THE PRIMER

An intricately structured soil bacterium, less than a micron in size, makes its home on the root surface of an *Arabidopsis* plant. The image is from a related DOE project at the Environmental Molecular Sciences Laboratory, a DOE national scientific user facility located at Pacific Northwest National Laboratory, to understand how carbon within the root zone impacts the diversity and function of the rhizosphere microbial community. (Courtesy of Pacific Northwest National Laboratory. Image was captured with the Helios Nanolab dual-beam focused ion beam/scanning electron microscope at EMSL and was created by Alice Dohnalkova.)

Regulating Root Microbiomes

"Plants grow in soil and soil is full of microbes. Many trade nutrients with the plant," said Jeff Dangl, a National Academy of Sciences member, Howard Hughes Medical Institute Investigator, and the John Couch Distinguished Professor of Biology at University of North Carolina. "Plants could select from among the complex community for strains that help them. But it is a crowded and complex ecosystem. And there are microbes that take plant nutrients and damage the plants. They are pathogens."

Understanding how plant yields can be optimized, in part by optimizing their microbial partners, is of fundamental interest to farmers and crop breeders working on developing sustainable crops for the production of food and advanced fuels from plant biomass amidst the pressure exerted by an ever-increasing global population. A plant's immune system can distinguish between friends and foes among these *continued on page 3*

Stresses for Sustainable Fuel Production

Like plants, algae can convert light into energy-rich chemical compounds; unlike plants, they require little space and don't need arable soil to grow. As part of the DOE Office of Science's efforts to study algae for energy and environmental applications, the U.S. Department of Energy Joint Genome Institute (DOE JGI), a DOE Office of Science User Facility, has published over 75 percent of all publicly available algal genomes. Some algae like *Chlamydomonas reinhardtii* (or "Chlamy," as it's known to its large research community) produce energy-dense oils or lipids when stressed, and these lipids can then be converted into fuels.

However, researchers walk a fine line in stressing the algae just enough to produce lipids, but not enough to kill them. Published online July 27, 2015 in the journal *Nature Plants*, a team led by DOE JGI scientists analyzed the genes that are being activated during algal lipid production, and in particular the molecular machinery that orchestrates these gene activities inside the cell. The work is expected to help algal *continued on page 2*
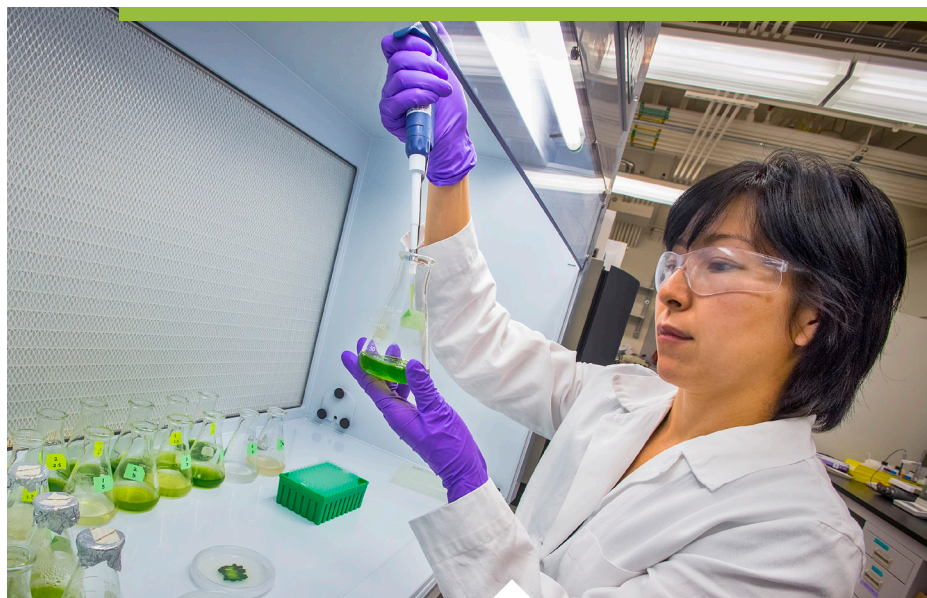
Stresses for Sustainable Fuel Production

bioenergy researchers develop more targeted approaches for producing lipids for fuels.

