2018
Progress Report
U.S. Department of Energy
Joint Genome Institute
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 strigosa-zonata (Robert Blanchette, University of Minnesota); Ivotuk range, Alaska (LANL–Cathy Wilson); Fistulated cow (Jonas Lavaas Gjerstad); poplar leaf (David Gilbert, JGI).
Table of Contents
<table>
<thead>
<tr>
<th>Page</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>JGI Mission</td>
</tr>
<tr>
<td>2</td>
<td>Director’s Perspective</td>
</tr>
<tr>
<td>8</td>
<td>Achieving the DOE Mission</td>
</tr>
<tr>
<td>10</td>
<td>Organizational Structure</td>
</tr>
<tr>
<td>12</td>
<td>Impact 2018</td>
</tr>
<tr>
<td>18</td>
<td>Case Study: JGI’s Community Science Program at Fifteen</td>
</tr>
<tr>
<td>20</td>
<td>Science: A Year in Review</td>
</tr>
<tr>
<td>21</td>
<td>Discovery</td>
</tr>
<tr>
<td>30</td>
<td>Bioenergy</td>
</tr>
<tr>
<td>46</td>
<td>Biogeochemistry</td>
</tr>
<tr>
<td>54</td>
<td>Computational Infrastructure</td>
</tr>
<tr>
<td>56</td>
<td>Appendices</td>
</tr>
<tr>
<td>57</td>
<td>Appendix A: Acronyms at a Glance</td>
</tr>
<tr>
<td>59</td>
<td>Appendix B: Glossary</td>
</tr>
<tr>
<td>62</td>
<td>Appendix C: 2018 User Program Supported Proposals</td>
</tr>
<tr>
<td>68</td>
<td>Appendix D: Advisory and Review Committee Members</td>
</tr>
<tr>
<td>70</td>
<td>Appendix E: 13th Annual Genomics of Energy and Environment Meeting</td>
</tr>
<tr>
<td>74</td>
<td>Appendix F: 2018 Publications</td>
</tr>
</tbody>
</table>
JGI Mission

View of the San Francisco Bay from Berkeley.
(Mark Lilly Photography)
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.
Director’s Perspective

Nigel Mouncey, Director, DOE Joint Genome Institute
In Support of Community-Driven Integrative Genome Science

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 outweights 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*)
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)

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
Achieving the DOE Mission
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

Steven Wilson  
Systems Engineering

Nikos Kyrpides  
Prokaryote Super Program

Tanja Woyke  
Microbial Program

Emiley Eloe-Fadrosh  
Metagenome Program

Alex Copeland  
Genome Assembly

Hugh Salamon  
Advanced Analysis

Yasuo Yoshikuni  
DNA Synthesis Science Program

Trent Northen  
Metabolomics Group

David Gilbert  
Senior Manager, Communications & Outreach
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.

1,882

**Spending Profile FY2018**

- **35%** Science Programs & Analysis
- **32%** Genomic Technologies
- **11%** Data Science & Informatics
- **6%** Compute Infrastructure and Support Team (@ NERSC)
- **6%** Management
- **4%** Operations
- **3%** Lease
- **3%** Project Management Office
- **1%** Emerging Technologies Opportunity Program (ETOP)
### Users on the Map: 1,882

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic</td>
<td>1,365</td>
</tr>
<tr>
<td>Other</td>
<td>352</td>
</tr>
<tr>
<td>DOE National Laboratory</td>
<td>143</td>
</tr>
<tr>
<td>Company</td>
<td>22</td>
</tr>
</tbody>
</table>

### North America: 1,358

<table>
<thead>
<tr>
<th>Country</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>1,276</td>
</tr>
<tr>
<td>Canada</td>
<td>76</td>
</tr>
<tr>
<td>Mexico</td>
<td>5</td>
</tr>
<tr>
<td>Panama</td>
<td>1</td>
</tr>
</tbody>
</table>

