 Jan 9. pii: 201715954. doi: 10.1073/ pnas.1715954115. [Epub ahead of print]

Kleiner M et al. Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. Proc Natl Acad Sci USA. 2018 Jun 12;115 (24):E5576-E5584. doi:10.1073/ pnas.1722325115

Kliver S et al. Assembly of the Boechera retrofracta Genome and Evolutionary Analysis of Apomixis-Associated Genes. Genes (Basel). 2018 Mar 28;9(4). pii: E185. doi: 10.3390/genes9040185.

Knapp DG et al. Comparative genomics provides insights into the lifestyle and reveals functional heterogeneity of dark septate endophytic fungi. Sci Rep. 2018 Apr 20;8(1):6321. doi: 10.1038/s41598-018-24686-4.

Kolinko S et al. A bacterial pioneer produces cellulase complexes that persist through community succession. Nat Microbiol. 2018 Jan;3(1):99-107. doi: 10.1038/ s41564-017-0052-z. Epub 2017 Nov 6.

Krassowski T et al. Evolutionary instability of CUG-Leu in the genetic code of budding yeasts. Nat Commun. 2018 May 14;9(1):1887. doi: 10.1038/s41467-018-04374-7.

Kroeger ME et al. New Biological Insights Into How Deforestation in Amazonia Affects Soil Microbial Communities Using Metagenomics and Metagenome-Assembled Genomes. Frontiers in Microbiology. 2018 Jul; 9: 1635. doi: 10.3389/fmicb.2018.01635

Kruse T et al. Comparative genomics of the genus Desulfitobacterium. FEMS Microbiol Ecol. 2017 Dec 1:93(12). doi: 10.1093/femsec/fix135.

Kuang MC et al. Repeated cis-regulatory tuning of a metabolic bottleneck gene during evolution. Mol Biol Evol. 2018 Aug 1;35(8):1968-1981. doi: 10.1093/ molbev/msy102. Epub 2018 May 21.

Lang D et al. The P. patens chromosome-scale assembly reveals moss genome structure and evolution. Plant J. 2017 Dec 13. doi: 10.1111/tpj.13801. [Epub ahead of print]

Lee JZ et al. Metagenomics reveals niche partitioning within the phototrophic zone of a microbial mat. PLoS One. 2018 Sep 11;13(9):e0202792. doi: 10.1371/journal. pone.0202792. eCollection 2018.

Lee LL et al. Genus-Wide Assessment of Lignocellulose Utilization in the Extremely Thermophilic Genus Caldicellulosiruptor by Genomic, Pangenomic, and Metagenomic Analyses. Appl Environ Microbiol. 2018 Apr 16;84(9). pii: e02694-17. doi: 10.1128/AEM.02694-17. Print 2018 May 1.

Levy A et al. Genomic features of bacterial adaptation to plants. Nat Genet. 2018 Jan;50(1):138-150. doi: 10.1038/s41588-017-0012-9. Epub 2017 Dec 18.

Lewin HA et al. Earth BioGenome Project: Sequencing life for the future of life. Proc Natl Acad Sci U S A. 2018 Apr 24;115(17):4325-4333. doi: 10.1073/ pnas.1720115115.

Li Y et al. Proteome and phospholipid alteration reveal metabolic network of Bacillus thuringiensis under triclosan stress. Sci Total Environ. 2018 Feb 15;615:508-516. doi: 10.1016/j.scitotenv.2017.10.004. Epub 2017 Oct 5.

Lin H et al. Combining Hadoop with MPI to Solve Metagenomics Problems that are both Data- and Compute-intensive. Int J Parallel Prog. 2018 Aug;46: 762. doi: 10.1007/s10766-017-0524-

Liu J et al. Novel reductive dehalogenases from the marine sponge associated bacterium Desulfoluna spongiiphila. *Environ Microbiol*. 2017 Jun 15. doi: 10.1111/1758-2229.12556

Looney BP et al. Russulaceae: a new genomic dataset to study ecosystem function and evolutionary diversification of ectomycorrhizal fungi with their tree associates. *New Phytol.* 2018 Apr;218(1):54-65. doi: 10.1111/nph.15001. Epub 2018 Jan 30.