For this study, the team made use of the Chlamy reference genome, which was released by the DOE JGI back in 2007. "We know how to stress the algae," said the study's first author Chew Yee Ngan of the DOE JGI. "What we don't know is how to keep the algae alive at the same time, until now."

Very little is known about the protein factor that can regulate lipid production. To find more of them, the team cultured Chlamy cells and starved them of nitrogen or sulfur, both of which are stress conditions to which Chlamy responds by producing lipids. They then analyzed the complex of DNA and proteins known as chromatin that defines what genes are being activated, as well as the expression profiles or transcriptome, and compared these to non-stressed Chlamy cells.

Study co-author Yuko Yoshinaga works with *C. reinhardtii* cells. The team identified a transcription factor that appears to play a pivotal role in lipid accumulation, which could be applied to other algal species for commercial biofuel production. (Roy Kaltschmidt, Berkeley Lab)

"We're looking for changes in starved cells vs. cells that are happily growing," Ngan explained. Through careful analysis of genome-wide data sets, they narrowed down their search to identify two transcription factors that appeared to play a pivotal role in lipid accumulation, and then studied one of them, *PSR1*, in detail. "In studying the chromatin modifications, we can read out changes in the proteins bound to DNA on a genome-wide scale and then specifically target those genes whose regulation profiles are changed under lipid-producing conditions."

"The study also demonstrated how cells can be tricked into producing lots of lipid without dying of starvation by overexpression of *PSR1*, which is a strategy that could potentially be

applied in other industrial algal species better suited for large-scale biofuel production," said study co-author Axel Visel, DOE JGI Deputy for Science Programs.

Corresponding author Chia-Lin Wei, head of DOE JGI's Sequencing Technologies Program, pointed out that this study also successfully demonstrated an effective strategy for the integration of gene expression data and epigenomic methods, such as the mapping of molecular tags that sit on top of the actual DNA sequence and affect its function, in an organism relevant to DOE missions in energy and environment. She added that the work "is expected to be widely applicable to more plants and fungi whose gene regulatory pathways still prove elusive."

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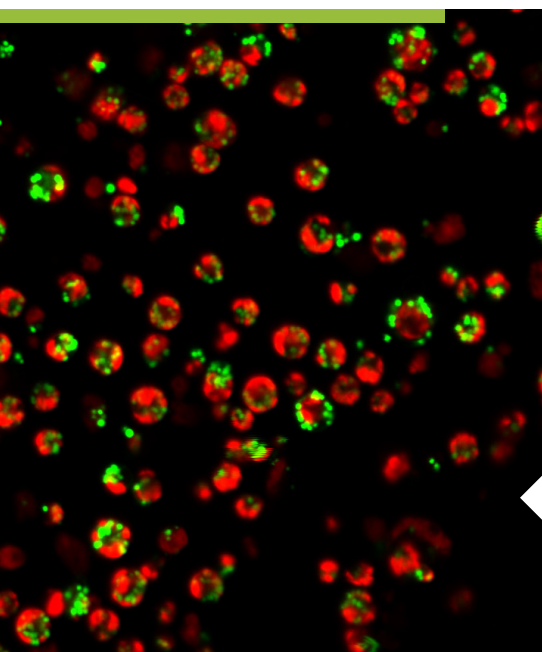
Regulating Root Microbiomes

microbes, and upon detecting pathogens, can produce regulatory chemicals called phytohormones to activate a defensive response.

In a study published August 21, 2015 in *Science*, a team led by Dangl and colleagues, including scientists from the DOE JGI, looked at roles of three phytohormones in controlling the composition of the root microbiome in the model plant *Arabidopsis thaliana*. The work was initiated as part of a DOE JGI Grand Challenge project studying the root-associated microbiomes of *Arabidopsis*, then grew into both a Community Science Program project and a NSF-funded Microbial Systems Biology grant investigating multiple plant species in the greenhouse and the field.