### South America: 17

<table>
<thead>
<tr>
<th>Country</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brazil</td>
<td>12</td>
</tr>
<tr>
<td>Chile</td>
<td>2</td>
</tr>
<tr>
<td>Colombia</td>
<td>2</td>
</tr>
<tr>
<td>Peru</td>
<td>1</td>
</tr>
</tbody>
</table>

### Europe: 372

<table>
<thead>
<tr>
<th>Country</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>21</td>
</tr>
<tr>
<td>Belgium</td>
<td>8</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>4</td>
</tr>
<tr>
<td>Denmark</td>
<td>11</td>
</tr>
<tr>
<td>Estonia</td>
<td>1</td>
</tr>
<tr>
<td>Finland</td>
<td>14</td>
</tr>
<tr>
<td>France</td>
<td>59</td>
</tr>
<tr>
<td>Germany</td>
<td>84</td>
</tr>
<tr>
<td>Greece</td>
<td>2</td>
</tr>
<tr>
<td>Hungary</td>
<td>4</td>
</tr>
<tr>
<td>Iceland</td>
<td>1</td>
</tr>
<tr>
<td>Italy</td>
<td>20</td>
</tr>
<tr>
<td>Netherlands</td>
<td>27</td>
</tr>
<tr>
<td>Norway</td>
<td>8</td>
</tr>
<tr>
<td>Poland</td>
<td>2</td>
</tr>
<tr>
<td>Portugal</td>
<td>4</td>
</tr>
<tr>
<td>Russian Federation</td>
<td>4</td>
</tr>
<tr>
<td>Slovenia</td>
<td>1</td>
</tr>
<tr>
<td>Spain</td>
<td>31</td>
</tr>
<tr>
<td>Sweden</td>
<td>16</td>
</tr>
<tr>
<td>Switzerland</td>
<td>11</td>
</tr>
<tr>
<td>Turkey</td>
<td>1</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>38</td>
</tr>
<tr>
<td>Africa</td>
<td>Asia</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
</tr>
<tr>
<td>Egypt</td>
<td>China</td>
</tr>
<tr>
<td>South Africa</td>
<td>India</td>
</tr>
<tr>
<td>Tunisia</td>
<td>Iran</td>
</tr>
<tr>
<td></td>
<td>Israel</td>
</tr>
<tr>
<td></td>
<td>Japan</td>
</tr>
<tr>
<td></td>
<td>Malaysia</td>
</tr>
<tr>
<td></td>
<td>Republic of Korea</td>
</tr>
<tr>
<td></td>
<td>Singapore</td>
</tr>
<tr>
<td></td>
<td>Taiwan</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Workshops and Meetings</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomics of Energy &amp; Environment 13th Annual User Meeting Participants:</td>
<td>602</td>
</tr>
<tr>
<td>VEGA Symposium</td>
<td>134</td>
</tr>
<tr>
<td>Other Workshop Participants</td>
<td>879</td>
</tr>
</tbody>
</table>

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.

Massively Parallel Short Read Sequencing

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>62</td>
<td>1,004</td>
<td>6,041</td>
<td>30,000</td>
<td>55,905</td>
<td>70,370</td>
<td>100,013</td>
<td>141,707</td>
<td>139,964</td>
<td>174,519</td>
<td>217,995</td>
</tr>
</tbody>
</table>

Single Molecule Long Read Sequencing

<table>
<thead>
<tr>
<th>Year</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>210</td>
<td>492</td>
<td>596</td>
<td>1,470</td>
<td>1,907</td>
<td>3,625</td>
<td>7,305</td>
</tr>
</tbody>
</table>
Cumulative Number of Scientific Publications

User Demand: Synthetic Biology
Requested & Approved (Mb)

User Program Growth Submitted & Accepted

Cumulative Number of Projects Completed
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.