Lopez D et al. Genome-Wide Analysis of Corynespora cassiicola Leaf Fall Disease Putative Effectors. Front. *Microbiol.* 02 March 2018. doi:10.3389/fmicb.2018.00276

López G et al. Draft genome sequence of Pseudomonas extremaustralis strain USBA-GBX 515 isolated from Superparamo soil samples in Colombian Andes. *Stand Genomic Sci.* 2017 Dec 15;12:78. doi: 10.1186/s40793-017-0292-9. eCollection 2017.

López G et al. Description of Alicyclobacillus montanus sp. nov., a mixotrophic bacterium isolated from acidic hot springs. *Int J Syst Evol Microbiol*. 2018 Mar 20. doi: 10.1099/ijsem.0.002718. [Epub ahead of print]

Mafu S et al. Discovery, Biosynthesis and Stress-Related Accumulation of Dolabradiene-Derived Defenses in Maize. *Plant Physiol.* 2018 Apr;176(4):2677-2690. doi: 10.1104/pp.17.01351. Epub 2018 Feb 23

Mackelprang R et al. Microbial Community Structure and Functional Potential in Cultivated and Native Tallgrass Prairie Soils of the Midwestern United States. *Front Microbiol.* 2018 Aug 15;9:1775. doi: 10.3389/ fmicb.2018.01775. eCollection 2018

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.

Mäkelä MR et al. Draft Genome Sequence of the Basidiomycete White-Rot Fungus Phlebia centrifuga. *Genome Announc.* 2018 Apr 5;6(14). pii: e01414-17. doi: 10.1128/genomeA.01414-17.

Marks PC et al. MetaHCR: a web-enabled metagenome data management system for hydrocarbon resources. *Database* (Oxford). 2018 Jan 1;2018:1-10. doi: 10.1093/database/bay087.

Martino E et al. Comparative genomics and transcriptomics depict ericoid mycorrhizal fungi as versatile saprotrophs and plant mutualists. *New Phytol.* 2018 Feb;217(3):1213-1229. doi: 10.1111/nph.14974. Epub 2018 Jan 7.

Matheus Carnevali PB et al. Distinct Microbial Assemblage Structure and Archaeal Diversity in Sediments of Arctic Thermokarst Lakes Differing in Methane Sources. *Front Microbiol.* 2018 Jun 7;9:1192. doi: 10.3389/fmicb.2018.01192. eCollection 2018.

McCormick RF et al. The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* 2017 Nov 21. doi: 10.1111/tpj.13781. [Epub ahead of print]

Mei R et al. Operation-driven heterogeneity and overlooked feed-associated populations in global anaerobic digester microbiome. *Water Res.* 2017 Nov 1;124(1):77-84. doi:10.1016/j.watres.2017.07.050

Meier-Kolthoff JP et al. Investigation of recombination-intense viral groups and their genes in the Earth's virome. *Sci Rep.* 2018 Jul 31;8(1):11496. doi: 10.1038/s41598-018-29272-2.

Meredith LK et al. Soil exchange rates of COS and CO18O differ with the diversity of microbial communities and their carbonic anhydrase enzymes. *ISME J.* 2018 Sep 13. doi: 10.1038/s41396-018-0270-2. [Epub ahead of print]

Meredith LK et al. Coupled biological and abiotic mechanisms driving carbonyl sulfide production in soils. *Soil Systems*. 2018 Jun 21;2(3):37, doi:10.3390/soilsystems2030037.

Milano ER et al. Quantitative trait loci for cell wall composition traits measured using near-infrared spectroscopy in the model C4 perennial grass Panicum hallii. *Biotechnol Biofuels*. 2018 Feb 3;11:25. doi: 10.1186/s13068-018-1033-z. eCollection 2018.

Miyauchi S et al. Integrative visual omics of the white-rot fungus Polyporus brumalis exposes the biotechnological potential of its oxidative enzymes for delignifying raw plant biomass. *Biotechnol Biofuels*. 2018 Jul 23;11:201. doi: 10.1186/s13068-018-1198-5. eCollection 2018.