"This is a key focus area for the DOE JGI, as we believe cultivation of beneficial microbiomes will be critical to sustainable bioenergy crop production in the future," said DOE JGI Metagenome Program Lead and newly appointed Deputy for User Programs Susannah Tringe.

To better understand the impact of phytohormones on root microbiomes, the team compared wild type root microbiomes with those of mutants lacking at least one of the following phytohormones: salicylic acid (SA), jasmonic acid or ethylene. Previous studies had shown that the compositions of root microbiomes can vary after infection, even if there were similar starting microbial communities.

Building off early findings, study co-first author Sarah Lebeis and her colleagues examined root microbiomes using bacterial isolates cultured from plants grown in the wild soils to colonize sterile roots in a sterile soil-like substrate. These synthetic bacterial communities in the soil-like substrate and within the plant roots

Arabidopsis **plants growing in wild soil from Mason Farm. (Derek Lundberg, Dangl Lab)**

were tracked for eight weeks following inoculation. Among the bacteria used for these communities were nearly 40 bacterial isolates whose genomes had been sequenced at the DOE JGI.

"Deployment of these genomes in fully controlled reconstruction experiments allowed us to go beyond a description of the microbes present on the root in wild soil, toward systems to define the mechanisms that control microbe abundance. This also demonstrates the power of obtaining the genomes of environmental isolates," said Lebeis, who added that the work confirmed their hypothesis that salicylic acid influences root microbiome composition. Even in this controlled setting, and importantly in the absence of pathogens, they were able to show the plant immune system has major effects on colonization by root-associated bacteria.

Knowing the roles of salicylic acid and other plant hormones in regulating the root microbiome composition could lead to an agricultural approach in which plants are regularly treated with the equivalent of an adult multivitamin, but how to improve pathogen defense without affecting growth remains to be determined.

"The immune system function for SA is defined largely in leaves, and yes, if you spray SA or chemical analogue of SA, one can make plants more resistant to a certain sort of pathogen," Dangl speculated. "But one does not want to induce the plant immune system all the time, in much the same way you don't want to always have a fever. We need to figure out how SA helps to determine the balance of bacterial families that accumulate on and inside a healthy root, and how that results in benefits to the plant."

Tools for Computational Challenges

The rapid explosion in the throughput of DNA sequencing due to new technology platforms is fueling an increase in the number of sequenced microbial genomes and driving much greater availability of these data to the research community. However, they also bring new challenges such as sequence decontamination of publicly available microbial genomes, and a more systematic method for classifying microbial species. In separate articles, DOE JGI-led teams described computational genomics tools developed to help researchers.

> *"With an emphasis on Biological Data Interpretation, the DOE JGI has played a leadership role in developing, standardizing and providing access for users to high-quality genome assemblies, annotations and other computational genomics tools."*
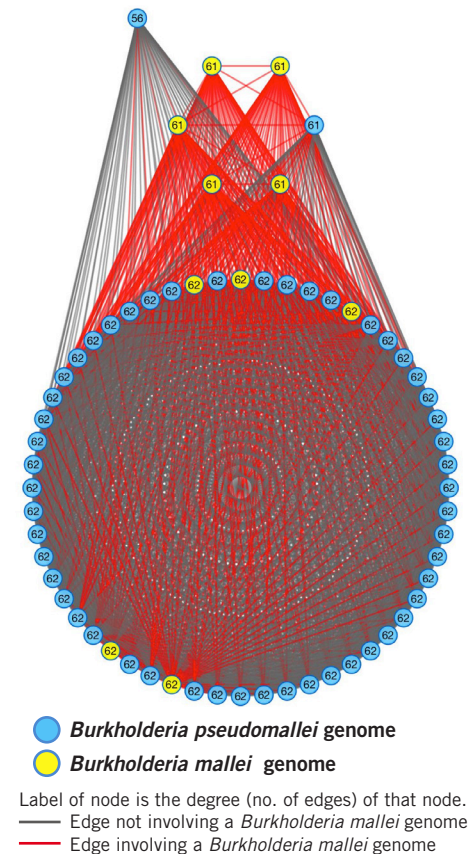>
> *–Nikos Kyrpides*

The development of computational tools is essential to characterizing complex biological and environmental systems in support DOE's research missions, as well as the DOE JGI's 10-Year Strategic Vision (http://bit.ly/JGI-Vision) and partnership with the National Energy Research Scientific Computing Center (NERSC). "With an emphasis on Biological Data Interpretation," said Nikos Kyrpides, DOE JGI Prokaryote Super Program head, "the DOE JGI has played a leadership role in developing, standardizing and providing access for users to high-quality genome assemblies, annotations and other computational genomics tools."