Under the umbrella of the Genomic Encyclopedia of Bacteria and Archaea (GEBA), the JGI has been systematically bringing to light and filling in uncharted branches in the bacterial and archaeal tree of life, the so-called “microbial dark matter.”
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 iron-oxide 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.
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**

Yeast 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](http://bit.ly/JGI2018Jeffries).)

Yeast 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](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)
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.
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.
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.
A Reference Catalog for the Rumen Microbiome

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 Hungate1000 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 sequence-based 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 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 \textit{de novo} synthesis of vitamin B\textsubscript{12}, and potential vertical inheritance of the rumen microbiome, are discussed in 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 moellendorffii* 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

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](http://brachypan.jgi.doe.gov). The *B. distachyon* genome is available on the JGI plant portal Phytozome: [phytozome.jgi.doe.gov](http://phytozome.jgi.doe.gov).
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 *Eucalyptus* 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.
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.
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**

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.

*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 endobacteria* 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-\textit{Burkholderia} relationship, \textit{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.

\textbf{All in the Family: Focused Genomic Comparisons}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{aspergillus_colonies}
\caption{Colonies of \textit{Aspergillus} \textit{A. campestris}, \textit{A. ochraceoroseus}, and \textit{A. steynii}. (Kirstine Ellen Lyhne, Technical University of Denmark)}
\end{figure}

Found in microbial communities around the world, \textit{Aspergillus} fungi are pathogens, decomposers, and important sources of biotechnologically-important enzymes. Each \textit{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 \textit{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 \textit{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+ \textit{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 \textit{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 \textit{Aspergillus} species. A comparative analysis involving these genomes and other \textit{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)
Geothermal outflow channels with pH ~ 8 like the ones seen at Yellowstone National Park’s Upper Geyser Basin are excellent habitats for high-temperature phototrophic microbial communities. Different types of phototrophic organisms living at temperatures from 60–70 °C give rise to a beautiful display of colors resulting from different photosystem pigments. These communities are excellent model environments for understanding light harvesting ecosystems, where numerous ecotypes colonize across gradients of light (depth) and temperature. (Miles O’Brien/William Inskeep)
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.
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 Nanohaloarchaeota.

“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.

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.
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
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.
Genepool Cluster
70+ million CPU-hours

Cori
15 million CPU-hours

JAMO
6.6 million files records

JAMO Archived Data Footprint
7.4 Petabytes

Data downloads in FY18
1PB, 1 million files
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.
Appendices
## Appendix A
### Acronyms at a Glance

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

**Annotation:** The process of identifying the locations of genes in a genome and determining what those genes do to improve 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 single-stranded clusters of template DNA.

**Carbon cycle:** The biogeochemical process by which carbon is exchanged among the planet’s atmosphere, land, and oceans.

**CAZymes:** Carbohydrate-active enzymes that can break down plant polysaccharides such as cellulose into small sugars.

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

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

**ChIP-Seq:** A method of analyzing protein interactions with DNA 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 sequence-tagged insertion mutant strains of a single-celled organism (typically bacteria) to simultaneously assess the functions of every gene in the genome in a variety of conditions by sequencing and counting the abundance of each tag.

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

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

**Sequence:** Order of nucleotides (base sequence) in a nucleic acid molecule. In the case of DNA sequence, it is the precise ordering of the bases (A, T, G, 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)

