

2017 Community Science Program Portfolio Announced



From deep within the Earth to the upper atmosphere, the organisms and ecosystems highlighted in the 37 projects selected for the 2017 Community Science Program (CSP) of the U.S. Department of Energy Joint Genome Institute (DOE JGI), a DOE Office of Science User Facility, reflect the breadth and depth of interests researchers are exploring to find solutions to energy and environmental challenges.

“These new CSP projects, selected through our external review process, exploit DOE JGI’s experimental and analytical “omics” capabilities and build our portfolio in key focus areas

The common or crystalline ice plant (*Mesembryanthemum crystallinum* L.) (Image by Krzysztof Ziarnik, Kenraiz CC BY-SA 4.0 Wikipedia)

including sustainable bioenergy production, plant microbiomes and terrestrial biogeochemistry,” said Susannah Tringe, DOE JGI User Programs Deputy.

The CSP 2017 projects were selected from 98 full proposals received, resulting from 123 letters of intent submitted. The full list of projects may be found at <http://bit.ly/JGI-CSP2017portfolio>.

continued on page 2

in this issue

2017 Community Science Program	2
Fungi Help Beat Drought Stress . .	3
Global Viral Diversity	4
2017 Collaborative Science Projects with EMSL	5
Expanding the Database of Workhorse Yeasts	6
Visualizing Uncultured Microbial Cell Activity	7
Science Highlights	8

Haystack of Datasets Yields Earth’s Viral Diversity

Microbes are known to play crucial roles in regulating carbon fixation, as well as maintaining global cycles involving nitrogen, sulfur, and phosphorus and other nutrients, but the majority of them remain uncultured and unknown. Plumbing the Earth’s microbial diversity requires learning more about the poorly-studied relationships between microbes and the viruses that infect them, viruses that impact the microbes’ abilities to regulate global cycles. Although the number of viruses is estimated to be at least two orders of magnitude more than the number of microbial cells on the planet, there are currently less than 2,200 sequenced DNA virus genomes, compared to the approximately 50,000 bacterial genomes, in sequence databases.

In a study published August 25, 2016 in *Nature*, DOE JGI researchers utilized the largest collection of assembled metage- *continued on page 4*

2017 Portfolio

continued from page 1

A number of the accepted proposals call for developing reference genomes for plants of relevance to bioenergy, either as potential feedstocks or because they possess biochemistries that could provide insights into better ways to derive biofuels from plant biomass. These will take advantage of powerful DOE JGI sequencing upgrades.

Karen Aitken of Australia's national science agency CSIRO is focused on producing the first genome assembly of a cultivated sugarcane plant, namely variety R570. Cultivated sugarcane produces 80 percent of the world's sugar, is already harvested in large volumes, and is optimized for transport from farm to market. These factors are helping to guide optimized production strategies for other renewable and sustainable biofuel crops.

Many projects focus on plants with demonstrated tolerance to stressors such as drought. The proposal from John Cushman of the University of Nevada seeks to establish the common or crystalline ice plant (*Mesembryanthemum crystallinum* L.) as a DOE JGI Flagship Genome species (more on these at <http://bit.ly/JGI-Plants>) like

poplar, soybean, and *Brachypodium distachyon*. Most plants use what is known as the C3 pathway to photosynthetically fix carbon (from CO₂) in the presence of abundant nutrients, while some plants in arid conditions rely on a different mechanism, the water-conserving CAM pathway. A rapidly growing desert annual native to the Namibian desert, the common ice plant is the first reported species that can switch from the C3 pathway to CAM when stressed by factors such as drought and salinity.

In a complementary project, Karolina Heyduk of the University of Georgia seeks genome sequences for two agave relatives: the C3 species *Yucca aloifolia* and the CAM species *Y. filamentosa*. Comparing the genomes of these species could illuminate the genes and pathways involved in these two forms of photosynthesis.

Several projects focus on plant-microbe interactions, both on micro- and macroscales. One of them comes from Sharon Doty of the University of Washington, who is investigating poplar endophytes, one that fixes nitrogen and one involved in phytoremediation. Plant root microbial

consortia are known to be critically important to plant growth and resilience to changing soil conditions but much remains to be learned about community composition and function.