In a study published online June 9, 2015 in The *ISME Journal,* a DOE JGI team developed the first computational protocol called ProDeGe (Protocol for Decontamination of Genomes) for quick and automated removal of contaminant sequences from draft genomes. The ProDeGe tool classifies sequences as either "clean," or "contaminant," and runs, the team reported, at a rate of 0.30 CPU core hours per megabase of sequence. ProDeGe is pre-calibrated to remove at least 84 percent of contaminant sequence, and the team found it performed best when it could compare the test sequence against homologs in the database that corresponded at the Class level or deeper. If the sequences belong to novel organisms, the team reported, ProDeGe removes contaminants solely by checking the sequence composition.

A second study published August 18, 2015 in *Nucleic Acids Research (NAR)* involved a team of DOE JGI researchers and their collaborators who developed and evaluated a new method called Microbial Species Identifier (MiSI) for classifying microbial species that could be supplemented — as needed — by traditional approaches relied on by microbiologists for decades.

The MiSI method developed at the DOE JGI relies primarily on genome sequencing and is a combination of two metrics for determining how closely related two genomes are: genome-wide Average Nucleotide Identity (gANI) and alignment fraction (AF). The original genome-wide gANI metric was developed by Jim Tiedje, Director of the Center for Microbial Ecology at Michigan State University and a DOE JGI collaborator on the Great Prairie Soil Metagenome Grand Challenge project, and his former student Kostas Konstantinidis, one of the study's co-authors, and was modified by the DOE JGI team as a

The 10 genomes of *Burkholderia mallei* and 53 genomes of *Burkholderia pseudomallei* form a single clique group.

🔵 *Burkholderia pseudomallei* genome
🟡 *Burkholderia mallei* genome

Label of node is the degree (no. of edges) of that node.
— Edge not involving a *Burkholderia mallei* genome
— Edge involving a *Burkholderia mallei* genome

The clique-group comprising of pathogenic strains of *Burkholderia mallei* and *Burkholderia pseudomallei* strains highlights the ability of MiSI to capture high similarity as well as subtle differences between closely related genomes. (Figure by Neha Varghese, DOE JGI)

basis for the MiSI method. MiSI speeds up the computations by about 10-fold compared to the gANI metric by using nucleotide sequences of genes and a modified BLAST-based similarity search.

The DOE JGI sequencing pipeline is already using the MiSI method to determine the similarity of newly sequenced genomes to existing reference genomes. MiSI is available for use by the general research community through the IMG system and the data used

Sifting Through Dark Matter Data Redefines Single Bacterial Phylum

To learn more about the uncultivated bacteria and archaea surrounding us, researchers are increasingly relying on culture-independent techniques such as single-cell genomics and metagenomics to fill in the still-unexplored branches on the Tree of Life. By carrying out such investigations, scientists hope not only to better understand the organization of microbial life on Earth but also to have more thorough surveys of potentially useful microbial genes, enzymes, pathways, and capabilities that could help DOE advance its energy and environmental missions.

In a study published online June 19, 2015 in *The ISME Journal,* a team including DOE JGI researchers offers compelling evidence that two groups of uncultivated bacteria actually belong to a single candidate phylum called 'Atribacteria.' These microorganisms are found globally distributed in "low energy" ecosystems where nutrients and oxygen are in short supply and shared metabolisms with other species — syntropy — is a viable survival strategy.