<table>
<thead>
<tr>
<th>INVESTIGATOR</th>
<th>AFFILIATION</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banfield, Jill</td>
<td>University of California, Berkeley</td>
<td>LBNL Watershed Function SFA: Biogeochemical Dynamics from Genome to Watershed Scales</td>
</tr>
<tr>
<td>Beam, Jake</td>
<td>Bigelow Laboratory for Ocean Sciences</td>
<td>Emergent Properties of Marine and Freshwater Sediment Chemolithoautotrophic Microbial Communities</td>
</tr>
<tr>
<td>Berka, Randy</td>
<td>Novozymes</td>
<td>Genus-wide Genomics of the Biomass-Degrading and Plant-Beneficial <em>Trichoderma</em></td>
</tr>
<tr>
<td>Bhaya, Devaki</td>
<td>Stanford University</td>
<td>Cyanophages and Cyanobacteria in Extreme Environments</td>
</tr>
<tr>
<td>Buckley, Dan</td>
<td>Cornell University</td>
<td>Microbial Metabolic Dependency and its Impacts on the Soil Carbon Cycle</td>
</tr>
<tr>
<td>Catalan, Pilar</td>
<td>Universidad de Zaragoza (Spain)</td>
<td>Genomic Characterization of the <em>Brachypodium</em> Polyploid Model to Unravel Bases of Success of Polyploidy in Flowering Plants</td>
</tr>
<tr>
<td>Coleman, Maureen</td>
<td>University of Chicago</td>
<td>Integrated Ecosystem Genomics across a Vast and Vital Freshwater System</td>
</tr>
<tr>
<td>Crump, Byron</td>
<td>Oregon State University</td>
<td>Combined Metagenomics and Metatranscriptomics to Evaluate Dissolved Organic Matter Metabolizing Microbes in Big Rivers of the U.S.</td>
</tr>
<tr>
<td>Cullen, Daniel</td>
<td>Forest Products Laboratory</td>
<td>Gene Expression in the Unusual White Rot Fungus <em>Phlebiopsis gigantea</em></td>
</tr>
<tr>
<td>Dijkstra, Paul</td>
<td>Northern Arizona University</td>
<td>Partitioning Flux between Entner-Doudoroff and Embden-Meyerhof-Parnas Glycolysis in Soil Communities</td>
</tr>
<tr>
<td>Duplessis, Sebastien</td>
<td>INRA (France)</td>
<td>Rust Pangenomics: Understanding the Diversity and Potential Impact of Rust Fungi — a Systematic Collection of Genomes &amp; Transcriptomes</td>
</tr>
<tr>
<td>Fukami, Tadashi</td>
<td>Stanford University</td>
<td>Genomic Basis of the Ecological Success of Nectar Yeasts in their Carbon-Stressed and Nitrogen-Limited Environments</td>
</tr>
<tr>
<td>INVESTIGATOR</td>
<td>AFFILIATION</td>
<td>DESCRIPTION</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Garre, Victoriano</td>
<td>University of Murcia (Spain)</td>
<td>Exploring the Role of DNA Methylation in Biofuel Production, Environmental Sensing and Development in Basal Fungi</td>
</tr>
<tr>
<td>Huffaker, Alisa</td>
<td>University of California, San Diego</td>
<td>Systems Analysis of Secondary-Metabolism-Mediated Microbial Community Interactions in Sorghum and Maize Enabled by Comparative Metabolomic and Transcriptomic Genome-Wide Association Studies</td>
</tr>
<tr>
<td>Kalluri, Udaya</td>
<td>Oak Ridge National Laboratory</td>
<td>Reciprocal Impacts of Modified Plant Cell Wall and Associated Microbiome</td>
</tr>
<tr>
<td>Kistler, Harold</td>
<td>USDA ARS Cereal Disease Lab, University of Minnesota</td>
<td>Mechanisms of Co-Evolutionary Adaptation of Soil Microbes</td>
</tr>
<tr>
<td>Leebens-Mack, James</td>
<td>University of Georgia</td>
<td>Open Green Genomes: a Framework for Comparative Plant Genomics</td>
</tr>
<tr>
<td>Liu, Yu</td>
<td>UT Southwestern Medical Center</td>
<td>Determination of Fungal Chromatin Regulatory Network and Its Impact on Gene Expression</td>
</tr>
<tr>
<td>Mandadi, Kranthi</td>
<td>Texas A&amp;M University</td>
<td>Gene Atlas of Diverse Grass-Microbe Interactions in Brachypodium and Setaria</td>
</tr>
<tr>
<td>Martinez-Gomez, Norma</td>
<td>Michigan State University</td>
<td>Dissecting the Genetic and Functional Diversity of Rare Earth-dependent Plant-Microbiome Interactions by Sequencing Analysis</td>
</tr>
<tr>
<td>Martiny, Jennifer</td>
<td>University of California, Irvine</td>
<td>Investigation of Diel Variation in Microbial Decomposition Processes</td>
</tr>
<tr>
<td>Reese, Brandi</td>
<td>Texas A&amp;M University-Corpus Christi</td>
<td>Dynamics of Seasonal and Diurnal Fluctuations in a Wetland Mangrove Ecosystem</td>
</tr>
<tr>
<td>Stuart, Rhona</td>
<td>Lawrence Livermore National Laboratory</td>
<td>Flipping the Switch: the Molecular Mechanisms Underlying Trophic Strategy Versatility in Parasitic Chytrids</td>
</tr>
<tr>
<td>Sullivan, Matt</td>
<td>Ohio State University</td>
<td>Elucidating Viral “Dark Matter” and Biogeochemical Impacts in Extreme Environments</td>
</tr>
</tbody>
</table>
Community Science Program (CSP) (continued)