Mondo SJ et al. Bacterial endosymbionts influence host sexuality and reveal reproductive genes of early divergent fungi. *Nat Commun*. 2017 Nov 29;8(1):1843. doi: 10.1038/s41467-017-02052-8.

Montero-Calasanz MDC et al. Genome-Scale Data Call for a Taxonomic Rearrangement of Geodermatophilaceae. Front Microbiol. 2017 Dec 19;8:2501. doi: 10.3389/ fmicb.2017.02501. eCollection 2017. Murat C et al. Draft Genome Sequence of Tuber borchii Vittad., a Whitish Edible Truffle. Genome Announc. 2018 Jun 21;6(25). pii: e00537-18. doi: 10.1128/ genomeA.00537-18.

Narrowe AB et al. Complex Evolutionary History of Translation Elongation Factor 2 and Diphthamide Biosynthesis in Archaea and Parabasalids. Genome Biol Evol. 2018 Sep 1;10(9):2380-2393. doi: 10.1093/gbe/ evy154.

Nash MV et al. Metagenomic insights into diazotrophic communities across Arctic glacier forefields. FEMS Microbiol Ecol. 2018 Sep 1;94(9). doi: 10.1093/femsec/ fiy114. Epub 2018 Jun 12

Neilson AD et al. Evolution and regulation of Bigelowiella natans light-harvesting antenna system. J Plant Physiol. 2017 Oct;217:68-76. doi: 10.1016/j. jplph.2017.05.019.

Nouioui I et al. Genome-Based Taxonomic Classification of the Phylum Actinobacteria. Front Microbiol. 2018 Aug 22;9:2007. doi: 10.3389/ fmicb.2018.02007. eCollection 2018.

O'Connor DL et al. Cross-species functional diversity within the PIN auxin efflux protein family. Elife. 2017 Oct 24;6. pii: e31804. doi: 10.7554/eLife.31804.

Pachiadaki MG et al. Major role of nitrite-oxidizing bacteria in dark ocean carbon fixation. Science. 2017 Nov 24:358(6366):1046-1051, doi: 10.1126/science. aan8260.

Palluk S et al. De novo DNA synthesis using polymerase-nucleotide conjugates. Nat Biotechnol. 2018 Aug;36(7):645-650. doi: 10.1038/nbt.4173. Epub 2018 Jun 18.

Pavlopoulos GA et al. Bipartite graphs in systems biology and medicine: a survey of methods and applications. Gigascience. 2018 Apr 1;7(4):1-31. doi: 10.1093/gigascience/giy014.

Pawlowska TE et al. Biology of Fungi and Their Bacterial Endosymbionts. Annu Rev Phytopathol. 2018 Aug 25;56:289-309. doi: 10.1146/annurevphyto-080417-045914.

Pelot KA et al. Functional Diversity of Diterpene Synthases in the Biofuel Crop Switchgrass. Plant Physiol. 2018 Sep;178(1):54-71. doi: 10.1104/ pp.18.00590. Epub 2018 Jul 15.

Perera D et al. Exploring the loblolly pine (Pinus taeda L.) genome by BAC sequencing and Cot analysis. Gene. 2018 Jul 15;663:165-177. doi: 10.1016/j. gene.2018.04.024. Epub 2018 Apr 12.

Perroud PF et al. The Physcomitrella patens gene atlas project: large scale RNA-seq based expression data. Plant J. 2018 Apr 22. doi: 10.1111/tpj.13940. [Epub ahead of print]

Pitt A et al. Polynucleobacter meluiroseus sp. nov., a bacterium isolated from a lake located in the mountains of the Mediterranean island of Corsica. Int J Syst Evol Microbiol. 2018 Jun;68(6):1975-1985. doi: 10.1099/ijsem.0.002777. Epub 2018 Apr 24.

Pold G et al. Draft Genome Sequences of Three Strains of a Novel Rhizobiales Species Isolated from Forest Soil. Genome Announc. 2018 Feb 1;6(5). pii: e01452-17. doi: 10.1128/genomeA.01452-17.