Other projects, including one from Christine Smart of Cornell University, focus on plant-fungal interactions. She is characterizing the mechanisms of willow rust pathogenesis (*Melampsora americana*), and conducting a comparative genomic study involving *M. americana* and other *Melampsora* genomes sequenced by the DOE JGI that are known plant pathogens of candidate bioenergy feedstocks.

Soil emissions are a focus for Jennifer Pett-Ridge of Lawrence Livermore National Laboratory. Pett-Ridge is determining the impact microbes in humid tropical forest soils have on carbon cycling, work that complements her DOE Early Career Research Program award.

Mark Dopson of Sweden's Linnaeus University has a project that deploys DOE JGI's single cell genomics resources on samples sourced from deep subsurface. Targeted microbiomes will come from deep bedrock waters including Mont Terri Underground Rock Lab in Switzerland and sites constructed to investigate the safety of deep geological storage of spent nuclear fuel. Data would provide knowledge contributing to informed decisions on safety.

Ashley Shade of Michigan State University plans to use a synthetic bacterial community to study the interactions among multiple community members, and then link community structure to activity and to exometabolite interactions.

Learn more about all of the selected projects at <http://bit.ly/JGI2017CS-Pannouncement>.



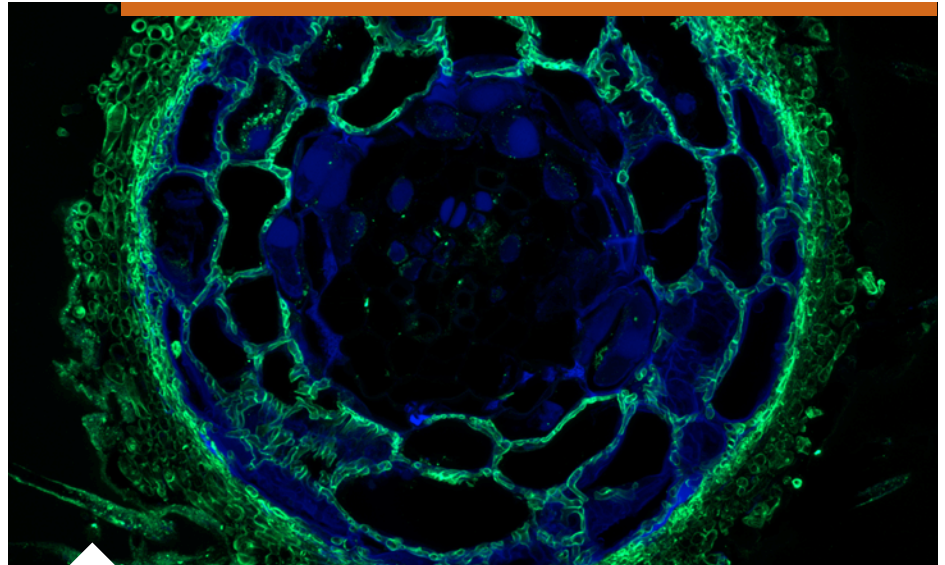
The 2017 portfolio includes a project that deploys DOE JGI's single cell genomics resources on samples sourced from the deep subsurface. (Courtesy of Mark Dopson)

Drought Resistance Conferred by Fungi on Trees

In forests, ectomycorrhizal (ECM) fungi are key players supporting the growth, health and stress tolerance of trees such as oak, pine, spruce, birch and beech, and help boost the productivity of bioenergy feedstock trees, including poplar and willow. The most common ECM fungus is *Cenococcum geophilum*, and it is the only mycorrhizal fungus in the Dothideomycetes, a large class comprised of some 19,000 fungal species, many of them plant pathogens.

To learn more about what ectomycorrhizal characteristics are dominant in *C. geophilum*, a team led by researchers at the French National Institute for Agricultural Research (INRA) and the Swiss Federal Institute for Forest, Snow and Landscape Research WSL compared its genome with the genomes of close relatives that are not ECM fungi. In the study published online September 7, 2016 in *Nature Communications*, they found specific adaptations in the *C. geophilum* transcriptome—the set of its genes that are expressed as messenger RNA molecules encoding proteins—that could help their hosts be more resistant to drought stress, a finding that could be useful in developing more plant feedstocks for bioenergy amidst the changing climate.