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

Small-Scale Microbial/Metagenome

<table>
<thead>
<tr>
<th>INVESTIGATOR</th>
<th>AFFILIATION</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atsumi, Shota</td>
<td>University of California, Davis</td>
<td>Genomic Analysis, and Reconstruction of Isobutyl Acetate Tolerance in Escherichia coli</td>
</tr>
<tr>
<td>Barnhart, Elliott</td>
<td>U.S. Geological Survey</td>
<td>Multi-omic Sequencing of Sulfate Transition Zones in the Terrestrial Subsurface with Recalcitrant Carbon</td>
</tr>
<tr>
<td>Epstein, Slava</td>
<td>Northeastern University</td>
<td>Microbial Stem Cell Hypothesis</td>
</tr>
<tr>
<td>Högfors-Rönnholm, Eva</td>
<td>Novia University of Applied Sciences (Finland)</td>
<td>Microbial Community Structure and Function of an Actual and Potential Acid Sulfate Soil</td>
</tr>
<tr>
<td>Izquierdo, Javier</td>
<td>Hofstra University</td>
<td>Plant Growth-Promoting Bacteria from the Rhizosphere of the Beachgrass Ammophila brevilligulata</td>
</tr>
<tr>
<td>INVESTIGATOR</td>
<td>AFFILIATION</td>
<td>DESCRIPTION</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Liu, Xiao-Jun Allen</td>
<td>University of Massachusetts Amherst</td>
<td>Disentangling the Relative Contributions of the Microbiome and Physical Protection in Soil Response to Long-Term Environmental Stress</td>
</tr>
<tr>
<td>Mason, Olivia</td>
<td>Florida State University</td>
<td>Coupled Plant: Microbe Interactions Mediate Carbon and Biogeochemical Cycling in the Marsh Rhizosphere</td>
</tr>
<tr>
<td>McMahon, Katherine</td>
<td>University of Wisconsin-Madison</td>
<td>Biogeochemical Gradients Structure Mercury-Methylating Microbial Communities in a Reservoir System</td>
</tr>
<tr>
<td>Nicol, Graeme</td>
<td>Ecole Centrale de Lyon (France)</td>
<td>Determining the Interaction of Viruses with Prokaryotic Hosts Controlling Nitrogen Cycling in Soil</td>
</tr>
<tr>
<td>Peay, Kabir</td>
<td>Stanford University</td>
<td>How Does Precipitation Impact the Taxonomic and Functional Diversity of the Populus Trichocarpa Soil Microbiome?</td>
</tr>
<tr>
<td>Potnis, Neha</td>
<td>Auburn University</td>
<td>Unraveling the Diversity of Plant-Associated Saprophytic/Non-Pathogenic Bacteria and Their Role in Plant Health and Plant-Pathogen Interactions</td>
</tr>
<tr>
<td>Richardson, Ruth</td>
<td>Cornell University</td>
<td>Metagenomic Exploration of Microbial Communities involved in Carbon and Sulfur Cycling in Two Central New York State Peatlands</td>
</tr>
<tr>
<td>Selbmann, Laura</td>
<td>University of Tuscia (Italy)</td>
<td>Metagenomic Reconstruction of Endolithic Communities from Victoria Land, Antarctica</td>
</tr>
<tr>
<td>Thamatrakoln, Kim</td>
<td>Rutgers University</td>
<td>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</td>
</tr>
<tr>
<td>Treusch, Alexander</td>
<td>University of Southern Denmark (Denmark)</td>
<td>Impacts of Climate Change Induced Flooding of Coastal Soils on Carbon Cycling and Sequestration – 3.0</td>
</tr>
<tr>
<td>Voges, Mathias</td>
<td>Stanford University</td>
<td>Elucidating the Biological Role of Root Exudates: Investigating the Transcriptional Response of Pseudomonas simiae WCS417 to Novel Specialized Plant Metabolites.</td>
</tr>
<tr>
<td>Ziels, Ryan</td>
<td>University of British Columbia (Canada)</td>
<td>Genome-Centric Metagenomics and Metatranscriptomics to Resolve Adaptive Capacities of Methane-Producing Biofilms to High Salinity Concentrations</td>
</tr>
</tbody>
</table>
### DNA Synthesis