Pold G et al. Genome Sequence of Verrucomicrobium sp. Strain GAS474, a Novel Bacterium Isolated from Soil. Genome Announc. 2018 Jan 25;6(4). pii: e01451-17. doi: 10.1128/genomeA.01451-17.

Polle JEW et al. Draft Nuclear Genome Sequence of the Halophilic and Beta-Carotene-Accumulating Green Alga Dunaliella salina Strain CCAP19/18. Genome Announc. 2017 Oct 26;5(43). pii: e01105-17. doi: 10.1128/genomeA.01105-17.

Powell JJ et al. Transcriptome analysis of Brachypodium during fungal pathogen infection reveals both shared and distinct defense responses with wheat. Sci Rep. 2017 Dec 8;7(1):17212. doi: 10.1038/s41598-017-17454-3.

Praet J et al. Large-scale cultivation of the bumblebee gut microbiota reveals an underestimated bacterial species diversity capable of pathogen inhibition. Environ Microbiol. 2017 Oct 27. doi: 10.1111/1462-2920.13973

Price MN et al. Mutant phenotypes for thousands of bacterial genes of unknown function. Nature. 2018 May;557(7706):503-509. doi: 10.1038/s41586-018-0124-0. Epub 2018 May 16.

Probst AJ et al. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. Nat Microbiol. 2018 Mar;3(3):328-336. doi: 10.1038/s41564-017-0098-y. Epub 2018 Jan 29.

Probst AJ et al. Homologous Recombination and Transposon Propagation Shape the Population Structure of an Organism from the Deep Subsurface with Minimal Metabolism. *Genome Biol Evol.* 2018 Apr 1;10(4):1115-1119. doi: 10.1093/gbe/evy067.

Proença DN et al. Arboriscoccus pini gen. nov., sp. nov., an endophyte from a pine tree of the class Alphaproteobacteria, emended description of Geminicoccus roseus, and proposal of Geminicoccaceae fam. nov. *Syst Appl Microbiol.* 2017 Dec 22. pii: S0723-2020(17)30172-8. doi: 10.1016/j. syapm.2017.11.006. [Epub ahead of print]

Props R et al. Flow cytometric monitoring of bacterioplankton phenotypic diversity predicts high population-specific feeding rates by invasive dreissenid mussels. *Environ Microbiol*. 2017 Oct 13. doi:10.1111/1462-2920.13953

Raudabaugh DB et al. Coniella lustricola, a new species from submerged detritus. *Mycol Progress*. 2018
Jan;17(1-2),191-203. doi:10.1007/s11557-017-1337-6

Robicheau BM et al. The homothallic mating-type locus of the conifer needle endophyte Phialocephala scopiformis DAOMC 229536 (order Helotiales). *Fungal Biol.* 2017 Dec;121(12):1011-1024. doi: 10.1016/j. funbio.2017.08.007.

Rodriguez-R LM et al. How much do rRNA gene surveys underestimate extant bacterial diversity? *Appl Environ Microbiol.* 2018 Jan 5. pii: AEM.00014-18. doi: 10.1128/AEM.00014-18. [Epub ahead of print]

Rojas-Rojas FU et al. Draft genome of Paraburkholderia caballeronis TNe-841T, a free-living, nitrogen-fixing, tomato plant-associated bacterium. *Stand Genomic Sci.* 2017 Dec 16;12:80. doi: 10.1186/s40793-017-0294-7. eCollection 2017.

Roux S et al. Ecogenomics of virophages and their giant virus hosts assessed through time series metagenomics. *Nat Commun.* 2017 Oct 11;8(1):858. doi: 10.1038/s41467-017-01086-2.