As part of a comparative genomic analysis done through the Mycorrhizal Genomics Initiative (MGI) headed by study senior author Francis Martin of INRA, the DOE JGI sequenced *C. geophilum* and its close relative *Lepidopterella palustris*, and annotated both of these genomes and another close relative, *Glomium stellatum*. “We showed that the genome of *C. geophilum*, the only known mycorrhizal symbiont within the largest fungal class Dothideomycetes, acquired the same genomic adaptations to the mycorrhizal lifestyle over



The crosscut shows the fungal tissues—the fungal mantle around the root tip and the fungal network of tendrils that penetrates the root of plants, or Hartig Net, between *Pinus sylvestris* plant root cells—in green. (Image by Maira de Freitas Pereira, INRA Nancy)

generations as the previously sequenced ectomycorrhizal basidiomycetes,” Martin said. “These include a strikingly reduced number of plant cell wall degrading enzymes (PCWDEs) and a large set of symbiosis-induced lineage-specific genes, including dozen of mycorrhiza-induced small secreted effector-like proteins (MiSSPs).” Unlike free-living saprotrophs, fungi that get their nutrients from decomposing organic matter in forest soils and so require PCWDEs, *Cenococcum* has come to rely heavily on its hosts for its carbon nutrition.

Noting that the tree root tips colonized with *C. geophilum* are highly resistant to desiccation, one of the team’s key findings is that two of the three most highly induced *C. geophilum* genes in symbiosis code for water channels. “*C. geophilum* population genomics should shed light on the mechanisms of host and environmental adaptation,” the team wrote in their paper. “It should facilitate the identification of

drought-adapted *C. geophilum* strains, which can be used to efficiently support their host trees threatened by the forecasted increase in drought periods in many parts of the world.”

“By combining genome sequences with rigorous physiological and ecological studies, we are entering a time where linking the presence, composition and abundance of soil mycorrhizal communities with important soil processes and forest productivity at an ecosystem scale is possible,” said Martin. Learn more at <http://bit.ly/JGICenococcumdrought>.

A 2015 *Nature Genetics* paper by the MGI in collaboration with the DOE JGI was recently recognized by *La Recherche*, the most prestigious French monthly popular science magazine, with the Lauréat du Prix La Recherche (mention environnement). Martin attended the ceremony and received the award on behalf of the team. Learn more about that project at <http://bit.ly/JGI15pfsymbioses>.

Haystack of Datasets Yields Earth's Viral Diversity

continued from page 1

onomic datasets from around the world to uncover over 125,000 partial and complete viral genomes, the majority of them infecting microbes. This single effort increased the number of known viral genes by a factor of 16, and provides researchers with a unique resource of viral sequence information.

"It is the first time that someone has looked systematically across all habitats and across such a large compendium of data," said study senior author and DOE JGI Prokaryote Super Program head Nikos Kyrpides. "A key to uncover all these novel viruses was the sensitive computational approach we have developed along with this work."

"We are constantly searching for regulatory DNA parts that will work across many different phyla, and that would allow us to build genes and pathways that can express in many different hosts."

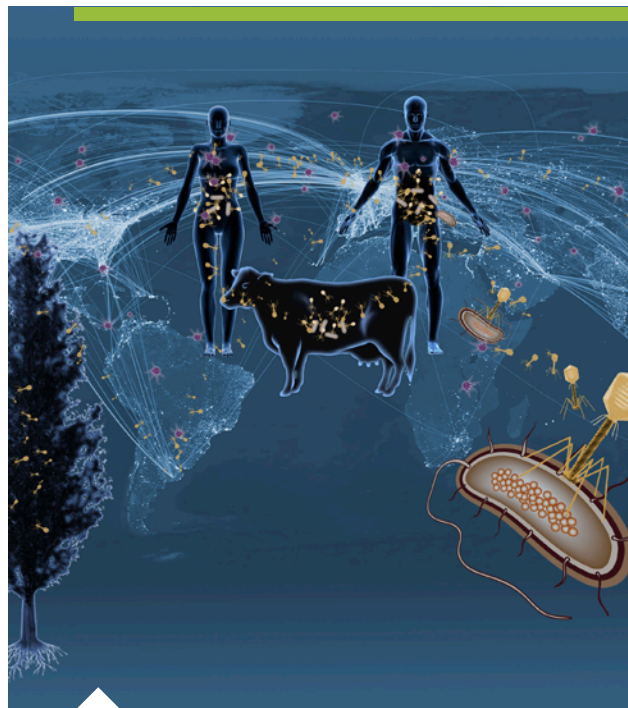
—Jan-Fang Cheng

That approach, explained first author David Paez-Espino, involved using a non-targeted metagenomic approach, referencing both isolate viruses and manually curated viral protein models, and what he described as "the largest and most diverse dataset to date." The team analyzed over 5 trillion bases of sequence available in the DOE JGI's Integrated Microbial Genomes with Microbiome Samples (IMG/M) system collected from 3,042 samples around