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

<table>
<thead>
<tr>
<th>INVESTIGATOR</th>
<th>AFFILIATION</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alper, Hal</td>
<td>University of Texas at Austin</td>
<td>Pre-Optimized Cell Free Lysates for Rapid Prototyping of Genes and Pathways</td>
</tr>
<tr>
<td>Jewett, Michael</td>
<td>Northwestern University</td>
<td></td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>INVESTIGATOR</th>
<th>AFFILIATION</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bowen, Jennifer</td>
<td>Northeastern University</td>
<td>Combining High Resolution Organic Matter Characterization and Microbial Meta-Omics to Assess the Effects of Nutrient Loading on Salt Marsh Carbon Sequestration</td>
</tr>
<tr>
<td>Mayes, Melanie</td>
<td>Oak Ridge National Laboratory</td>
<td>Linking Proteogenomics, Metabolomics, and Soil Organic Chemistry of Tropical Wetlands to a Soil Nutrient Cycling Model</td>
</tr>
<tr>
<td>Onstott, Tullis</td>
<td>Princeton University</td>
<td>Detecting Seismically-Sustained Deep Subsurface CH₄-Cycling Chemolithoautotrophic Microbial Communities Using Multi-Omic Analyses and NanoSIMS</td>
</tr>
<tr>
<td>Saleska, Scott</td>
<td>University of Arizona</td>
<td>Investigating the Carbon Cycling Implications of Changing Microbial Leaf Litter Decomposition across a Permafrost Thaw Gradient</td>
</tr>
<tr>
<td>Talbot, Jennifer</td>
<td>Boston University</td>
<td>Scaling Molecular Mechanisms of Mycorrhizal-Decomposer Interactions to Emergent Ecosystem Carbon Balance</td>
</tr>
<tr>
<td>Tiemann, Lisa</td>
<td>Michigan State University</td>
<td>Tracking Switchgrass Photosynthesize via 13CO₂ Pulse-Chase into the Rhizosphere Microbiome and Metabolome</td>
</tr>
</tbody>
</table>
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 Tresee, 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, Lawrence 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

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

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

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.

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.”


Nigel Mouncey, Kristala Prather, Susannah Tringe
John Ioannidis 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 Ioannidis, the goal is the same: to enhance trust in scientific findings.


“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

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

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

Other Featured Speakers (in order of appearance):

Tim James, University of Michigan
Mercè Montoliu Nerin, 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
Ian 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.

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


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


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]


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]


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]


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.