Sade N et al. Salt tolerance of two perennial grass Brachypodium sylvaticum accessions. *Plant Mol Biol.* 2018 Jan 10. doi: 10.1007/s11103-017-0696-3. [Epub ahead of print]

Sancho R et al. Comparative plastome genomics and phylogenomics of Brachypodium: flowering time signatures, introgression and recombination in recently diverged ecotypes. *New Phytol.* 2017 Dec 5. doi: 10.1111/nph.14926. [Epub ahead of print]

Saski CA et al. Sub genome anchored physical frameworks of the allotetraploid Upland cotton (Gossypium hirsutum L.) genome, and an approach toward reference-grade assemblies of polyploids. *Sci Rep.* 2017 Nov 10;7(1):15274. doi: 10.1038/s41598-017-14885-w.

Sasse J et al. Feed Your Friends: Do Plant Exudates Shape the Root Microbiome? *Trends Plant Sci.* 2017 Oct 17. pii: S1360-1385(17)30199-1. doi: 10.1016/j. tplants.2017.09.003. [Epub ahead of print]

Schmid MW et al. Extensive epigenetic reprogramming during the life cycle of Marchantia polymorpha. *Genome Biol.* 2018 Jan 25;19(1):9. doi: 10.1186/s13059-017-1383-z.

Scholthof KG et al. Brachypodium: A Monocot Grass Model Genus for Plant Biology. *Plant Cell.* 2018 Aug;30(8):1673-1694. doi: 10.1105/tpc.18.00083. Epub 2018 Jul 11. Review.

Schulz F et al. Towards a balanced view of the bacterial tree of life. *Microbiome*. 2017 Oct 17;5(1):140. doi: 10.1186/s40168-017-0360-9.

Schwessinger B et al. A Near-Complete Haplotype-Phased Genome of the Dikaryotic Wheat Stripe Rust Fungus Puccinia striiformis f. sp. tritici Reveals High Interhaplotype Diversity. *MBio*. 2018 Feb 20;9(1). pii: e02275-17. doi: 10.1128/mBio.02275-17.

Scott KM et al. Genomes of ubiquitous marine and hypersaline Hydrogenovibrio, Thiomicrorhabdus, and Thiomicrospira spp. encode a diversity of mechanisms to sustain chemolithoautotrophy in heterogeneous environments. *Environ Microbiol.* 2018 Mar 9. doi: 10.1111/1462-2920.14090. [Epub ahead of print]

Sczyrba A et al. Critical Assessment of Metagenome Interpretation-a benchmark of metagenomics software. *Nat Methods.* 2017 Nov;14(11):1063-1071. doi: 10.1038/nmeth.4458. Epub 2017 Oct 2.

Seaver SMD et al. PlantSEED enables automated annotation and reconstruction of plant primary metabolism with improved compartmentalization and comparative consistency. *Plant J.* 2018 Sep;95(6):1102-1113. doi: 10.1111/tpj.14003.

Seshadri R et al. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat Biotechnol.* 2018 Apr;36(4):359-367. doi: 10.1038/nbt.4110. Epub 2018 Mar 19.

Sieber CMK et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol. 2018 Jul;3(7):836-843. doi: 10.1038/s41564-018-0171-1. Epub 2018 May 28.

Siletti CE et al. Distributions of fungal melanin across species and soils. Soil Biol and Biochem. 2017 Oct;113:285-293. doi:10.1016/j.soilbio.2017.05.030

Sipos G et al. Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi Armillaria. Nat Ecol Evol. 2017 Dec;1(12):1931-1941. doi: 10.1038/s41559-017-0347-8. Epub 2017 Oct 30.

Škraban J et al. Draft genome sequence of Chryseobacterium limigenitum SUR2T (LMG 28734T) isolated from dehydrated sludge. Braz J Microbiol. 2018 Jan - Mar;49(1):5-6. doi: 10.1016/j.bjm.2017.03.009. Epub 2017 Jul 18.

Slaby BM et al. Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization. ISME J. 2017 Nov;11(11):2465-2478. doi: 10.1038/ismej.2017.101.

Solomon KV et al. Catabolic repression in earlydiverging anaerobic fungi is partially mediated by natural antisense transcripts. Fungal Genet Biol. 2018 Sep 14. pii: \$1087-1845(18)30119-1. doi: 10.1016/j. fgb.2018.09.004. [Epub ahead of print]

Spindel JE et al. Association mapping by aerial drone reveals 213 genetic associations for Sorghum bicolor biomass traits under drought. BMC Genomics. 2018 Sep 17;19(1):679. doi: 10.1186/s12864-018-5055-5.