the world from 10 different habitat types. Their efforts to sift through the veritable haystack of datasets yielded over 125,000 viral sequences containing 2.79 million proteins.

The team matched viral sequences against multiple samples in multiple habitats. For example, one viral group they identified was found in 95 percent of all samples in the ocean's twilight zone—a region located between 200 and 1,000 meters below the ocean surface where insufficient sunlight penetrates for microorganisms to perform photosynthesis.

By analyzing a CRISPR-Cas system—an immune mechanism in bacteria that confers resistance to foreign genetic elements by incorporating short sequences from infecting viruses and phages—the team was able to generate a database of 3.5 million spacer sequences in IMG. These spacers, fragments of phage genetic sequences retained by the host, can then be used to explore viral and phage metagenomes for where the fragments may have originally come from. Also, using mainly this approach, the team computationally identified the host for nearly 10,000 viruses. "The majority of these connections were previously unknown, and include the identification of organisms serving as viral hosts from 16 prokaryotic phyla for which no



DOE JGI researchers utilized the largest collection of assembled metagenomic datasets from around the world to uncover over 125,000 partial and complete viral genomes. (Graphic by Zosia Rostomian, Berkeley Lab)

viruses have previously been identified," they reported.

Jan-Fang Cheng, head of the DOE JGI's Functional Genomics group, said the work being done by Kyrpides' group in identifying new viral sequences will help the Synthetic Biology group develop novel promoters that can work in many bacterial hosts. "We are constantly searching for regulatory DNA parts that will work across many different phyla, and that would allow us to build genes and pathways that can express in many different hosts."

The work also used resources at the National Energy Research Scientific Computing Center, another DOE Office of Science User Facility at Lawrence Berkeley National Laboratory. More at <http://bit.ly/JGI17viraldiversity>.

Joint DOE User Facility Initiative Announces 10 New Projects



PhD student Moira Hough from the Rich lab at OSU surveying plant community composition. (Credit: Amelia McClure)

The DOE JGI and the Environmental Molecular Sciences Laboratory (EMSL) accepted 10 projects submitted during the 2017 call for proposals for their joint “Facilities Integrating Collaborations for User Science” (FICUS) initiative. The accepted proposals began on October 1, 2016 and fall under the following focused topic areas: Plant-Microbe Interactions, Biofuels and Bioproducts, and Biogeochemistry of Select Inorganics. The full list of approved projects is available at <http://bit.ly/JGI2017FICUSportfolio>.

These new research projects all involve collaboration between two user facilities that are stewarded by the DOE Office of Biological and Environmental Research and integrate the expertise and capabilities to address pressing questions in support of DOE’s energy, environment and basic research missions. The FICUS initiative represents a unique opportunity for scientists to harness the combined power of genomics and molecular characterization in a single research project. The researchers, whose

projects were selected by outside reviewers, will access the capabilities of both user facilities to generate datasets unique to these two providers of critical infrastructure—deriving benefits beyond what could be generated by either facility by itself.

Among the selected projects:

- Virginia Rich of The Ohio State University is using a well-studied “model ecosystem” of permafrost thaw in subarctic Sweden. She and her colleagues intend to look at the pathways for carbon loss in thawing permafrost, and metabolic changes in peatland microbial communities that are regulating the carbon flux.
- Petr Baldrian of the Academy of Sciences of the Czech Republic is studying soil and leaf litter microbes in coniferous forests to learn more about the impact of nitrogen availability on carbon cycle-related processes. Through metabolomics analyses, researchers hope to identify drivers of seasonal changes

in microbial function—both inside and outside the tree root—as well as the interplay between tree roots and their symbiont fungi. Additionally, the fungal genomes that are being obtained through the DOE JGI’s 1000 Fungal Genomes Project (more at <http://bit.ly/1000-JGI-Fungi>) are expected to provide ecological perspectives.