"This paper is the first comprehensive look into the genomic diversity within the 'Atribacteria', a candidate bacterial phylum that's found in a wide variety of habitats such as geothermal springs, wastewater digesters and bioreactors, sub-seafloor environments, petroleum reservoirs, hypersaline microbial mats, and landfill leachates," said Brian Hedlund of the University of Nevada, Las Vegas and a DOE JGI collaborator. "The work was only possible with close collaboration at JGI, particularly Tanja Woyke, who coordinated the Microbial Dark Matter project and helped to bring together several CSPs with significant 'Atribacteria' data."

One candidate bacterial group or phylum called OP9 was first identified from samples collected at Yellowstone National Park's Obsidian Pool. Similar microbial diversity studies from a marine sediment yielded a candidate bacterial phylum called JS1. Samples of these microbes were also found in various locations around the world as part of the DOE JGI's efforts to shed light on what are considered "microbial dark matter," and generate reference genomes of uncultivated bacteria and archaea in order to discover novel genes that could have applications in the fields of energy and environment, as well as to learn more about microbial diversity. Early efforts in this field led to a DOE JGI publication that incorporated collaborators from around the world and multiple Community Science Program projects.

Building on this foundation, and on other succeeding studies, Hedlund's group proposed that the OP9 bacteria belonged to a candidate phylum to be called 'Atribacteria,' while JS1 was thought to belong to a sister candidate bacterial phylum. In this more recent paper, however, researchers combined analysis of 16S ribosomal RNA genes and other highly conserved genes (i.e. phylogenomics) and reported that, despite significant phylogenetic distance and broad habitat range, the OP9 and JS1 lineages share some common features and are closely related enough to be lumped into the phylum 'Atribacteria'.

"There is not a compelling argument for designation of OP9 and JS1 as separate phyla, and the most parsimonious analysis of the available data would suggest that the 'Atribacteria', inclusive of OP9 and JS1, is a single candidate phylum within the Bacteria," the team noted in their report.

Aside from providing the research community with a "first comprehensive view of this group of microbes and prediction of their shared structural and physiological traits," added Jeremy Dodsworth, a co-first author of the study who did his postdoctoral studies with the Hedlund lab and is now at California State University San Bernardino, "we also hope that it will give us clues as to how to grow members of the 'Atribacteria' in the laboratory and study them in isolation." Knowing "what genes are out there" offers considerable promise for better tools for both growing difficult-to-grow microbes, as well as bioproducts relevant to energy and environmental challenges faced by DOE.

Great Boiling Spring, Nevada (Robert Dodsworth)

Liquid gold being poured into a cast at Gold Reef City in Johannesburg, South Africa. (Dan Brown via Flickr CC BY 2.0 license.)

Seeking "Gold Standard" Wastewater Treatments

Natural ecosystems and the diverse communities of microbes (and other organisms) within them are extremely difficult to study because they are not closed systems, where all inputs and outputs can be measured. In a study published July 28, 2015 in *Environmental Microbiology*, a team led by DOE JGI collaborator Jill Banfield at the University of California, Berkeley focused on making the compositions and activities of microbial communities in one specific system more tractable for analyses with genomic methods: microbial communities in laboratory bioreactors that were being studied as a potential method of treating wastewater contaminated by gold ore processing.

Banfield's project comes out of the DOE JGI's Emerging Technologies Opportunity Program (ETOP), launched in 2013 to bring new technologies developed at other institutions into the fold and make them available to its users for energy and environment applications. A half dozen proposals were approved during the Program's first year, among them

one from metagenomics pioneer Banfield. She proposed building and characterizing a pipeline that would allow researchers to isolate and study both near-complete and complete microbial genomes from environmental samples.

Using the laboratory bioreactors, the team reconstructed draft and curated microbial genomes using high-throughput metagenomic sequencing of biofilm and supernatant samples. In one bioreactor, a mixture of cyanide (CN^-) and thiocyanate (SCN^-) was being degraded, while in the other bioreactor, only thiocyanate was being degraded. Cyanide is used for processing gold ore while thiocyanate is a byproduct of the process.

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

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

Expanding Barley Genetics Resources

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

Craig Nagy via Flickr CC BY-SA 2.0 license.