Spriggs A et al. Assembled genomic and tissue-specific transcriptomic data resources for two genetically distinct lines of Cowpea (Vigna unquiculata (L.) Walp). Gates Open Res. 2018 Feb 9;2:7. doi: 10.12688/ gatesopenres.12777.1.

Stein JC et al. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza. Nat Genet. 2018 Feb;50(2):285-296. doi: 10.1038/s41588-018-0040-0. Epub 2018 Jan 22.

Stolze Y et al. Targeted in situ metatranscriptomics for selected taxa from mesophilic and thermophilic biogas plants. Microb Biotechnol. 2018 Jul;11(4):667-679. doi: 10.1111/1751-7915.12982. Epub 2017 Dec 4. Stritt C et al. Recent activity in expanding populations and purifying selection have shaped transposable element landscapes across natural accessions of the Mediterranean grass Brachypodium distachyon. Genome Biol Evol. 2017 Dec 21. doi: 10.1093/gbe/ evx276. [Epub ahead of print]

Taş N et al. Landscape topography structures the soil microbiome in arctic polygonal tundra. Nat Commun. 2018 Feb 22;9(1):777. doi: 10.1038/s41467-018-03089-z.

Tian L et al. Enhanced ethanol formation by Clostridium thermocellum via pyruvate decarboxylase. Microb Cell Fact. 2017 Oct 4;16(1):171. doi: 10.1186/ s12934-017-0783-9.

Torres-Beltrán M et al. A compendium of geochemical information from the Saanich Inlet water column. Scientific Data. 2017 Oct 31. doi:10.1038/ sdata.2017.159

Tran P et al. Microbial life under ice: Metagenome diversity and in situ activity of Verrucomicrobia in seasonally ice-covered Lakes. Environ Microbiol. 2018 Jun 19. doi:10.1111/1462-2920.14283

Tschitschko B et al. Genomic variation and biogeography of Antarctic haloarchaea. Microbiome. 2018 Jun 20;6(1):113. doi: 10.1186/s40168-018-0495-3.

Tsemetzis N et al. Microbial community analysis of three hydrocarbon reservoir cores provides valuable insights for the assessment of reservoir souring potential. Int. Biodeterior. Biodegrad. 2018 Jan;(126):177-188. doi:10.1016/j.ibiod.2016.09.002

van der Schuren A et al. Broad spectrum developmental role of Brachypodium AUX1. New Phytol. 2018 Sep;219(4):1216-1223. doi: 10.1111/ nph.15332. Epub 2018 Jun 27.

Varriale S et al. Talaromyces borbonicus, sp. nov., a novel fungus from biodegraded Arundo donax with potential abilities in lignocellulose conversion. Mycologia. 2018 Apr 23:1-9. doi: 10.1080/00275514.2018.1456835. [Epub ahead of print]

Vavourakis CD et al. A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. Microbiome. 2018 Sep 19;6:168. doi:10.1186/s40168-018-0548-7

Vaz-Moreira I et al. Oryzisolibacter propanilivorax gen. nov., sp. nov., a propanil-degrading bacterium. Int J *Syst Evol Microbiol.* 2017 Oct;67(10):3752-3758. doi: 10.1099/ijsem.0.002184.

Vigneron A et al. Comparative metagenomics of hydrocarbon and methane seeps of the Gulf of Mexico. *Sci Rep.* 2017 Nov 22;7(1):16015. doi: 10.1038/s41598-017-16375-5.

Vigneron A et al. Beyond the tip of the iceberg; a new view of the diversity of sulfite- and sulfate-reducing microorganisms. *ISME J.* 2018 Aug;12(8):2096-2099. doi: 10.1038/s41396-018-0155-4. Epub 2018 May 28.