- Tom Bianchi of University of Florida is investigating a phenomenon known as the “priming effect” that is well-studied in soils but poorly understood in the marine environment. Along the Florida coastline are mangroves, salt marshes and seagrasses that trap “blue carbon,” organic carbon derived from and stored in coastal and marine environments. Lab experiments show that in the presence of highly reactive substrates, dissolved organic carbon breaks down up to 75 times faster during the first 24 hours. One of the team’s questions is how the priming effect might influence blue carbon stores impacted by rising sea levels and the corresponding loss of wetlands.
- Scott Fendorf of Stanford University is targeting the persistence of contaminants in groundwater. Focusing on the Colorado River Basin, where legacy contamination has left floodplains with uranium traces, his team aims to learn about the metabolic constraints of organic matter decomposition in floodplain sediments, and how this regulation of the carbon cycle impacts metal cycling.

Read the full story to learn about all of the projects at <http://bit.ly/JGI2017FICUSannouncement>.

Expanding the Stable of Workhorse Yeasts

Yeasts are physically hard to distinguish, and it is easy to think they are all the same. Metabolically, genetically and biochemically, however, yeasts are highly diverse. Characteristics such as their thick cell walls and tolerance of pressure changes that could rupture other cells mean yeasts are easily scaled up for industrial processes. In addition, they are easy to grow and modify and, with notable exceptions, most are not associated with human illness. While these capabilities can be used for a wide range of biotechnological applications, including biofuel production, so far industry has only harnessed a fraction of the diversity available among yeast species.

To help boost the use of a wider range of yeasts and to explore the use of genes and pathways encoded in their genomes, a team led by DOE JGI researchers conducted a comparative genomic analysis of 29 yeasts, including 16 whose genomes were newly sequenced and annotated. In the study published August 30, 2016 in the *Proceedings of the National Academy of Sciences (PNAS)*, the team mapped various metabolic pathways to yeast growth profiles.

Yeasts (which belong to the “kingdom” of fungi) can use a wide range of carbon and energy sources, ranging from cellulosic (6-carbon) and hemicellulosic (5-carbon) sugars to methanol, glycerol, and acetic acid. Products include ethanol and other alcohols, esters, organic acids, carotenoids, lipids, and vitamins. “We sequenced these diverse genomes to expand the catalog of genes, enzymes, and pathways encoded in these genomes for producing biofuels and bio-based products we use in daily life,” said Igor Grigoriev, JGI Fungal Program Head and co-senior author of the manuscript.

Sequencing these less-known yeasts and characterizing their

metabolic pathways, added study first author Robert Riley of the DOE JGI, helps fill in knowledge gaps regarding the fungal enzymes that can help convert a wide range of sugars into biofuel. One of the newly-sequenced yeasts is *Pachysolus tannophilus*, which can ferment xylose, otherwise known as wood sugar as it is derived hemicellulose, an abundant constituent of woody biomass. It is only distantly related to well-studied xylose fermenters such as *Scheffersomyces stipitis*—another yeast sequenced by the DOE JGI.

These distances are huge. “We might think of yeasts as simple unicellular creatures similar to each other, but in fact their genetic diversity is like the difference between human and invertebrate sea squirt,” said Riley. “We sequenced these diverse genomes to discover and facilitate the next generation of biotechnological workhorse

yeasts for producing the fuels and products we use in daily life. We also discovered a genetic code change that, if not understood, will impede the yeasts’ biotechnological use.” Read the full story at <http://bit.ly/JGI16workhorseyeasts>.

In *P. tannophilus*, the team found a change in one of the three-letter codons that represent one of the 20 regularly used amino acids. That change from CUG-Ser to CUG-Ala is only the second observed case of a non-stop codon reassignment (a change from one amino acid to another, rather than from one amino acid to a stop codon) in nuclear genomes. First author Riley talked about finding a fungus in which the 3-letter codon CUG is interpreted as the amino acid serine rather than the expected lysine during the 2016 DOE JGI Genomics of Energy & Environment Meeting. Watch his talk at <http://bit.ly/JGI2016Riley>.