Building off of worldwide efforts, a team involving DOE JGI researchers recently reported that nearly two-thirds of the barley genome's gene space has been mapped. In the study published ahead online August 7, 2015 in *The Plant Journal*, the team identified and sequenced over 15,000 bacterial artificial chromosomes (BACs), comprising roughly 1.7 billion base pairs (Gbp) of sequence out of the

estimated 5.1 Gbp that makes up the barley genome.

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

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

Eucalyptus sole topic of New Phytologist issue

As a fast growing tree found in multiple continents, eucalyptus is of interest to the U.S. Department of Energy as a potential biofuel feedstock. For this reason, the DOE JGI approved a 2008 Community Science Program project to sequence its genome. In June 2014, an international consortium of researchers, including DOE JGI scientists, released the reference genome of *Eucalyptus grandis*.

Showcasing how the global research community has already

University of Pretoria

begun harnessing the eucalyptus genomic resources made available by the DOE JGI through the plant portal Phytozome, the June 2015 issue of *New Phytologist* was dedicated to all things eucalyptus. The effort was spearheaded by DOE JGI collaborator Zander Myburg of the University of Pretoria, one of the leads on the eucalyptus genome CSP project, and a co-author of several of the papers published in the special issue.

"Compared to plant genomes sequenced to date, [eucalyptus] represents an independent evolutionary experiment on what it means to be a large woody perennial plant evolving in diverse, and often stressful, habitats," he and co-author Steven Strauss wrote in the lead commentary for the issue. "The papers in this issue show that the genome sequencing investment by

the US Department of Energy — and the associated inputs and collaborative efforts from many scientific and industrial communities around the globe — is truly an accomplishment to celebrate."

Several studies in the special issue focus on various features of the eucalyptus genome, primarily with an eye toward improving wood production for bioenergy applications, while others are analyses or tools that can further improve the reference genome. One such paper comes from French researchers who developed genetic maps that helped improve the eucalyptus reference assembly last year, and expands on that work. Ultimately, *Eucalyptus* could provide both a sustainable source of bioenergy through its wood, as well as multiple bioproducts based on its well-known oil content.

Susannah Tringe and Axel Visel Named DOE JGI Deputies

Effective August 1, 2015, Susannah Tringe (left) has been named the DOE JGI Deputy for User Programs, and Axel Visel (right) is the DOE JGI Deputy for Science Programs. According to DOE JGI Director Eddy Rubin, "these new roles will enable the JGI to take advantage of a unique opportunity to promote two very talented JGI scientists into senior management positions and also allow them the opportunity to continue running their successful independent research programs." As the User Programs Deputy, Dr. Tringe oversees external user programs and varied DOE JGI external scientific partnerships. As the Science Programs Deputy, Dr. Visel primarily oversees various internal DOE JGI science activities. These new positions fill the previous Deputy Director for Science Programs role held by Jim Bristow prior to his retirement.

Tools for Computational Challenges

by the DOE JGI team for this study are publicly available at https://ani.jgi-psf.org. Learn more about the MiSI method at http://jgi.doe.gov/mising-piece-revealed-classifying-microbial-species-in-the-genomics-era/.

ProDeGe's web interface for uploading and analyzing datasets can be found at http://prodege.jgi-psf.org. Standalone software for ProDeGe can be downloaded from http://prodege.jgi-psf.org/downloads/src can be run on a system with Perl, R, and NCBI Blast. Learn more about ProDeGe at http://jgi.doe.gov/automating-microbial-genome-sequence-decontamination/.

SAVE THE DATE

11th Annual Genomics of Energy & Environment Meeting

March 21 – 24, 2016
Walnut Creek, California

http://bit.ly/JGI-UM11

Who should attend?

All current Community Science Program (CSP) users as well as investigators considering an application for future CSP calls. Also, we welcome any and all researchers and students interested in energy and environmental genomics.

Topics:

Microbial genomics, fungal genomics, metagenomics, and plant genomics; genome editing, secondary metabolites, pathway engineering, synthetic biology, high-throughput functional genomics, high-performance computing applications and societal impact of technological advances. State-of-the-art presentations by invited speakers as well as short talks selected from poster abstracts.