Vögeli B et al. Combining Promiscuous Acyl-CoA Oxidase and Enoyl-CoA Carboxylase/Reductases for Atypical Polyketide Extender Unit Biosynthesis. *Cell Chemical Biology*. 2018 May 3;25(7):833-839.e4. doi:10.1016/j.chembiol.2018.04.009

Walters WA et al. Large-scale replicated field study of maize rhizosphere identifies heritable microbes. *Proc Natl Acad Sci U S A.* 2018 Jul 10;115(28):7368-7373. doi: 10.1073/pnas.1800918115. Epub 2018 Jun 25.

Wang W et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature*. 2018 May;557(7703):43-49. doi: 10.1038/s41586-018-0063-9. Epub 2018 Apr 25.

Warshan D et al. Genomic changes associated with the evolutionary transitions of Nostoc to a plant symbiont. *Mol Biol Evol.* 2018 Mar 15. doi: 10.1093/molbev/msy029. [Epub ahead of print]

Wehking J et al. Community composition and seasonal changes of archaea in coarse and fine air particulate matter. *Biogeosciences*. 2018 Jul 11;15, 4205-4214. doi:10.5194/bg-15-4205-2018

West PT et al. Genome-reconstruction for eukaryotes from complex natural microbial communities. *Genome Res.* 2018 Apr;28(4):569-580. doi: 10.1101/gr.228429.117. Epub 2018 Mar 1.

Weston DJ et al. The Sphagnome Project: enabling ecological and evolutionary insights through a genus-level sequencing project. *New Phytol.* 2018 Jan;217(1):16-25. doi: 10.1111/nph.14860. Epub 2017 Oct 27.

Wilhelm RC et al. Biogeography and organic matter removal shape long-term effects of timber harvesting on forest soil microbial communities. *ISME J.* 2017 Nov;11(11):2552-2568. doi: 10.1038/ismej.2017.109.

Woodcroft BJ et al. Genome-centric view of carbon processing in thawing permafrost. *Nature*. 2018 Aug;560(7716):49-54. doi: 10.1038/s41586-018-0338-1. Epub 2018 Jul 16.

Woyke T et al. The trajectory of microbial single-cell sequencing. *Nat Methods*. 2017 Oct 31;14(11):1045-1054. doi: 10.1038/nmeth.4469. Review.

Wu GA et al. Genomics of the origin and evolution of Citrus. *Nature*. 2018 Feb 15;554(7692):311-316. doi: 10.1038/nature25447. Epub 2018 Feb 7.

Wu B et al. Substrate-Specific Differential Gene Expression and RNA editing in the Brown Rot Fungus Fomitopsis pinicola. *Appl Environ Microbiol.* 2018 Aug 1;84(16). pii: e00991-18. doi: 10.1128/AEM.00991-18. Print 2018 Aug 15.

Xie L et al. Methylglucosylation of aromatic amino and phenolic moieties of drug-like biosynthons by combinatorial biosynthesis. *Proc Natl Acad Sci USA*. 2018 May 29;115(22):E4980-E4989. doi:10.1073/pnas.1716046115

Xie M et al. A 5-enolpyruvylshikimate 3-phosphate synthase functions as a transcriptional repressor in Populus. *Plant Cell.* 2018 Jun 11. pii: tpc.00168.2018. doi: 10.1105/tpc.18.00168. [Epub ahead of print]

Yang J et al. Brassinosteroids Modulate Meristem Fate and Differentiation of Unique Inflorescence Morphology in Setaria viridis. *Plant Cell.* 2018 Jan;30(1):48-66. doi: 10.1105/tpc.17.00816. Epub 2017 Dec 20.

Yang S et al. Complete genome sequence and the expression pattern of plasmids of the model ethanologen Zymomonas mobilis ZM4 and its xylose-utilizing derivatives 8b and 2032. *Biotechnol Biofuels*. 2018 May 2;11:125. doi: 10.1186/s13068-018-1116-x

Yang X et al. Amplification and adaptation of centromeric repeats in polyploid switchgrass species. *New Phytol.* 2018 Jun;218(4):1645-1657. doi: 10.1111/nph.15098. Epub 2018 Mar 25.