Scheffersomyces stipitis. (Courtesy of Tom Jeffries, University of Wisconsin, Madison)

Novel Techniques to Visualize Uncultured Microbial Cell Activity

Many uncultured microbes play unknown roles in regulating Earth's biogeochemical processes; everything from regulating plant health to driving nutrient cycles in both terrestrial and marine environments, processes that can impact global climate. While researchers are harnessing multiple approaches to identify these microbes, referred to as "microbial dark matter," and determine what they're doing, most techniques don't allow them to do both at once.

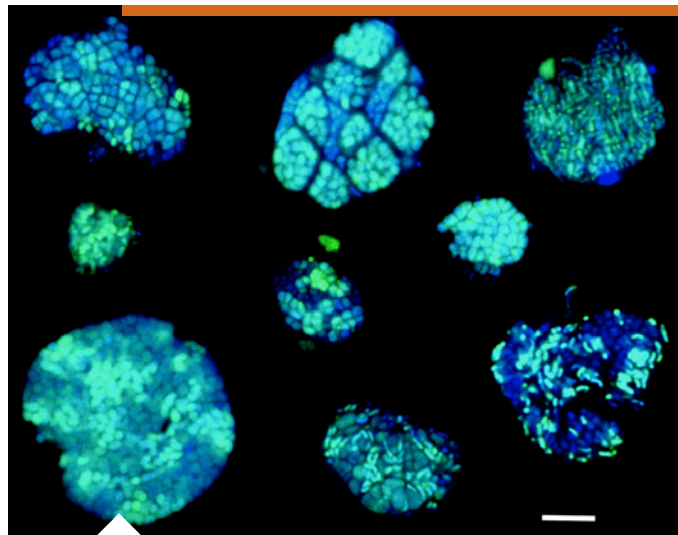
In a study published July 12, 2016 in the *Proceedings of the National Academy of Sciences (PNAS)*, researchers at the California Institute of Technology (Caltech) and the DOE JGI utilized a recently refined technique to identify both individual active cells, and single clusters of active bacteria and archaea within microbial communities. Understanding the true scope of the planet's microbial diversity is of interest to the DOE in order to learn how they can be harnessed for a wide range of energy and environmental challenges.

"When we deal with environmental or host-associated organisms that live within complex communities – typically thousands of species or more co-occur at the same time – it is very hard to determine what an individual species or cell is up to," said study first author Roland Hatzenpichler, then a postdoctoral researcher working in the Caltech lab of microbial ecologist Victoria Orphan. As part of the DOE Biological & Environmental Research (BER) program, Orphan's lab has been studying the ecology and physiology of uncultured microorganisms catalyzing the anaerobic oxidation of methane in the environment. Orphan herself has been named a 2016 MacArthur Fellow for her work studying microbial communities in

extreme environments. "Historically, this has been accomplished using single cell resolved isotope labeling. This is a comparatively tedious and expensive approach that is, however, very specific and sensitive. However, if we want to better understand microbiome functioning we need new, complementary approaches that are more high-throughput, that can be parallelized, and that are cheap. BONCAT [the new technique] is such a method."

Short for **B**io**O**rthogonal **N**on-**C**anonical **A**mino acid **T**agging, BONCAT is a technique developed at Caltech for bioengineering studies, but was adapted by Hatzenpichler and Orphan for use in microbial ecology investigations. BONCAT uses synthetic amino acids to label protein-making cells. These amino acids can then be stained with a fluorescent tag, lighting up active cells in their habitat.

The Caltech researchers combined BONCAT with fluorescent *in situ* hybridization (FISH) to analyze active microbes within simulated environments and identify under which conditions they were active. DOE JGI researchers then helped develop a process that allowed the team to apply BONCAT to flow cytometry. "It's the same process as for single cell genomics, but for aggregates of symbiotic bacteria and archaea," said DOE JGI



The depicted cell consortia are composed of archaea and bacteria that, by combining their individual metabolic powers, together achieve oxidation of the greenhouse gas methane in the absence of oxygen. For the first time, these consortia have now been analyzed for their protein synthesis activity under a variety of conditions. In this image, nucleic acids within individual cells are stained in blue. Green color indicates that new proteins have been made. The scale bar equals 10 micrometer, or 1/8th the width of a human hair. (Credit: Roland Hatzenpichler)

Microscale Applications Group head Rex Malmstrom of the technique called BONCAT-FACS (BONCAT — **F**luorescence-**A**ctivated **C**ell **S**orting). "We're sorting individual aggregates full of cells that are active in a simulated environment and with the cell sorter we can grab active cells to study them. We had to figure out how to adapt the process for the flow cytometer. After the sorting, we performed whole genome amplification on individual aggregates then sequenced the 16S rRNA marker genes to identify cells comprising the aggregate. We're now doing whole genome sequencing on these samples." Sequencing could then provide researchers the genetic information explaining why some microbes were active under certain conditions, and why some others were not. Full story at <http://bit.ly/JGI16BONCAT>.

Microbial Metabolism Impacts Fracking Sustainability

By 2040, shale formations are expected to be the source of half of all the natural gas used in the United States. Extracting these natural gas reserves is made dramatically easier by hydraulic fracturing, or “fracking.” While the economic benefits and some environmental impacts are being studied, less attention has been focused on the impacts to the deep subsurface and the microbial communities therein.

Published September 6, 2016 in *Nature Microbiology*, a team led by the Ohio State University’s Kelly Wrighton has been able to reconstruct microbial genomes for the first time from shale formations. Coupled with microbial metabolic information, the data shed light on the impacts of hydraulic fracturing on microbial communities in the deep subsurface, and offers insight on the sustainability of energy extraction through this approach.

Wrighton’s work was done through a collaborative science program called Facilities Integrating Collaborations for User Science (FICUS) led by the DOE JGI and the Environmental Molecular Sciences Laboratory (EMSL) at Pacific Northwest National Laboratory. She spoke about her FICUS project at the DOE JGI 2016 Annual Genomics of Energy & Environment Meeting. Watch her talk at <http://bit.ly/JGI2016Wrighton>.

Modeling Microbial Networks in an Oxygen Minimum Zone

With help from the DOE JGI and EMSL, a team at the University of British Columbia (UBC) has developed a math model that could help researchers and policy makers track the impact of climate change on the microbial networks that drive the world’s marine ecosystems. The model, reported in the October 4, 2016 issue of the *Proceedings of the National Academy of Sciences*, couples omic information and biogeochemical data from the

Saanich Inlet ecosystem in British Columbia, Canada, which has long been used as a model for studying oxygen minimum zones.

“It takes a network to study networks. A drop of seawater can contain millions of single-celled microbes that collectively form the basis for nutrient and energy cycles in the ocean,” said DOE JGI collaborator and UBC microbial ecologist Steven Hallam. “Understanding how microbial processes contribute to these cycles is vital in a time of climate change. Our model provides a step change in more accurate projections of microbial processes with potential feedback on climate change and ecosystem health.”

SAR11 Lineages Adapted to Oxygen-Poor Ocean Zones

Oxygen minimum zones (OMZs) are regions in the ocean where very low levels of oxygen are available and nitrate reduction to nitrite becomes the dominant oxidative process, impacting the composition of microbial communities and their effect on global biogeochemical cycles. Though SAR11 bacteria are estimated to make up as much as half of the total microbial community in the ocean’s oxygen-rich surface layers, they have also been found to be abundant in OMZs. To understand how SAR11 bacteria can thrive in what was thought to be an inhospitable environment for them, a team led by Georgia Tech scientists analyzed single amplified genomes (SAGs) generated from water samples collected in the nitrogen reduction zone of the Eastern Tropical North Pacific OMZ. DOE JGI researchers conducted a series of single-cell sorting and synthetic biology experiments that helped to demonstrate how SAR11 cells adapted to anoxic environments. The study appeared in *Nature* on August 11, 2016.

Read these and other DOE JGI Science Highlights at <http://bit.ly/JGIsciHighlights>.

REGISTER NOW

12th Annual Genomics of Energy & Environment Meeting

March 20–23, 2017
Walnut Creek, California
<http://usermeeting.jgi.doe.gov/>

Microbial Genomics & Metagenomics (MGM) Workshop

February 27–March 3, 2017
Walnut Creek, California
<http://mgm.jgi.doe.gov>

NeLLi: From New Lineages of Life to New Functions

April 5–6, 2017
Walnut Creek, California
<http://jgi.doe.gov/events/nelli/>

Contact The Primer

Massie Santos Ballon, Editor
m1ballon@lbl.gov