Yang X et al. The Kalanchoë genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. *Nat Commun.* 2017 Dec 1;8(1):1899. doi: 10.1038/s41467-017-01491-7.

Yao Q et al. Community proteogenomics reveals the systemic impact of phosphorus availability on microbial functions in tropical soil. Nat Ecol Evol. 2018 Mar;2(3):499-509. doi: 10.1038/s41559-017-0463-5. Epub 2018 Jan 22.

Yassin AF et al. Draft genome sequence of Actinotignum schaalii DSM 15541T: Genetic insights into the lifestyle, cell fitness and virulence. PLoS One. 2017 Dec 7;12(12):e0188914. doi: 10.1371/journal. pone.0188914. eCollection 2017.

Yin H et al. Diel rewiring and positive selection of ancient plant proteins enabled evolution of CAM photosynthesis in Agave. BMC Genomics. 2018 Aug 6;19(1):588. doi: 10.1186/s12864-018-4964-7.

Yoo CG et al. Insights of biomass recalcitrance in natural Populus trichocarpa variants for biomass conversion. Green Chem. 2017 Oct 9;19:5467-5478. doi: 10.1039/C7GC02219K

Zhalnina K et al. Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. Nat Microbiol. 2018 Apr;3(4):470-480. doi: 10.1038/ s41564-018-0129-3. Epub 2018 Mar 19.

Zhang J et al. Genome-wide association studies and expression-based quantitative trait loci analyses reveal roles of HCT2 in caffeoylquinic acid biosynthesis and its regulation by defense-responsive transcription factors in Populus. New Phytol. 2018 Jul 11. doi: 10.1111/nph.15297. [Epub ahead of print]

Zhou EM et al. Thermus sediminis sp. nov., a thiosulfate-oxidizing and arsenate-reducing organism isolated from Little Hot Creek in the Long Valley Caldera, California. Extremophiles. 2018 Sep 15. doi: 10.1007/s00792-018-1055-2. [Epub ahead of print]

Zhou R et al. Characterization of a large sex determination region in Salix purpurea L. (Salicaceae). Mol Genet Genomics. 2018 Jul 18. doi: 10.1007/ s00438-018-1473-y. [Epub ahead of print.]

Zhu T et al. Analysis of Brachypodium genomes with genome-wide optical maps. Genome. 2018 Aug;61(8):559-565. doi: 10.1139/gen-2018-0013. Epub 2018 Jun 8.

Zhu Y et al. Subnanogram proteomics: Impact of LC column selection, MS instrumentation and data analysis strategy on proteome coverage for trace samples. Int J Mass Spectrom. in press, available online 1 September 2017 doi:10.1016/j.ijms.2017.08.016

Zuo C et al. Revealing the transcriptomic complexity of switchgrass by PacBio long-read sequencing. Biotechnol Biofuels. 2018 Jun 20;11:170. doi: 10.1186/ s13068-018-1167-z.

Comments?

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

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

Follow Us















Ongoing construction of the Integrative Genomics Building (IGB), future home of the JGI and KBase at Berkeley Lab, as of February 2019. (Marilyn Chung, Berkeley Lab)

Facing page:

Top (left to right): 2018 Women @ the Lab awardee Danielle Goudeau, WSEC Empower subcommittee chair and event emcee Esther Singer, Lab Director Mike Witherell, 2018 awardee Alicia Clum, and 2013 awardee Susannah Tringe; Jenifer Kaplan.

Center (left to right): 2017 LBNL Director's Awards for Exceptional Achievement recipients Ray Turner for Operations and Christine Naca for Safety; Juna Lee.

Bottom (left to right): Nikos Kyrpides, the USFCC/J. Roger Porter Award recipient at the 2018 ASM Microbe Meeting; Megan Pawlowski; Tanja Woyke at the Annual Foothill Middle School STEAM (Science, Technology, Engineering, Arts, and Mathematics) Day.

(Images by Marilyn Chung and Thor Swift, Berkeley Lab; David Gilbert, JGI)

Greta Lorge and Cindi Hoover contributed to the Science: Year in Review section.

Layout and Design: Creative Services, Information Technology Division, 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.

